

Communications of the Association for Information Systems

Volume 41

Article 14

9-2017

Using Mechanical Turk Data in IS Research: Risks, Rewards, and Recommendations

Ronnie Jia

Illinois State University, ronniejia@gmail.com

Zachary R. Steelman

University of Arkansas

Blaize Horner Reich

Simon Fraser University

Follow this and additional works at: <https://aisel.aisnet.org/cais>

Recommended Citation

Jia, Ronnie; Steelman, Zachary R.; and Reich, Blaize Horner (2017) "Using Mechanical Turk Data in IS Research: Risks, Rewards, and Recommendations," *Communications of the Association for Information Systems*: Vol. 41 , Article 14.

DOI: 10.17705/1CAIS.04114

Available at: <https://aisel.aisnet.org/cais/vol41/iss1/14>

This material is brought to you by the AIS Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Communications of the Association for Information Systems by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.



Using Mechanical Turk Data in IS Research: Risks, Rewards, and Recommendations

Ronnie Jia

Illinois State University
United States
ronniejia@gmail.com

Zachary R. Steelman

University of Arkansas
United States

Blaize Horner Reich

Simon Fraser University
Canada

Abstract:

With the increasing use of crowdsourced data in behavioral research fields, it is important to examine their appropriateness and desirability for IS research. Extending recent work in the IS literature, this tutorial discusses the risks and rewards of using data gathered on Amazon's Mechanical Turk. We examine the characteristics of MTurk workers and the resulting method biases that may be exacerbated in MTurk data. Based on this analysis, we present a 2x2 matrix to illustrate the categories of IS research questions that are and are not amenable to MTurk data. We suggest that MTurk data is more appropriate for generalizing studies that examine diverse cognition than for contextualizing studies or those involving shared cognition. Finally, we offer a set of practical recommendations for researchers who wish to collect data on MTurk.

Keywords: MTurk, Mechanical Turk, Crowdsourced Data, Research Method, Method Bias.

This manuscript underwent peer review. It was received 01/29/2017 and was with the authors for 2 months for one revision. The Associate Editor chose to remain anonymous.

1 Introduction

The advent of the Internet has continued to afford researchers with new and innovative ways to recruit research participants. One such method is Amazon's Mechanical Turk, or MTurk, an online crowdsourcing labor market open to requesters (those who request work¹) and workers (those who complete work) from around the world. Since its inception in 2005, over 15,000 published papers have referenced MTurk in 10 years (Chandler & Shapiro, 2016). As a fast, inexpensive, and convenient sampling method, MTurk is appealing to behavioral researchers as one can obtain a large sample in hours at a cost as low as a few cents per participant (Behrend, Sharek, Maede, & Wiebe, 2011). Though few studies that have used crowdsourced data have appeared in top IS journals (e.g., Deng & Joshi, 2016; Jenkins, Anderson, Vance, Kirwan & Eargle, 2016; Mamonov & Koufaris, 2014; O'Leary, Wilson & Metiu, 2014; Steelman, Hammer & Limayem, 2014)², its use has risen rapidly in IS conferences and in fields such as accounting (e.g., Brandon, Long, Loraas, Mueller-Phillips & Vansant, 2014; Farrell, Grenier & Leiby, 2017), management (e.g., Chua, 2013), marketing (e.g., Hur, Koo & Hofmann, 2015), and psychology (e.g., Gosling & Mason, 2014).

However, using crowdsourced data is not without controversy. Some authors have cautioned that: "The time has come for organization researchers to begin seriously investigating whether the recent enthusiasm for crowdsourced data is warranted or desirable" (Harms & DeSimone, 2015, p. 189). Though the IS field encompasses more than just organization research, this call for action is still highly relevant. IS researchers must understand the risks and rewards of crowdsourced data and decide whether it is appropriate for their research questions.

This work is further motivated by Steelman et al.'s (2014) comparison of crowdsourced data with other types of data. They found that MTurk samples generally provide data composition and quality equal to that of traditional approaches when adequate controls are put in place. Noting unanswered questions as to why non-U.S. participants led to different results, they called on future research to better understand MTurk workers' characteristics and motivations and the method's validity for other types of research designs and models.

Building on their results, we review literature across several fields and examine the following questions regarding the use of MTurk data in IS research:

1. Who are MTurk workers?
2. What method biases may be exacerbated in MTurk data?
3. Is MTurk data more appropriate for some IS research questions than others?
4. If IS researchers gather MTurk data, how can they ensure its validity?

This paper proceeds as follows: in Section 2, we examine the characteristics of MTurk workers and explore how these unique characteristics can generate method biases that may be amplified through data collection on MTurk. Based on this analysis, we present a 2x2 matrix to illustrate the types of IS research that are and are not amenable to MTurk data in Section 3. In Section 4, we offer a set of practical recommendations for researchers who wish to gather data on MTurk. Finally, in Section 5, we conclude the paper.

2 MTurk Data and Method Bias

Much literature outside the IS field has debated whether the use of MTurk data is legitimate or desirable. While advocates argue that such data is in many ways equivalent to, or even better than, traditional student and organizational samples (e.g., Landers & Behrend, 2015), others have advised caution (e.g., Fleischer, Mead & Huang, 2015; Woo, Keith & Thornton, 2015). Skeptics have warned that MTurk and traditional samples "really do differ both in terms of their quality and in the degree to which conclusions can be drawn from them" and argued that MTurk data are "not of much use" other than in "studies of job-seekers, the recently terminated, [or] the chronically unemployed" (Harms & DeSimone, 2015). The *Journal of Vocational Behavior* even declines to publish research that uses crowdsourced data because they "may threaten the integrity of research samples and the validity of results" due to questions related to

¹ While academic surveys and experiments are the focus of our discussion, most work completed on MTurk are "human intelligence tasks" (HITs) such as tagging pictures, transcribing documents, and testing websites.

² See Table 1 for a more complete list.

“the qualification and identification of respondents, authenticity of their responses, selection bias, representativeness of the sample, and the generalizability of the results” (Elsevier, n.d.). Thus, we must ask: is MTurk data appropriate for all areas of IS research? To answer this question, we first consider the characteristics of MTurk workers and the method biases associated with their context and composition.

2.1 MTurk Worker Characteristics

MTurk offers a diverse, ever-changing group of over 500,000 individuals (Bohannon, 2016) who mostly come from the US and India (MTurk-Tracker.com). Prior research reports that their demographics are similar to that of the general U.S. population (Paolacci, Chandler & Ipeirotis, 2010) and more diverse than standard Internet samples and typical U.S. college student samples (Goodman, Cryder & Cheema, 2013). MTurk also allows researchers to access otherwise hard-to-reach subjects (e.g., addicts; Smith, Sabat, Martinez, Weaver & Xu, 2015) and generalize to broader populations (Buhrmester, Kwang & Gosling, 2011) and across cultures (Eriksson & Simpson, 2010), which avoids oversampling participants from WEIRD (i.e., Western, educated, industrialized, rich, and democratic) backgrounds (Henrich, Heine, & Norenzayan, 2010). Initial research on the validity of MTurk data focused on demographic and psychometric properties and found that samples typically met or exceeded the customary standards (Buhrmester et al., 2011; Goodman et al., 2013), which has led some to conclude that MTurk is no better or worse than student and organizational samples (Landers & Behrend, 2015).

However, mounting evidence has found that MTurk workers *do* differ from U.S. community participants, such as in basic personality traits (i.e., higher in introversion and neuroticism; Feitosa, Joseph & Newman, 2015; Gosling, Rentfrow & Swann, 2003). MTurk workers also value money more than time compared with U.S. community participants because, perhaps, a large proportion of workers are from low-income economies (Goodman et al., 2013) and the majority are underemployed or unemployed (Ross, Irani, Silberman, Zaldivar & Tomlinson, 2010).

Though MTurk boasts over 500,000 members, the real number of active workers is likely much lower: between 15,059 and 42,912 (Fort, Adda & Cohen, 2011). A recent estimate using MTurk worker IDs from 114,000 experimental sessions run over a three-year period by seven psychology labs in the US, Europe, and Australia has similarly placed the number of unique participants at around 30,000 (Bohannon, 2016).

A non-trivial portion of the tasks completed on MTurk has been attributed to these professional workers who treat the site as a full-time job or primary occupation (Berinsky, Huber & Lenz, 2012; Harms & DeSimone, 2015). A recent MTurk sample found that the average participant (mean age = 35) completed 958 tasks (SD = 1,539) over 26.1 hours (SD = 16.1) each week and that 47 percent considered MTurk a career (Deng, Joshi, & Galliers, 2016). A meta-analysis of 132 MTurk studies found that the top one percent of the most prolific workers were responsible for 11 percent of the completed tasks (Chandler, Mueller & Paolacci, 2014). Others have estimated that 20 percent of the active workers perform 80 percent of the tasks (Deneme, 2009), which translates into 3,011 to 8,582 active Turkers (or 0.6 to 1.7% of all registered members) (Fort et al., 2011). The worldwide collaborative effort to estimate the effective MTurk research population similarly found that “the true number of Turkers that are willing to take part in an experiment at any one time is only about 7,300” (Bohannon, 2016). One likely member of this group is a U.S.-based stay-at-home mother, who has completed 20,000 academic surveys on MTurk in five years (i.e., averaging 11 surveys per day) while caring for her toddler son. She earns half of her income on MTurk mostly through completing academic surveys (Marder & Fritz, 2015).

In Section 2.2, we discuss how the specific attributes of this anonymous MTurk workforce, if not addressed, can exacerbate several method biases that limit the validity of MTurk data and the techniques that one can use to alleviate them.

2.2 Method Bias Resulting from Gathering Data on MTurk

Method bias refers to the difference between the measured score and the true score and stems from one or more elements of the method (i.e., rater, instrument, procedure) used to obtain the measured score (Burton-Jones, 2009). If researchers fail to mitigate method biases, they may erroneously attribute results caused by the method to hypothesized effects.

Data collected on MTurk, similar to those gathered using traditional methods, are subject to a variety of method biases. We highlight five key biases that we believe are particularly pertinent in the MTurk context: 1) self-selection (by both eligible and ineligible participants), 2) non-independence, 3) attentiveness, 4) ability, and 5) social desirability. While some of these biases (e.g., self-selection by ineligible participants,

ability) have not typically been a significant concern in traditional IS samples, one must carefully manage and address them when using crowdsourcing markets such as MTurk. Though traditional samples are still subject to many of these biases, we argue that their extent and severity in MTurk data can be much higher due to the unique characteristics of the MTurk platform and its workers.

2.2.1 Self-selection

Since researchers usually cannot gather data from the entire population of interest, they typically seek a randomly selected sample, where any particular participant's inclusion or exclusion is entirely due to chance (Yates, Moore & Starnes, 2008). Self-selection bias arises when individuals with certain characteristics select themselves into the study due to various motivations, which can lead to a biased sample (Winship & Mare, 1992). Though researchers have often discussed self-selection in the context of eligible participants' selecting themselves into (or out of) a study, it also applies to ineligible individuals who misrepresent their identities or background to gain inclusion. We discuss these two types of self-selection in turn.

Self-selection by eligible participants: to understand how an MTurk worker chooses a particular task from all available tasks (over 100,000 at any given time per MTurk-Tracker.com), we take a deeper look at the MTurk platform and its online community (e.g., MTurk Forum, Turkopticon). At its core, MTurk is a marketplace for freelance individuals (i.e., independent contractors) who expect to be paid a fair amount for each task completed (Deng, Joshi & Galliers, 2016). Workers have autonomy over which tasks they choose and typically base this decision on the level of pay in relation to the time and effort required. This incentivized motivation creates an initial self-selection bias in both the MTurk platform and in which specific tasks a worker completes on the platform. Though money is not the only motivation for workers, it is the most prominent one. Research has shown that money, fun, and learning (in that order) are the primary motivating factors for MTurk workers (e.g., Behrend, Sharek, Meade & Wiebe, 2011; Litman, Robinson, Rosenzweig, 2015). However, regardless of worker motivation, the mechanism is the same (namely, self-selection).

To help optimize their time and pay, MTurk workers as a collective have developed a variety of online tools and resources such as community forums, auto-completion techniques, and browser plug-ins to help search, select, and rate which tasks to complete. These third party forums typically provide a running list of threads that link to tasks on MTurk (Schmidt, 2015) and reviews about instructional manipulation checks, scales, and experimental conditions (Chandler et al., 2014; Gosling & Mason, 2015). Using these forums can influence workers' behavior and, subsequently, affect sample composition and research results in that better-paying and less-effortful studies will attract workers who frequently visit such forums (Schmidt, 2015). Some forums report not only the rates of pay and the amount of time needed to complete tasks but also whether the study employs eligibility or data quality checks³ (Chandler et al., 2014). Seeking and disseminating information about the "worthiness" of studies in this way can lead to biased samples and pose a threat to a study's validity (Gosling & Mason, 2015; Harms & DeSimone, 2015; Woo et al., 2015)—especially those that rely on specific manipulation and validity checks for participation. These forums can also cause an unintended shift in the sample composition (Chandler & Shapiro, 2016). For example, Chandler et al. (2014) found an unusually high proportion of men in their MTurk sample, which they later attributed to a post about the study on Reddit, a website that more men than women frequent.

Though self-selection may also be present in traditional samples such as with student participants' sharing their experience with others after their completion, MTurk workers share information "in a more systematic, permanent, and searchable manner, with more dramatic consequences for data validity" (Chandler et al., 2014). We next examine self-selection by ineligible participants, which can present an even greater challenge to MTurk data.

Self-selection by ineligible participants: existing IS studies have not closely examined ineligible participation due possibly to its not being a prominent concern in traditional IS samples and research designs since researchers typically have reasonable knowledge of and control over which participants are included⁴. For example, an investigator who gathers student data can target either specific types of students (e.g., IS majors) or the broader student body. Similarly, researchers who desire an organizational

³ Forum moderators typically monitor for individuals posting detailed information about a task (e.g., attention checks) that may invalidate future work when possible.

⁴ Ineligible participation has been a significant concern in medical fields, where researchers have long dealt with participants in medical trials who falsify their inclusion criteria—especially when these requirements are explicitly stated (Kramer et al., 2014).

sample can work with their contacts in participating organizations to identify appropriate respondents (e.g., programmers, project managers) or directly recruit participants using membership directories from industry or professional associations (e.g., ACM).

The lack of attention to ineligible participation may also be related to the fact that, in contrast to the payment mechanism on MTurk, many IS research designs that use traditional samples (e.g., students) often provide non-financial incentives (e.g., extra course credit) that may have limited appeal to an ineligible person who may seek inclusion. While this fact certainly does not imply that all the participants in traditional IS samples are definitively eligible, we do argue that researcher control, participant anonymity, and incentive mechanisms differ with MTurk compared to traditional student or organization samples.

Prior research indicates that data from non-U.S. workers can have poor quality (Feitosa et al., 2015) and even fail to support long-established theories (e.g., TAM; Steelman et al., 2014), which is a major concern given the high percentage of non-U.S. workers on MTurk (Paolacci et al., 2010). Therefore, researchers have recommended that one should use U.S.-only MTurk samples until the causes of these differences are determined (Steeleman et al., 2014). However, filtering non-U.S. workers out of the sample using an MTurk country-of-origin screen can be problematic because they can access several websites that show them how to set up fake addresses and accounts to make themselves eligible for U.S.-only studies (Harms & DeSimone, 2015). Indeed, research has shown that a significant proportion of self-described U.S. participants misrepresent their country of origin (e.g., Berinsky et al., 2012; Feitosa et al., 2015; Rand, 2012). Since the MTurk platform is built around worker anonymity, researchers cannot easily establish participant eligibility, which is highly important in certain types of research.

Besides country of origin, anonymity also makes an MTurk worker's employment status unverifiable. For many workers, MTurk is their primary source of income (Paolacci et al., 2010), and many report that they are "employed" part-time or full-time, which means they are either interpreting working on MTurk as employment or simply misrepresenting their identities in order to participate and receive pay (Harms & DeSimone, 2015; Smith et al., 2015). Thus, using self-report questions to screen workers' employment status and type may be ineffective (Siegel, Navarro & Thomson, 2015).

In sum, though the same factors (e.g., money) motivate both self-selection types on MTurk and though both are undesirable, ineligible participants, who are typically not an issue in traditional IS samples, can be particularly problematic if the research requires relevant background knowledge to ensure data validity. While MTurk workers are no more dishonest than participants in traditional samples (Chandler & Shapiro, 2016), the anonymity and incentives built into the MTurk platform do reduce researcher control and provide a fertile ground for identity falsification and self-selection.

2.2.2 Non-independence

Non-independence arises when observations in a sample or across samples are not independent from one other (Kenny & Judd, 1996). Since third party sites offer workers a place for social connections (Deng et al., 2016), data collection on MTurk will always be somewhat of a snowball method in that at least a subset of participants will have weak social ties with each other, which may raise concerns of non-independence across observations in a sample (Schmidt, 2015). However, such social ties among participants may be present in traditional samples as well.

Non-independence can be further exacerbated when a small percentage of workers complete a disproportionately large number of tasks as on MTurk (e.g., Chandler et al., 2014; Deneme, 2009; Fort et al., 2011). Berinsky et al. (2012) found that, even in their own set of seven studies conducted in a four-month period, 30 percent of their total sample appeared in at least two of the studies and two percent were present in five or more. These individuals create a significant concern for non-independence not only across other researchers' studies but also across one's own repeated data collections.

Concerns for non-independence can also exist across studies in the form of practice effects and participant non-naïveté (Chandler et al., 2014). MTurk constantly presents workers with opportunities to participate in academic studies. In fact, 98.2 percent of MTurk respondents in a recent survey reported such prior experience (Deng et al., 2016). Though an undergraduate student in the US may be about 4,000 times more likely than an average U.S. person to participate in a research study (Henrich et al., 2010), a survey of both college students and MTurk workers (Rand et al., 2014) found that the median Turker participates in more experimental studies in a week (20 per week or 300 in total) than the median traditional lab subject does in a lifetime (15 in total or 1 per week).

Since researchers typically assume that participants are naïve about the method, instruments, or manipulations in the study (Burton-Jones, 2009), repeated exposure can have significant implications for the validity of the results (e.g., reduction in effect size)—especially in studies that use foundational measures or classical experiments (Chandler et al., 2014; Chandler & Shapiro, 2016). For example, research has found MTurk worker performance on cognitive tasks to correlate with the number of tasks they have completed (Chandler et al., 2014). Repeated exposure to attitudinal measures also provides participants with an opportunity to elaborate on their beliefs and change subsequent responses (Sturgis et al., 2009). Having observed a change in her own survey performance, one professional survey taker noted that: “It’s hard to reproduce a gut response when you’ve answered a survey that’s basically the same 200 times” (Marder & Fritz, 2015). Other workers report increased self-awareness, such as a better grasp of their political and religious attitudes, through research participation. One worker said: “You just can’t fill out 40,000 surveys and not get a better sense of what you do and don’t think” (Marder & Fritz, 2015). Even the percentage of workers who correctly answer attention check questions has significantly increased over time (Chandler et al., 2014). Indeed, it would be hard not to if one has encountered the “Have you ever had a fatal heart attack?” question hundreds, if not thousands, of times before (Marder & Fritz, 2015).

In sum, non-independence can be a more significant problem in MTurk data than in traditional samples both across repeated data collections in the same study and across different studies. Though perhaps less of a concern for an individual researcher who gathers MTurk data, the cumulative effect of building a research stream (e.g., TAM) or a field (e.g., marketing) with data from a small group of MTurk workers can be an alarming prospect for an academic field as a whole (Marder & Fritz, 2015).

2.2.3 Attentiveness

Attentiveness is a function of a rater’s motivation (Podsakoff, MacKenzie, & Podsakoff, 2012). It can range from occasional lapses in attention to intentional random responses throughout a survey (Huang et al., 2015). Attentiveness is not a problem limited to professional MTurk workers, though they are known for providing inferior-quality data (Mason & Suri, 2012). In a rush to complete tasks quickly and maximize income, these workers are less likely to make the necessary cognitive effort to provide accurate answers and, thus, are more prone to using a smaller range of answers than casual users do, which can result in greater peakedness in the distribution (Deneme, 2009).

According to Fleisher et al. (2015), 15 to 20 percent of MTurk participants respond carelessly and fail data validity checks; in some instances, this ratio can reach as high as 42 percent. Another researcher has also reported rejecting about 20 percent of American and 50 percent of Indian worker responses for the same reason (Bohannon, 2011). Such widespread inattentiveness can produce systematic bias that inflates correlations between variables, increases type I error rates, and threatens the internal validity of findings (Huang, Curran, Keeney, Poposki, & DeShon, 2015).

Evidence continues to mount that crowdsourcing workers consistently multitask during their participation by watching TV, listening to music, sending instant messages, or using their cellphone while completing tasks (Chandler et al., 2014; Clifford & Jerit, 2014). One professional survey taker, a young mother, worked on MTurk “while negotiating her toddler’s milk bottles and giving him hugs” (Marder & Fritz, 2015). Though such distraction and inattentiveness may be less of a concern for a survey study that comprises randomized items, it can be a particular challenge for experimental research with a complex cognitive or learning element that requires reaction times or manipulation checks. Unfortunately, unlike lab settings, researchers have reduced control in online studies and must take additional measures to address this concern.

Though researchers recognize the need to identify inattentive workers (e.g., Crowston, 2011), these workers do not typically respond with robotic-style answers and are not easily detectable using existing methods. The effectiveness of attention checks diminishes over time (Chandler et al., 2014), especially for workers who see them possibly dozens of times each week (Marder & Fritz, 2015). In fact, even in our own research, we have had varied success with many of these screening methods and quality checks because workers typically know about and search for them in advance (through their own experience or via reviews on third party websites posted by prior participants) and specifically monitor them while ignoring or skimming over focal items.

To be fair, everyone’s life has distractions. Though participants in traditional samples (e.g., students who fill out a survey in a classroom or perform a task in a lab, employees who complete a survey at work) may

also be subject to inattentiveness, its variety, extent, and severity is likely lower and more remediable through research design. In addition, traditional student and organizational participants typically do not know the various quality control checks in advance or as easily as many MTurk workers do, and survey respondents in a traditional sample typically have no incentive to speed through the questions so as to move onto the next (paid) task.

2.2.4 Ability

Ability bias occurs when a rater has difficulty understanding instructions or questions or performing required tasks (Podsakoff et al., 2012). As we discuss above, ineligible participants who have nonetheless selected themselves into a study often lack the necessary background to provide valid responses (e.g., a non-smoker completing a survey of smokers), which creates an ability-related bias.

This bias also results from the diverse MTurk population, which comprises workers with varied abilities. With an anonymous platform, researchers have little information about worker ability and must rely on workers' prior approval ratings, which only indicate the percentage of tasks they have completed successfully. Though most research on MTurk uses English, researchers have found that non-native English speakers produce poor-quality data (Deneme, 2009; Feitosa et al., 2015; Mason & Suri, 2012). In particular, Indian workers struggle with reverse-coded items and instructional manipulation checks (Litman et al., 2015), which evidences key language issues (Chandler & Schapiro, 2016). Because those low in verbal ability tend to respond in a non-differentiated manner (Krosnick, 1999), many non-native English-speaking workers may not be able to comprehend instructions and survey items, which leads to ability-related response biases (Podsakoff et al., 2012).

In sum, researchers who use traditional methods can reasonably manage ability bias with their research design and choice of participants. However, these controls are often unavailable or more difficult to implement on an anonymous platform such as MTurk.

2.2.5 Social Desirability

Social desirability bias occurs when a rater responds negatively to socially undesirable traits or positively to desirable traits in an untruthful fashion (Nederhof, 1985). While MTurk workers' anonymity may encourage them to respond truthfully and, thus, reduce social desirability, other aspects of MTurk may serve to heighten concerns for such bias in relation to traditional samples. For example, since workers only receive pay after the task requester (researcher) has approved their work, they may respond in ways that they think the requester desires to ensure payment (McGonagle, 2015). A requester's denying a worker's submission (which means the worker does not receive pay) reduces the worker's approval rating, which many studies use as a primary filter and requirement and has direct implications on the future tasks available to the worker in the marketplace.

Similar to participant eligibility and ability issues, social desirability becomes a larger problem for those who frequent third party forums. Workers with prior knowledge about the study, requirements, and research objectives may falsify their data in order to receive a pay incentive for their time completing a biased survey. While biased or falsified responses may also affect traditional samples, the presence of a similar system to share information across the population, along with a payment mechanism that requires the task requester's approval, represents a unique challenge to MTurk data.

In summary, though traditional samples may also suffer from some of the method biases we examine here, MTurk greatly exacerbates them due to the characteristics of the platform, its workers, and its community. The one underlying factor for each of these biases is financial incentive: it is primarily because of money that workers join MTurk and select themselves, regardless of eligibility, into a study (self-selection) or multiple phases of the same or similar research (non-independence) and speed through a study to quickly move on to the next task (attentiveness) even when they lack the verbal skills to understand instruction and questions (ability). Workers may also respond to questions in ways that they think will ensure payment (social desirability). Thus, financial gain is the primary motive of MTurk workers, which is exacerbated by a cloak of anonymity and access to other community members with similar goals.

Could these biases be so problematic that they render the MTurk platform infeasible for certain research topics? We examine this question in Section 3 and discuss the types of research questions that are and are not amenable to MTurk data.

3 MTurk Data and the Research Question

Each sampling strategy has implications for both internal and external validity (Landers & Behrend, 2015). In many cases, the internal validity of the results rests on sampling specific populations (McGonale, 2015). In these situations, the underlying assumptions and boundaries of the research question are essential. In this section, we develop a 2x2 matrix based on Zhu, Barnes-Farrell, and Dalal's (2015) categorization of research questions and illustrate the appropriateness of MTurk data for each cell.

3.1 "Generalizing" vs. "Contextualizing" Research

Broadly speaking, research questions can span the continuum that ranges from "generalizing" to "contextualizing" (Zhu et al., 2015). We argue that, while MTurk samples may be appropriate in a generalizing study in which one investigates generic attitudes and behaviors (e.g., TAM, Steelman et al., 2014), they can threaten the validity of the results in a contextualizing study in which the particular setting and the identity of the participants are critical (Zhu et al., 2015).

For example, if a researcher examines a cross-cultural phenomenon or one that requires a broad range of socioeconomic statuses (Woo et al., 2015), such as social media adoption and use, then MTurk represents a legitimate or even an ideal source of data. However, if one studies IT employee burnout in Silicon Valley firms and does not seek to generalize to the overall IT workforce, sourcing participants from that particular population of interest is indispensable to the study's internal validity. It would be inappropriate to recruit MTurk workers whose employment status and professional background are not verifiable. Woo et al. (2015) recommend that researchers should avoid MTurk if participants' misrepresenting their identities would invalidate their studies.

Since participant eligibility and ability are critical in contextualizing work such as alignment (e.g., Reich & Benbasat, 2000) and other organizational IS research, these concerns may restrict such studies to only using samples from actual organizations. Even researchers who advocate using MTurk data (e.g., Landers & Behrend, 2015) acknowledge that there exists scant evidence that demonstrates the usefulness of crowdsourced data in research on organizational issues. Little wonder the *Journal of Vocational Behavior* cites challenges in "qualification and identification of respondents" as a reason to not publish crowdsourced data.

A major strength of organizational samples is that researchers can be reasonably confident that the respondents actually work for that organization (McGonagle, 2015) and have the necessary background and ability to respond to questions. Researchers may also establish what jobs the participants hold and, with careful design, verify which work unit they work in and other pertinent job-related information (Jia & Reich, 2013).

Of the five method biases we discuss above, ineligible participation and ability biases likely present the greatest challenges when using MTurk data in contextualizing studies. MTurk may be better suited for generalizing research that imposes less stringent criteria for inclusion.

3.2 Diverse vs. Shared Cognition

One can also categorize research questions along a second continuum that ranges from "diverse" to "shared" experiences (Zhu et al., 2015). If a study is on the former end of the continuum (i.e., one interested in understanding how people with different experiences and backgrounds exhibit the same psychological processes, such as how individual perceptions of a technology influences adoption), then using an intentionally diverse sample can rule out context-specific explanations and enhance external validity (Zhu et al., 2015). In this case, crowdsourced samples may be desirable given the workers' varied past experiences.

However, if the research focuses on "psychological processes that need participants to have a shared experience, such as emergence of culture or leadership", then it would require one to use "an organizational sample wherein employees experience the same event" (Zhu et al., 2015, p. 230). Such is the case for research on employees' shared cognition or perceptions in IT teams (e.g., Faraj & Lee, 2000; Guinan, Cooperider & Faraj, 1998; Jia & Reich, 2013).

In studies of shared mental models, advantages of organizational samples over crowdsourced data are again clear: one can gather data from multiple raters from the same work unit and assess sharedness across responses. As a marketplace of independent, anonymous workers, the current MTurk platform

does not support data collection from an organization that exists outside MTurk⁵. This limitation concerns the platform itself and does not relate to the method biases we discuss above.

3.3 A 2x2 Matrix

The 2x2 matrix (Figure 1) depicts the two continua and the appropriateness of MTurk data in different types of inquiry. Given the potential for heightened method biases in MTurk data (particularly self-selection by ineligible participants and participant ability) and the nature of the MTurk platform as a marketplace of individual workers, MTurk is not recommended for research on participants' shared cognition in a specific context (cell III). However, research that deals with diverse cognition and is generalizing to a wider population can be a great opportunity to use MTurk data (cell I). Researchers should exercise caution with topics in the other two cells, explicitly justify why they used MTurk data, and describe the subsequent controls and techniques they used to address any potential validity concerns. To further explain the matrix, we present a single IS research stream (i.e., TAM) that spans all four cells depending on the nature of the research question. We discuss the appropriateness of using MTurk data in each cell and then use the matrix to categorize existing IS work that has used MTurk data.

<i>Generalizing study</i>	Cautioned	Appropriate
	IV	I
<i>Contextualizing study</i>	Not Recommended	Cautioned
	III	II
	<i>Shared cognition</i>	<i>Diverse cognition</i>

Figure 1. Is MTurk Data Viable for Your Research Topic?

3.3.1 Cell I: Generalizing Study, Diverse Cognition

Early TAM research (e.g., Davis, 1989; Davis, Bagozzi & Warshaw, 1989) focused on how perceived usefulness and ease of use influence user intentions. The research generalized to a variety of users and technologies and dealt with diverse individual cognition. The research questions imposed few requirements on participant eligibility and ability: participants did not need to work in the same organization or have experience with the examined technologies. Therefore, one could replicate the core TAM theory (e.g., Steelman et al., 2014) via MTurk and provide both empirically and theoretically valid results.

3.3.2 Cell II: Contextualizing Study, Diverse Cognition

As research on TAM expanded, researchers began to contextualize the theory. For example, Brown and Venkatesh (2005) examined Internet adoption in households. Their focal research question concerned not only those who had adopted technologies in their households but also those who had chosen not to adopt.

⁵ However, an experimental researcher could conceivably recruit an ad hoc group of MTurk workers, introduce some stimuli, and then examine the shared cognition that may arise in the group.

While MTurk does offer a diverse set of individuals, no worker on MTurk could credibly claim to be a non-adopter, which would undermine an MTurk sample's validity due to the absence of non-adopters in the sampling frame. The lack of external validity resulting from the selection bias in the MTurk data would prevent a valid test of the research question.

However, if the level of contextualization is such that including MTurk workers in the sample would be theoretically possible due to their *known* characteristics (e.g., Internet use, country of origin), researchers may be able to justify using MTurk data by carefully creating and reporting the design methods used to validate their inclusion in the sample (see examples in Section 4).

3.3.3 Cell III: Contextualizing Study, Shared Cognition

As the TAM literature matured, research questions became further contextualized and involved shared perceptions. For example, Venkatesh and Sykes (2013) used social network analysis to examine the adoption of Internet kiosks in rural villages in India. In this study, villagers had to be familiar with not only the context but also other participants in the study. Therefore, the research question relied on a highly context-specific sample of villagers with shared cognition regarding the use of the specific Internet kiosks. Due to potential for ineligible participation, it would be inappropriate to use an MTurk sample regardless of whether it provides psychometrically sound data or confirming evidence for TAM.

3.3.4 Cell IV: Generalizing Study, Shared Cognition

As mounting evidence pointed to the power of social influences in guiding IT adoption and use (e.g., Venkatesh, Morris, Davis, & Davis, 2003), subsequent TAM research began to study social impacts in groups. Sykes, Venkatesh, and Gosain (2009) investigated how peer support in one's social network influences whether one adopts and uses an organizational system. While the research question involved a generalized approach, it required the individual respondents to have shared cognition of their environment to construct a valid social network map. Since MTurk participants are anonymous and independent, one could not construct a network map. Therefore, while MTurk does allow researchers to tap a diverse set of perceptions from individuals, MTurk data brings significant validity issues when examining a research question concerning shared cognition.

Although shared cognition presents a significant problem in using MTurk data because it requires both the focal and aggregate participants to be eligible for the study, research that uses a quasi-experimental design to create short-term, small-group trials in an online environment (e.g., Horton, Rand, & Zeckhauser, 2011) might use MTurk participants successfully. However, in examining shared cognition that exists with more specificity (e.g., witnesses of the Tunisian uprising), at a higher level (e.g., department or organizational), or over a longer time period (e.g., outsourcing relationships), researchers should take specific measures to ensure participant eligibility.

3.3.5 Implications of the Matrix

Based on the above set of examples, we argue that researchers should not accept MTurk as a data-collection tool without first examining the nature of the research question. Fields that have largely embraced crowdsourced data (e.g., marketing) are primarily situated in cell I and typically focus on generalizing studies of individuals with diverse attitudes, which impose minimal participant eligibility and ability requirements. Fields that go beyond cell I (e.g., organizational research) have exercised more caution with crowdsourced data due to concerns of participant eligibility and ability in an anonymous crowd, which explains why some of their outlets (e.g., *Journal of Vocational Behavior*) reject crowdsourced data.

As this set of TAM examples shows, IS research questions span all four cells: they cover both generalizing and contextualizing research questions that involve diverse and shared cognition. Therefore, IS researchers must examine their research question to ensure that crowdsourced data is a good fit for their specific study. This is particularly important for studies in cells II, III, and IV, which require additional measures to ensure data validity. Researchers should explicitly discuss their strategies to mitigate method biases and clearly present the rationale and justification to readers.

Table 1 summarizes publications in top IS journals that have used MTurk participants in survey-based, quantitative research designs. These papers span cells I and II, and the authors exhibit varied levels of awareness of the biases typically associated with MTurk data. A separate keyword search of "Mechanical Turk" in *ICIS*, *AMCIS*, and *PACIS* proceedings yielded another 155 papers (as of March 2017). Therefore,

we can expect many more studies that use MTurk data to appear in top IS journals in the near future. We hope this discussion provides researchers and reviewers with a useful guide to use and evaluate MTurk data across a variety of research questions.

Table 1. MTurk-related Research in Senior Scholars' Basket and Other AIS Journals

Author	Journal	Research focus	Location in the matrix	Method bias discussed
Mamonov & Koufaris (2014)	<i>Communications of the AIS</i>	The impact of privacy breaches on mobile phone users' attitudes and intentions to terminate their relationship.	Cell I	None
Steelman et al. (2014)	<i>MIS Quarterly</i>	A replication of TAM models in student samples, professional panels, and online crowdsourcing workers.	Cell I	Attentiveness, self-selection
O'Leary et al. (2014)	<i>MIS Quarterly</i>	The impacts of actual and perceived proximity on relationship quality in collocated and virtual team members.	Cell II	Ability, attentiveness
Soror, Hammer, Steelman, Davis, & Limayem (2015)	<i>Information Systems Journal</i>	The effect of mobile phone use habits and self-regulation on mobile phone use and negative consequences.	Cell I	Attentiveness, non-independence, social Desirability
Kehr, Kowatsch, Wentzel, & Fleisch (2015)	<i>Information Systems Journal</i>	The effect of general privacy concerns, general institutional trust, and affect in privacy calculus estimations	Cell I	Attentiveness
Chen & Horton (2016)	<i>Information Systems Research</i>	The behavioral response to wage changes in online labor markets.	Cell II	Attentiveness, non-independence, self-selection
Jenkins et al. (2016)	<i>Information Systems Research</i>	The effect of dual task interferences on the interrupting task.	Cell I	None
Baird, Miller, Raghu, & Sinha (2016)	<i>Information Systems Research</i>	The impact of piracy controls and product line extensions on welfare in a consumer software market context.	Cell I	Attentiveness
Lowry, Zhang, Wang, & Siponen (2016)	<i>Information Systems Research</i>	Examination of adult cyberbullying on social media platforms.	Cell II	Ability, attentiveness, non-independence, self-Selection, social Desirability
Lowry & Wilson (2016)	<i>Journal of Strategic Information Systems</i>	The influence of IT service climate on IT service quality and IT agility in organizations.	Cell II	Attentiveness, self-selection
Deng & Joshi (2016)	<i>Journal of the AIS</i>	Examination of individual motivations to participate in micro-task crowdsourcing platforms.	Cell II	None
Hibbeln, Jenkins, Schneider, Valacich, & Weinmann (2017)	<i>MIS Quarterly</i>	Examination of user emotions through mouse-movement estimations.	Cell I	Ability

Note: we exclude studies that involve MTurk workers but do not use them as participants in survey-based quantitative research (e.g., Deng et al., 2016; Kane & Ransbotham, 2016) from this summary.

In sum, we do not argue that crowdsourced data is invalid. To the contrary, we believe that crowdsourced platforms such as MTurk can be an efficient tool for many research questions. However, concerns over certain method biases (e.g., participant eligibility, ability) may prevent one from using MTurk data in certain types of research. After all, the research question should drive the sampling strategy rather than the sheer convenience that a particular method may provide. Next, we make some practical recommendations for researchers who use MTurk in their designs.

4 Practical Recommendations for Gathering MTurk Data

Researchers who have considered the risks and rewards and elected to use crowdsourced data should work to ensure the validity of their research design and clearly present their methods and related decisions to reviewers. In Table 2, we present a set of safeguards that one can implement before, during, and after collecting data to help minimize method bias. These suggestions are specific for users of the MTurk platform and complement general recommendations for data collection.

Table 2. Recommendations for Collecting Data on MTurk

Timing	Recommendation	Bias addressed	Rationale
Before data collection	Screen participants by IP address and restrict access to specific areas (e.g., US).	<ul style="list-style-type: none"> • Self-selection • Ability 	<ul style="list-style-type: none"> • Non-U.S. participants may lead to different conclusions (Steelman et al., 2014). • Less fluent non-native English speakers produce low-quality data (Feitosa et al., 2015). • However, this screen is not always effective (Harms & DeSimone, 2015).
	Offer moderate compensation.	<ul style="list-style-type: none"> • Self-selection 	<ul style="list-style-type: none"> • Higher pay does not necessarily lead to better data quality (Smith et al., 2015) and may attract a disproportionate number of professional workers and incentivize non-U.S. workers to misrepresent their identity (Schmidt, 2015). • Lower pay may prolong data-collection periods (Buhmester et al., 2011; Steelman et al., 2014).
	Restrict participants to those with high approval ratings on MTurk.	<ul style="list-style-type: none"> • Attentiveness • Ability 	<ul style="list-style-type: none"> • Workers with an approval rating of 95 percent or higher tend to produce higher-quality work (Peer, Vosgerau & Acquisti, 2014).
	Warn participants that inattentive respondents will not be paid ⁶ .	<ul style="list-style-type: none"> • Attentiveness 	<ul style="list-style-type: none"> • Participants who fail to receive payment will receive a negative mark on MTurk, which likely affects their ability to get work in the future (Fleischer et al., 2015).
	Revise items to use neutral wording.	<ul style="list-style-type: none"> • Social desirability 	<ul style="list-style-type: none"> • This strategy is one of many that Nederhof (1985) recommends to mitigate this bias.
During data collection	Explain the importance of the study.	<ul style="list-style-type: none"> • Attentiveness 	<ul style="list-style-type: none"> • This recommendation may reduce inattentive responding (Fleischer et al., 2015; Goodman et al., 2013).
	Filter eligible and ineligible participants into different areas of the study based on their individual characteristics or identities.	<ul style="list-style-type: none"> • Self-selection • Social desirability 	<ul style="list-style-type: none"> • This recommendation may reduce participants' motivation to misrepresent their identity (since they will get paid either way) though this strategy may increase the overall cost of data collection (Smith et al., 2015).
	Gather a large sample.	<ul style="list-style-type: none"> • Attentiveness • Self-selection 	<ul style="list-style-type: none"> • A large sample (e.g., $n > 800$) can keep the proportion of professional workers low (Harms & DeSimone, 2015).
	Embed quality-control measures <ul style="list-style-type: none"> • attention checks, and • comprehension checks. 	<ul style="list-style-type: none"> • Attentiveness • Ability 	<ul style="list-style-type: none"> • These measures may reduce inattentive responding (Goodman et al., 2013). • However, such measures are likely less effective with experienced workers (Marder & Fritz, 2015).
	Include additional questions (e.g., qualitative) to establish participant identity and ability.	<ul style="list-style-type: none"> • Ability • Self-selection 	<ul style="list-style-type: none"> • Additional questions allow the researchers to further assess whether participants are appropriate for the sample.
After data collection	Remove all responses that fail quality control.	<ul style="list-style-type: none"> • Attentiveness • Ability • Self-selection 	<ul style="list-style-type: none"> • This recommendation can improve data quality (Harms & DeSimone, 2015).
	Use workers' IDs to remove those who have responded in more than one phase of a multi-phase study or prior study.	<ul style="list-style-type: none"> • Non-independence • Self-selection 	<ul style="list-style-type: none"> • This recommendation can reduce non-independence between samples (Berinsky et al., 2012) • This recommendation can limit participation by professional workers. However, it may also eliminate prior participants who are not professional workers (false positives) (Aust, Diedenhofen, Ullrich, & Musch, 2013).

Additionally, researchers who find their work positioned in potentially inappropriate cells should further detail and justify their approach to reviewers and readers. For example, verifying an individual's employment status on MTurk requires additional efforts and controls in place to provide adequate evidence because self-reported employment status is unreliable (Siegel et al., 2015). Approaches to consider may include asking for a verifiable work email address⁷. In order to ensure participants' identities

⁶ Note that some institutional review boards (IRB) may view this practice as violating participants' right to withdraw without penalty or loss of benefit (Fleischer et al., 2015) as outlined in *Withdrawal of Subjects from Research Guidance (2010)* from U.S. Department of Health & Human Services (<http://www.hhs.gov/ohrp/policy/subjectwithdrawal.html>). Due to these human subject rules, many researchers pay all participants, including the inattentive ones (e.g., Crowston, 2012).

⁷ While this method may provide the necessary validation, it violates Amazon's terms of service for the MTurk platform (see https://requester.mturk.com/help/faq#restrictions_use_mturk). In studies where participant validation is critical, researchers may wish to use a crowdsourcing platform designed with less worker anonymity (e.g., uTest, SurveyMonkey; Lease et al., 2013).

and abilities, researchers might collect detailed qualitative insights from them (e.g., their job roles) and use this information to rigorously validate their responses.

The burden of proof is on researchers to demonstrate how they have addressed method biases in their research design and data analysis. If they fail to adequately address these biases, reviewers could reasonably request additional data collection from a traditional sample to provide corroborating evidence.

5 Concluding Remarks

Deciding whether a sampling strategy is acceptable or desirable often requires both theoretical and empirical considerations. While MTurk samples may provide satisfactory psychometrics and demographic compositions (e.g., Steelman et al., 2014), few have examined the theoretical validity itself of using a crowdsourced sample for their specific research questions.

In this paper, we examine the characteristics of the MTurk platform and its workers and discuss how they can exacerbate a set of method biases compared to traditional sampling techniques. Of particular concern are participant eligibility and ability, which likely pose the greatest challenges to “contextualizing” studies. We also note that research on “shared cognition” in organizations cannot use MTurk data due to the platform’s independent and anonymous nature. In these situations, the potential benefit from a convenient crowd may be offset by the internal validity issues it creates.

However, MTurk data may be more appropriate for “generalizing” studies that investigate a generic attitude or behavior exhibited by people with “diverse cognition”. In these situations, crowdsourced data are likely appropriate (e.g., Reis & Gosling, 2010), though researchers still need to be aware of the increased risks for non-independence, attentiveness, and social desirability biases in such data vis-à-vis traditional samples.

Note that we describe the continuum between generalizing and contextualizing studies in relative terms because there is no fine line between them. Steelman et al.’s (2014) results offer such a cautionary tale: a seemingly generalizing TAM study can end up appearing quite contextualized when the model is supported by U.S.-based workers, but not by the worldwide MTurk sample. Is TAM more culture specific than we realize? Or do the differing results arise due to the biases from non-U.S. workers? Which is the problem: the theory or the MTurk data?

If TAM turns out to be culture specific, we may have to ask ourselves to what extent other IS theories also suffer from U.S.-centric parochialism (e.g., Boyacigiller & Adler, 1991) and have implicit cultural assumptions not applicable elsewhere. Before we can answer that question, researchers will experience uncertainty when using MTurk samples—particularly worldwide ones. Ironically, excluding non-U.S. workers from our samples could only perpetuate parochialism in our research. Regardless, we must be sensitive to cultural differences that may affect how MTurk workers interpret and answer questions (Smith et al., 2015).

Each sampling strategy brings its own risks and rewards along multiple dimensions in terms of both internal and external validity (Landers & Behrend, 2015). A sample that best answers a research question, regardless of its convenience, is the best sample (Vogt, Gardner & Haeffele, 2012). Despite the temptation to collect data quickly and inexpensively, we must understand the participants in these online populations and ensure we use them for both their value and representativeness, not merely their convenience (Rapp & Hill, 2015). In cases where it is sensible to use MTurk data, researchers must take steps to ensure data quality and report data-gathering and cleansing procedures in a detailed way to ensure future replication.

While we focus on the methodological and theoretical concerns related to gathering MTurk data in this paper, it involves significant ethical issues as well (e.g., low pay in an unregulated labor market). While some reports have begun to provide an initial look into the value that MTurk workers see in tasks and what academics can do to provide more ethical treatment for the workers (Deng et al., 2016; Marder & Fritz, 2015), we clearly need more research to better understand when and how crowdsourced data constitutes a viable sampling strategy for IS research and ensure the viability of this platform for future studies. We hope this paper stimulates discussion on this important issue and provides researchers a set of questions to ask themselves when planning their next study.

References

- Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods, 45*(2), 527-535.
- Baird, A., Miller, C. J., Raghu, T. S., & Sinha, R. K. (2016). Product line extension in consumer software markets in the presence of free alternatives. *Information Systems Research, 27*(2), 282-301.
- Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavioral Research Methods, 43*(3), 1-14.
- 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*, 351-368.
- Bohannon, J. (2011). Human subject research: Social science for pennies. *Science, 334*(6054), 307.
- Bohannon, J. (2016). Mechanical Turk upends social sciences: Growing pains arise for researchers using online platform. *Science, 352*(6291), 1263-1264.
- Boyacigiller, N. A., & Adler, N. J. (1991). The parochial dinosaur: Organizational science in a global context. *Academy of Management Review, 16*(2), 262-290.
- Brandon, D. M., Long, J. H., Loraas, T., Mueller-Phillips, J., & Vansant, B. (2014). Online instrument delivery and participant recruitment services: Emerging opportunities for behavioral accounting research. *Behavioral Research in Accounting, 26*(1), 1-23.
- Brown, S. A., & Venkatesh, V. (2005). A model of adoption of technology in the household: A baseline model test and extension incorporating household life cycle. *MIS Quarterly, 29*(3), 399-426.
- Buhrmester, M. D., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5.
- Burton-Jones, A. (2009). Minimizing method bias through programmatic research. *MIS Quarterly, 33*(3), 445-471.
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods, 46*(1), 112-130.
- Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology, 12*, 53-81.
- Chen, D. L., & Horton, J. J. (2016). Are online labor markets spot markets for tasks? A field experiment on the behavioral response to wage cuts. *Information Systems Research, 27*(2), 403-423.
- Chua, R. Y. J. (2013). The costs of ambient cultural disharmony: Indirect intercultural conflicts in social environment undermine creativity. *Academy of Management Journal, 56*(6), 1545-1577.
- Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science, 1*(2), 120-131
- Crowston, K. (2012). Amazon Mechanical Turk: A research tool for organizations and information systems scholars. In A. Bhattacharjee & B. Fitzgerald (Eds.), *IFIP advances in information and communication technology* (vol. 389, pp. 210–221). Berlin: Springer.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*, 982-1003.
- Deneme. (2009). *Deneme: A blog of experiments on Amazon Mechanical Turk*.
- Deng, X., & Joshi, K. D. (2016). Why individuals participate in micro-task crowdsourcing work environment: Revealing crowdworkers' perceptions. *Journal of the Association for Information Systems, 17*(10), 648-673.

- Deng, X., Joshi, K. D., & Galliers, R. D. (2016). The duality of empowerment and marginalization in Microtask crowdsourcing: Giving voice to the less powerful through value sensitive design. *MIS Quarterly*, 40(2), 279-302.
- Eriksson, K. & Simpson, B. (2010). Emotional reactions to losing explain gender differences in entering a risky lottery. *Judgment and Decision Making*, 5(3), 159-163.
- Elsevier. (n.d.). *Guide for authors*. Retrieved from <https://www.elsevier.com/journals/journal-of-vocational-behavior/0001%E2%80%93938791/guide-for-authors>
- Faraj, S., & Lee, S. (2000). Coordinating expertise in software development teams. *Management Science*, 46(12), 1554-1568.
- Farrell, A. M., Grenier, J. H., & Leiby, J. (2016). Scoundrels or stars? Theory and evidence on the quality of workers in online labor markets. *The Accounting Review*, 92(1), 93-114.
- Feitosa, J., Joseph, D., & Newman, D. (2015). Crowdsourcing and personality measurement equivalence: A warning about countries whose primary language is not English. *Personality and Individual Differences*, 75, 47-52.
- Fleischer, A., Mead, A. D., Huang, J. (2015). Inattentive responding in MTurk and other online samples. *Industrial and Organizational Psychology*, 8(2), 196-201.
- Fort, K., Adda, G., & Cohen, K. B. (2011). Amazon Mechanical Turk: Gold mine or coal mine? *Computational Linguistics*, 37(2), 413-420.
- Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, 26, 213-224.
- Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual Review of Psychology*, 66, 877-902.
- Gosling, S. D., Rentfrow, P. J., Swann, W. B., Jr. (2003). A very brief measure of the Big Five personality domains. *Journal of Research in Personality*, 37, 504-528.
- Guinan, P. J., Coopridge, J. G., & Faraj, S. (1998). Enabling software development team performance during requirements definition: A behavioral versus technical approach. *Information Systems Research*, 9(2), 101-125.
- Harms, P. D., & DeSimone, J. A. (2015). Caution! MTurk workers ahead—fines doubled. *Industrial and Organizational Psychology*, 8(2), 183-189.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33, 61-83.
- Hibbeln, M., Jenkins, J. L., Schneider, C., Valacich, J. S., & Weinmann, M. (2017). How is your user feeling? Inferring emotion through human-computer interaction devices. *MIS Quarterly*, 41(1), 1-A7.
- 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.
- Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business Psychology*, 27, 99-114.
- Hur, J. D., Koo, M., & Hofmann, W. (2015). When temptations come alive: How anthropomorphism undermines self-control. *Journal of Consumer Research*, 42(2), 340-358.
- Jenkins, J. L., Anderson, B. B., Vance, A., Kirwan, C. B., & Eargle, D. (2016). More harm than good? How messages that interrupt can make us vulnerable. *Information Systems Research*, 27(4), 880-896.
- Jia, R., & Reich, B. H. (2013). IT service climate, antecedents and IT service quality outcomes: Some initial evidence. *Journal of Strategic Information Systems*, 22(1), 51-69.
- Kane, G. C., & Ransbotham, S. (2016). Content and collaboration: An affiliation network approach to information quality in online peer production communities. *Information Systems Research*, 27(2), 424-439.

- Kehr, F., Kowatsch, T., Wentzel, D., & Fleisch, E. (2015). Blissfully ignorant: The effects of general privacy concerns, general institutional trust, and affect in the privacy calculus. *Information Systems Journal*, 25(6), 607-635.
- Kenny, D. A., & Judd, C. M. (1996). A general procedure for the estimation of interdependence. *Psychological Bulletin*, 119(1), 138-148.
- Kramer, J., Rubin, A., Coster, W., Helmuth, E., Hermos, J., Rosenbloom, D., Moed, R., Dooley M., Kao, Y.-C., Liljenquist, K., Brief, D., Enggasser, J., Keane, T., Roy, M., & Lachowicz, M. (2014). Strategies to address participant misrepresentation for eligibility in Web-based research. *International Journal of Methods in Psychiatric Research*, 23(1), 120-129.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50, 537-567.
- Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology*, 8(2), 142-164.
- Lease, M., Hullman, J., Bigham, J. P., Bernstein, M. S., Kim, J., Lasecki, W., Bakhshi, S., Mitra, T., & Miller, R. C. (2013). *Mechanical Turk is not anonymous* (SSRN Scholarly Paper ID 2228728). Rochester, NY: Social Science Research Network.
- 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. *Behavioral Research Methods*, 47, 519-528.
- Lowry, P. B., & Wilson, D. (2016). Creating agile organizations through IT: The influence of internal IT service perceptions on IT service quality and IT agility. *Journal of Strategic Information Systems*, 25(3), 211-226.
- Lowry, P. B., Zhang, J., Wang, C., & Siponen, M. (2016). Why do adults engage in cyberbullying on social media? An integration of online disinhibition and deindividuation effects with the social structure and social learning model. *Information Systems Research*, 27(4), 962-986.
- Mamonov, S., & Koufaris, M. (2014). The impact of perceived privacy breach on smartphone user attitudes and intention to terminate the relationship with the mobile carrier. *Communications of the Association for Information Systems*, 34, 1157-1174.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, 44(1), 1-23.
- McGonagle, A. K. (2015). Participant motivation: A critical consideration. *Industrial and Organizational Psychology*, 8(2), 208-213.
- Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3), 263-280.
- O'Leary, M. B., Wilson, J. M., & Metiu, A. (2014). Beyond being there: The symbolic role of communication and identification in perceptions of proximity to geographically dispersed colleagues. *MIS Quarterly*, 38(4), 1219-1243.
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63, 539-569.
- Marder, J., & Fritz, M. (2015). The Internet's hidden science factory. *PBS NewsHour*. Retrieved from <http://www.pbs.org/newshour/updates/inside-amazons-hidden-science-factory/>
- Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5, 411-419.
- Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavioral Research Methods*, 46(4), 1023-1031.
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 65, 539-569.
- Rapp, J. M., & Hill, R. P. (2015). Lordy, lordy, look who's 40! The *Journal of Consumer Research* reaches a milestone. *Journal of Consumer Research*, 42(1), 19-29.

- Rand, D. G. (2012). The promise of Mechanical Turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, 299, 172-179.
- Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, 5.
- Reis, H. T. & Gosling, S. D. (2010). Social psychological methods outside the laboratory. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., vol. 1). Hoboken, NJ: Wiley Hoboken.
- Reich, B. H., & Benbasat, I. (2000). Factors that influence the social dimension of alignment between business and information technology objectives. *MIS Quarterly*, 24(1), 81-111.
- Ross, J., Irani, I., Silberman, M. S., Zaldivar, A., & Tomlinson, B. (2010). Who are the crowdworkers? Shifting demographics in Amazon Mechanical Turk. In K. Edwards & T. Rodden (Eds.), *Proceedings of the ACM Conference on Human Factors in Computing Systems* (pp. 2863-2872). New York, NY: ACM.
- Siegel, J. T., Navarro, M. A., & Thomson, A. L. (2015). The impact of overtly listing eligibility requirements on MTurk: An investigation involving organ donation, recruitment scripts, and feelings of elevation. *Social Science & Medicine*, 142, 256-260.
- Schmidt, G. B. (2015). Fifty days as an MTurk worker: The social and motivational context for Amazon Mechanical Turk workers. *Industrial & Organizational Psychology*, 8(2), 165-171.
- Smith, N. A., Sabat, I. E., Martinez, L. R., Weaver, K., & Xu, S. (2015). A convenient solution: Using MTurk to sample from hard-to-reach populations. *Industrial & Organizational Psychology*, 8(2), 220-228.
- Soror, A. A., Hammer, B. I., Steelman, Z. R., Davis, F. D., & Limayem, M. M. (2015). Good habits gone bad: Explaining negative consequences associated with the use of mobile phones from a dual-systems perspective. *Information Systems Journal*, 25(4), 403-427.
- Steelman, Z. R., Hammer, B. I., & Limayem, M. (2014). Data collection in the digital age: Innovative alternatives to student samples. *MIS Quarterly*, 38(2), 355-A20.
- Sturgis, P., Allum, N., & Brunton-Smith, I. (2009). Attitudes over time: The psychology of panel conditioning. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 113-126). Hoboken, NJ: Wiley.
- Sykes, T. A., Venkatesh, V., & Gosain, S. (2009). Model of acceptance with peer support: A social network perspective to understand employees' system use. *MIS Quarterly*, 33(2), 371-393.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425-478.
- Venkatesh, V., & Sykes, T. A. (2013). Digital divide initiative success in developing countries: A longitudinal field study in a village in India. *Information Systems Research*, 24(2), 239-260.
- Vogt, P. W., Gardner, D. C., & Haeffele, L. M. (2012). *When to use what research design*. New York, NY: Guilford Press.
- Winship, C., & Mare, R. D. (1992). Models for sample selection bias. *Annual Review of Sociology*, 18, 327-350.
- Woo, S. E., Keith, M., & Thornton, M. A. (2015). Amazon Mechanical Turk for industrial and organizational psychology: Advantages, challenges, and practical recommendations. *Industrial and Organizational Psychology*, 8(2), 171-179.
- Yates, D. S., Moore, D. S., & Starnes, D. S. (2008). *The practice of statistics* (3rd ed.). New York: Freeman.
- Zhu, X., Barnes-Farrell, J. L., & Dalal, D. K. (2015). Stop apologizing for your samples, start embracing them. *Industrial and Organizational Psychology*, 8(2), 228-232.

About the Authors

Ronnie Jia is an Associate Professor of Information Systems at Illinois State University. His research interests include intrinsic interests in IT, negative impacts of technology use, and IT service management. He has published in *Journal of Association for Information Systems*, *Journal of Strategic Information Systems*, *Communications of the Association for Information Systems*, *Computers in Human Behavior*, and a number of other journal and conference outlets.

Zachary Steelman is an Assistant Professor of Information Systems in the Walton College of Business at the University of Arkansas. He has authored refereed publications in prominent IS journals and conferences such as *MIS Quarterly*, *MIS Quarterly Executive*, *Information Systems Journal*, *Communications of the Association of Information Systems*, *Computers in Human Behavior*, *Americas Conference on Information Systems*, and the *Hawaii International Conference on System Sciences*. His research interests include the selection, development, and management of IT portfolios for individuals, teams, and organizations. Additionally, he has examined the “dark side of IT” focusing on the impacts, both positive and negative, on individuals.

Blaize Horner Reich is the RBC Professor of Technology and Innovation at the Beedie School of Business, Simon Fraser University, Canada. She has authored many publications in top IS journals (including *MIS Quarterly*, *Journal of Information Technology*, *Information Systems Research*, the *Journal* and the *Communications of the Association of Information Systems*, *Journal of Management Information Systems*) as well as leading journals in the project management domain. Her research interests include governance of the intersection of IT and Business, and agile approaches to projects and management.

Copyright © 2017 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from publications@aisnet.org.