Diverse but not representative: On the occupational and demographic diversity of the labor force that is accessible via MTurk

Christopher M. Castille[1], Kathryn Daigle[1], Sarah Avet[1], Rachel Williamson Smith[2], James DeLeon[3], Bodour H. Mahmoud[4], & John E. Buckner V[5]

[1] Nicholls State University

[2] Louisiana State University

[3] APTMetrics

[4] Villanova University

[5] AlixPartners

Author note

We have sought to make this project fully reproducible. All project materials can be found on the Open Science Framework (https://osf.io/r7mjt/), where elements of this study were pre-registered (https://osf.io/fdgew/). A previous version of this manuscript was presented at the 2017 annual conference for the Society for Industrial & Organizational Psychology, Inc., in Orlando, FL. It should also be noted that data collection for one of the 2015 samples was funded by the Graduate School Dean's Award Program at the University of Georgia.

Correspondence concerning this article should be addressed to Christopher M. Castille, 906 E. 1st St., Thibodaux, LA, 70301. E-mail: chris_castille@mac.com

Abstract

Scholars have relied upon Amazon Mechanical Turk (MTurk) to sample from the broader labor force. While doing so allows scholars to sample from a variety of occupational settings, it remains unclear whether occupational settings sampled from MTurk differ from the broader labor force. To address this problem, we compare the occupational and demographic diversity of three conveniently gathered MTurk samples (two in 2015, one in 2017) against the broader labor force from which they have been indirectly sampled in the United States. We consistently found differences suggesting that the MTurk population is overrepresented by certain white-collar professions (e.g., Management, Business, and Financial; Computer, Engineering, and Science; Education, Legal, Community Service, Arts, and Media) and underrepresented by blue-collar professions (e.g., Production; Transportation and Material Moving). Other demographic trends mirrored broader demographic trends identified with prior MTurk research with regard to age (younger), gender (female-biased), ethnicity (e.g., African Americans underrepresented), and education (generally overeducated). Overall, we contribute to the published literature by indicating that the labor force that is accessible via MTurk does not adequately represent the broader labor force.

Keywords: MTurk, Occupational Diversity, Demographics

Word count: 3936

Diverse but not representative: On the occupational and demographic diversity of the labor force that is accessible via MTurk

Introduction

Amazon's Mechanical Turk (MTurk), an online labor market created by Amazon, has been popular among social scientists interested in collecting inexpensive, high-quality data from a diverse population for well (Aguinis & Lawal, 2012; Ahler, Roush, & Sood, 2018; Behrend, Sharek, Meade, & Wiebe, 2011; Bergman & Jean, 2016; Kennedy, Clifford, Burleigh, Jewell, & Waggoner, 2018; Landers & Behrend, 2015; Paolacci & Chandler, 2014). The platform allows for easier subject pool access at an affordable cost and for subject prescreening (Mason & Suri, 2012), subject anonymity or identifiability (which is important for longitudinal studies), and offers a simple and supportive infrastructure (Paolacci, Chandler, & Ipeirotis, 2010). Leveraging the strengths of this tool, MTurk has been used to replicate a variety of well-known phenomena in the social sciences (e.g., reaction times, priming, task switching, see Crump, McDonnell, & Gureckis, 2013; memory, see Simons & Chabris, 2012).

Scholars who study organizational phenomena (e.g., organizational behaviorists, industrial and organizational psychologists, management scholars) have also adopted MTurk. Keith, Tay, and Harms (2017) recently reported that published organizational science studies utilizing MTurk increased by 800% between 2012 and 2015. When comparing crowdsourcing tools like MTurk with traditional methods of studying organizational phenomena (e.g., gaining consent from both institutional review boards and local organizations; see Ilgen & Bell, 2001), several benefits are particularly salient, such as rapidly gaining access to multiple organizational settings and enhanced external validity. Such benefits can greatly reduce the cost of studying organizational phenomena, leveling the research playing field (Goodman, Cryder, & Cheema, 2013).

As enthusiasm surrounding the use of MTurk in the organizational sciences has grown, so have scholars' concerns over its use (e.g., Cheung, Burns, Sinclair, & Sliter, 2017; Keith et al., 2017; Landers & Behrend, 2015; Woo, Keith, & Thornton, 2015). One problematic area relates to an assumption underlying research into organizational phenomena that draws on MTurk; namely, that the MTurk workers themselves represent a broad cross-section of the labor force population. This assumption is key to a commonly cited benefit of using crowdsourcing tools like MTurk; that is, increased generalizability (e.g., Aguinis & Lawal, 2012; Behrend et al., 2011; Highhouse & Zhang, 2015; Keith et al., 2017; Roulin, 2015). In other words, organizational scholars may assume that MTurk can be used to indirectly and *adequately* sample from the broader working population. Unfortunately, little research has compared MTurk worker demographics to that of the broader labor force. If MTurk does not adequately reflect the broader labor force, then the generalizability of research drawing on MTurk will have its own set of constraints (Keith et al., 2017; Landers & Behrend, 2015).

Here, we compare the working subpopulation of the U.S. MTurk workforce to the broader labor force from which they were indirectly sampled. Given the importance of context in organizational research (Johns, 2006; Judge & LePine, 2007; Newman, Joseph, & Feitosa, 2015), we focus specifically on the extent to which the labor force that is sampled via MTurk adequately represents the broader labor force. Our study, which fits into the broader body of literature regarding who we study and the generalizability of claims made in the social sciences (e.g., Paolacci et al., 2010; Henrich, Heine, & Norenzayan, 2010), contributes to the wealth of knowledge that exists on the demographics of the MTurk population by clearly pointing out which occupations, if any, have been overrepresented or underrepresented. It also informs future research on the study of organizational phenomena via MTurk by calling researchers using MTurk to take the role of occupational context seriously when attempting to generalize across occuaptional settings.

Overview of the Literature

Over half of the MTurk population is employed (Behrend et al., 2011; Keith et al., 2017; Levay, Freese, & Druckman, 2016). If the total MTurk population contains over 100,000 people (Difallah, Filatova, & Ipeirotis, 2018), there is a sizable subpopulation of individuals whom we may invite to participate in our research. Even if only a few thousand individuals can be sampled from (see Stewart et al., 2015), this can still be helpful for organizational scholarship (e.g., measurement refinement). It might even be useful for examining understudied phenomena or populations of workers (e.g., job search behavior, underemployment, unemployed; see Roulin, 2015).

By using MTurk to sample workers from a wide range of occupational settings, it has widely been believed that scholars' ability to generate inferences with greater external validity from sample data has increased (Aguinis & Lawal, 2012; Highhouse & Zhang, 2015; Smith, Sabat, Martinez, Weaver, & Xu, 2015). MTurk's diversity on variables such as age, ethnicity, education, employment status, job experience, profession, nationality, and personality (Behrend et al., 2011) is commonly cited as a strength that sets it apart from other commonly studied populations (e.g., university students, community samples), making it an attractive subject pool for organizational scholars to draw upon (Buhrmester, Kwang, & Gosling, 2011; Eriksson & Simpson, 2012; Paolacci et al., 2010).

While several studies have examined the attributes of the MTurk population (e.g., demographic, occupational, psychographic, etc.; see Behrend et al., 2011; Michel, O'Neill, Hartman, & Lorys, 2018), only one has compared the occupational makeup of MTurk (i.e., Michel et al., 2018). By comparing the occupational makeup (or occupational heterogeneity or diversity) of a sample of U.S. MTurk respondents to the labor force, Michel et al. (2018) inferred that MTurk workers are broadly sampled from various occupational areas. For instance, they noted that the largest MTurk subpopulation was from the "Education and Training" career cluster, comprising approximately 13% of responses. They point out that a majority of the career clusters were within a 5% difference range when compared with corresponding career clusters as defined by the Bureau of Labor Statistics. They also point out that the most undersampled population was "Business Management and Administration."

Michel et al.'s (2018) study is important for prompting conversation about how the occupational diversity of MTurk compares to the broader labor force. However, while informative, Michel et al.'s (2018) study has notable shortcomings that we sought to overcome. First, Michel et al. sought to determine if MTurk workers are broadly distributed across the labor market. While they did indeed answer this question, a statistical test would challenge a key assumption that might be adopted (implicitly or explicitly) by organizational researchers who use MTurk; namely, that the labor force accessed through MTurk is comparable to the broader labor force. Examining this assumption statistically is important because if we cannot assume that MTurk gives us a comparable cross section of the labor force, then our conclusions will be biased in favor of over- or under-sampled sampled areas. It is also worth pointing that that, since we have data spread across multiple years (2015 and 2017), we have an opportunity to examine the pervasiveness of trends in occupational and demographic makeup of the workforce that is accessed via MTurk. If MTurk has a particular and pervasive bias in favor of certain occupational areas, then increasing our reliance on MTurk might will not bring about the purported benefits. Rather, such a growing reliance will prevent us from overcoming concerns that our knowledge base applies only to certain kinds of workers (e.g., core, salaried, managerial; Bergman & Jean, 2016).

Second, Michel et al. relied on labor force descriptions from 2013. If we assume that Michel et al.'s data were also gathered in the same year, then - given that roughly half of the workers leave MTurk within each year (Difallah et al., 2018; Stewart et al., 2015) - Michel et al.'s conclusions may no longer accurately describe the labor force that is accessible via MTurk. Having some idea regarding the composition of the labor force that is accessible via MTurk can inform sampling practices. For instance, if a researcher hopes to examine sales professionals, then knowing whether or not this population is historically underrepresented by MTurk can be useful for deciding whether or not to sample workers from MTurk.

In sum, though prior work suggests that the labor force that is accessible via MTurk is quite diverse in terms of occupations, our work expands the literature by (1) clarifying the extent to which this heterogeneity aligns with the broader and indirectly sampled labor force and (2) showing the pervasiveness of these trends. If the occupational diversity of the broader U.S. population is indeed reflected by MTurk, then findings from MTurk's labor force should better generalize across occupations. Alternatively, if MTurk's occupational diversity consistently over- or under-represents certain occupations, then the generalizability of findings from MTurk may be more limited than has been appreciated. Therefore, we begin our investigation by clarifying our focal research question.

Research Question 1: Does the occupational diversity of the labor force that is accessible via MTurk adequately represent the broader U.S. labor force?

In addition to examining the occupational makeup of the labor force that is accessible via MTurk, we examine other demographic attributes, focusing in on age, gender, ethnicity, and education, as these are commonly studied in the organizational sciences (e.g., Berry, Ones, & Sackett, 2007; Collins & Gleaves, 1998; Dalal, 2005). While there is much prior research on MTurk, it describes broad trends associated with the MTurk population in general, which may or may not generalize to the subpopulation that is also part of the labor force. For the U.S., the general MTurk population seems to be younger and slightly female-biased (Difallah et al., 2018). Compared to college students and community samples, it is more diverse in terms of education (Behrend et al., 2011; Goodman et al., 2013). In comparing MTurk samples to nationally representative samples (e.g., U.S. Census, Survey USA), Keith et al. (2017) also note that MTurk samples are younger, more educated. Knowing the extent to which the labor force that is accessible via MTurk mirrors the broader labor force, demographically speaking, would also be informative for the same reasons noted for RQ1. Given that no prior work has conducted this comparison, we sought to answer a second research question.

Research Question 2: Does the demographic diversity (age, gender, ethnicity, and education) of the labor force that is accessible via MTurk adequately represent of the broader U.S. labor force?

Methods

U.S. Population Occupational Information

Occupational Employment Statistics were obtained from the BLS (BLS) both from years 2015 and 2017 to correspond with the MTurk samples, which were also gathered from the same years. These years for chosen because those are the years for which we have relevant data on MTurk. The BLS data are released annually in the spring and describe the previous year.

Statistics from these reports allowed us to define expectations regarding the proportion of workers falling into a particular occupational category. For the sake of consistency, we adopted the Intermediate Aggregation and Standard Occupational Classification (SOC) used by the BLS for this study.

MTurk Sample Information

In 2015, participants' data were collected from two independently conducted studies (n_1 = 727; n_2 = 1,046) for purposes that were not directly related to the purpose of this study (i.e., investigating psychological phenomena in organizational life). Data collected in 2017 (n_3 = 616) can be similarly described.[1] For each sample, data collection was restricted to employed individuals in the United States. Small monetary incentives were used ($.75-$1.30) to encourage participation. Following guidance from the literature, we used worker reputation to pre-screen individuals prior to granting them access to our surveys (e.g., approval ratings greater than 90%; Peer, Vosgerau, & Acquisti, 2014). For all samples, participants self-reported their job titles. For two samples (n_1 = 727; n_2 = 1,046), the six-digit standard occupational classification (SOC) codes were obtained by utilizing trained research assistants, who matched the self-reported job titles to those in the ONET database. In a third sample, (n_3 = 616) participants were then brought to the Occupation Network Online (ONET) website (www.onetonline.org) and asked to retrieve the six-

---

[1] Our samples are what remained after filtering out inattentive responders (see Meade & Craig, 2012), individuals who were outside of the target population (that is, the labor force that is accessible via MTurk), completing MTurk work while located outside of the U.S. (and therefore being of questionable origin), and for using a questionable IP address (Kennedy et al., 2018).

digit occupational classification code that most closely corresponded to their job title.[2] Only individuals who could be classified using the Intermediate Aggregation were included in our analyses. Using this data, we described the occupational and demographic diversity of each MTurk sample using proportions of category representation. We also estimated 95% confidence intervals to allow ease of plotting and analysis.

Analytical Approach

A chi-square test of goodness-of-fit was used to determine whether proportions observed from MTurk samples deviated from expected proportions as derived from the BLS. Specifically, we tested whether MTurk's occupational diversity deviated significantly from that expected if we assume the BLS population parameters generalize to the MTurk population. Gender, ethnicity, and education were examined in the same fashion. Given that multiple occupational and ethnic groups were examined, significant tests were followed by a post-hoc test. More specifically, we examined the standardized residuals (std. res.) from our chi-square tests (see Agresti, 2007), which in this context defines occupations that were overrepresented (values ≥ 3), underrepresented (values ≤ 3), or consistent with expectations ($\leq |2.99|$). To facilitate interpretation, we also calculated 95% confidence intervals for our sample proportions and plotted these values (see Figure 1 for 2015 and Figure 2 for 2017). Age differences were

---

[2] To conserve resources, future researcher wishing to replicate our results should consider using the "ONETr" package for this purpose.

examined used a one-sample t-test with BLS data defining the average age of the labor force (the sample standard deviations were used to estimate the population standard deviation).[3]

Results

Occupational Diversity of the Broader Labor Force Accessible via MTurk in 2015

All relevant statistics are described by Tables 1 through 4. Alternatively, Figures 1 (2015) and 2 (2017) plot these occupational diversity for ease of analysis. Chi-square testing indicated that MTurk's occupational diversity does not consistently align with expectations in both sample 1 ($\chi^2(12) = 333.86$, $p < 0.001$) and sample 2, ($\chi^2(12)= 464.62$, $p < 0.001$). Therefore, both samples deviate from expectations. Across both samples, overrepresented occupations included (i) 11-13 Management, Business, and Financial (std. res.: study 1 = 6.96; study 2 = 14.47), (ii) 15-19 Computer, Engineering, and Science (std. res.: study 1 = 10.74; study 2 = 8.41), and (iii) 21-27 Education, Legal, Community Service, Arts, and Media (std. res.: study 1 = 8.08; study 2 = 7.18). Occupations that were consistently under-represented included (i) 31-39 Service (std. res.: study 1 = -5.94; study 2 = -5.22), (ii) 47 Construction and Extraction (std. res.: study 1 = -3.52; study 2 = -4.64), (iii) 49 Installation, Maintenance, and Repair (std. res.: study 1 = -4.28; study 2 = -6.10), (iv) 51 Production (std. res.: study 1 = -5.60; study 2 = -5.22), and (v) 53 Transportation and Material Moving (std. res.: study 1 = -5.52; study 2 = -6.99). Lastly, occupations that consistently appeared in a proportion that conformed with expectations included (i) 41 Sales and

[3] Analysis of the occupational diversity of MTurk was conducted on the 2015 samples and were not pre-registered. However, analysis of the occupational diversity of the 2017 sample were pre-registered on the Open Science Framework. It is worth pointing out that post-hoc testing was not pre-registered nor were the analyses regarding gender, ethnicity, education, and age.

Related (std. res.: study 1 = -2.14; study 2 = -1.40), (ii) 43 Office and Administrative Support (std. res.: study 1 = 1.11; study 2 = 0.47), (iii) 45 Farming, Fishing, and Forestry (std. res.: study 1 = -0.12; study 2 = -0.03), and (iv) 55 Military Specific (std. res.: study 1 = -1.22; study 2 = -1.63). For 29 Healthcare Practitioners and Technical, sample 1 suggests that these occupations are represented as expected (std. res.: study 1 = 0.61). However, this group was under-represented in sample 2 (std. res.: study 2 = 7.18).

Insert Table 1, Table 2, and Figure 1 about here

Demographic Diversity of the Broader Labor Force Accessible via MTurk in 2015

The average worker in 2015 was 42.30 years of age, which was significantly older than the average MTurk worker across both samples (n_1: M = 34.39, SD = 10.52, $t(785) = -21.07$, $p < 0.001$, $d = -0.75$; n_2: M = 35.55, SD = 10.52, $t(1277) = -21.07$, $p < 0.001$, $d = -0.75$).

With regard to gender, the BLS suggests a male-biased labor force (Male = 53.0%; Female = 47.0%). The opposite was the case for both MTurk samples (n_1: Male = 43.0%; Female = 57.0%; n_2: Male = 37.0%; Female = 63.0%). These differences were statistically significant in both cases (n_1: $\chi^2(1) = 34.43$, $p < 0.001$; n_2: $\chi^2(1) = 129.88$, $p < 0.001$), which suggests that MTurk is overrepresented by females (std. res.: n_1 = 5.87; n_2 = 11.40).

With regard to ethnicity, we examined only the published proportions for Caucasians, African Americans, and Asians.[4] BLS summary statistics suggest that the labor force was composed of approximately 79.2% Caucasians, 11.7% African Americans, and 5.80% Asians. This distribution deviated significantly from both of our MTurk samples (n_1: $\chi^2(3) = 136.23$, $p < 0.001$; n_2: $\chi^2(3) = 335.99$, $p < 0.001$). Post-hoc testing revealed that African Americans ($n_1 = 5.00\%$; n_2 4.00%) were consistently under-represented (std. res.: $n_1 = -5.85$; $n_2 = -8.49$). Additionally, the "other" category ($n_1 = 9.00\%$; n_2 11.0%) was consistently overrepresented by MTurk (std. res.: $n_1 = 10.33$; $n_2 = 16.62$). Caucasians ($n_1 = 81.0\%$, std. res.: $n_1 = 1.23$; n_2 80.0%; std. res. = 1.04) and Asians ($n_1 = 5.00\%$, std. res.: $n_1 = -1.53$; n_2 5.00%; std. res. = -2.11) were represented in a manner that was consistent with expectations.

With regard to education, due to a scaling issue with sample 1, we were only able to examine sample 2. Furthermore, to be consistent with the BLS, only a subset of individuals of age 25 or older were analyzed. Nevertheless, we detected a significant overall deviation from expectations (n_2: $\chi^2(4) = 262.66$, $p < 0.001$). Specifically, both the "no high school" (MTurk = 0%; BLS = 8.00%; std. res. = -9.06) and the "high school diploma/GED" groups (MTurk = 10.0%; BLS = 26.0%; std. res. = -11.51) were under-represented in the MTurk sample. Two education levels were overrepresented: specifically, individuals with some college (MTurk = 24.0%, BLS = 17.0%, std. res. = 5.60) and individuals with a bachelors degree or higher (MTurk = 53.0%; BLS = 38.0%, std. res. = 9.77). Individuals holding a 2 year degree emerged in a manner that was consistent with expectations (MTurk = 13.0%, BLS = 11.0%, std. res. = 2.12).

---

[4] An "other" category was created as the remainder of these three categories, which contained Hispanics and other ethnic groups.

Occupational Diversity of the Broader Labor Force Accessible via MTurk in 2017

Many of our findings regarding the occupational diversity of the 2017 sample mirrored those from 2015. Chi-square testing indicated that MTurk's occupational diversity does not consistently align with expectations ($\chi^2(12) = 268.15$, $p < 0.001$). Overrepresented occupations included (i) 11-13 Management, Business, and Financial (std. res. = 7.76), (ii) 15-19 Computer, Engineering, and Science (std. res. = 9.03), and (iii) 21-27 Education, Legal, Community Service, Arts, and Media (std. res. = 5.86). Occupations that were underrepresented included (i) 31-39 Service (std. res. = -5.53), (ii) 47 Construction and Extraction (std. res. = -4.66), (iii) 49 Installation, Maintenance, and Repair, (iv) 51 Production (std. res. = -4.28), and (v) 53 Transportation and Material Moving (std. res. = -5.23). Lastly, occupations that appeared in a proportion that conformed with expectations included (i) 29 Healthcare Practitioners and Technical (std. res. = -2.11), (ii) 41 Sales and Related (std. res. = -1.77), (iii) 43 Office and Administrative Support (std. res. = 2.21), (iv) 45 Farming, Fishing, and Forestry (std. res. = 1.59), and (v) 55 Military Specific (std. res. = -1.32).

Insert Table 3, Table 4, and Figure 2 about here

Demographic Diversity of the Broader Labor Force Accessible via MTurk in 2017

Many of our findings regarding the demographic diversity of the 2017 sample (n_3) mirrored those from 2015. The average worker in 2017 was 42.20 years of age, which was significantly older than the average MTurk worker (M = 35.91, SD = 11, $t(666) = -14.77$, $p < 0.001$, $d = -0.57$).

With regard to gender, the BLS suggests a male-biased labor force (Male = 53.0%; Female = 47.0%). The opposite was the case for the MTurk sample (Male = 34.0%; Female = 34.0%). This difference was statistically significant in ($\chi^2(1) = 97.86$, $p < 0.001$), suggesting that MTurk is overrepresented by females (std. res. = 9.89).

With regard to ethnicity, we examined only the published proportions for Caucasians, African Americans, and Asians. BLS summary statistics suggest that the labor force was composed of approximately 78.4% Caucasians, 12.1% African Americans, and 6.16% Asians. This distribution deviated significantly from our our MTurk sample ($\chi^2(3) = 80.71$, $p < 0.001$). Post-hoc testing revealed that African Americans (MTurk = 7.00%) were significantly underrepresented (std. res. = -4.24). Additionally, the "other" category (MTurk = 9.00%) was overrepresented by MTurk (std. res. = 8.02). Caucasians (MTurk: 80.0%; std. res.: 0.85) and Asians (MTurk: ; std. res.: -1.66) were represented in a manner that was consistent with expectations.

Discussion

Consistent with the broader literature on MTurk demographics (Behrend et al., 2011; Berinsky, Huber, & Lenz, 2012; Difallah et al., 2018; Keith et al., 2017; Levay et al., 2016), we identified consistent trends in the occupational makeup of MTurk, which suggest that the labor force available for study via MTurk does not mirror the broader labor force. Rather, the labor force that is accessible via MTurk contains a disproportionate share of certain white-collar occupations and a dearth of blue-collar occupations. Specifically, MTurk workers were consistently more likely to come from occupational areas related to (i) Management, Business, and Financial, (ii) Computer, Engineering, and Science, and (iii) Education, Legal, Community Service, Arts, and Media occupations. They were also less likely to come from occupational areas related to (i) Service, (ii) Construction and Extraction, (iii) Installation, Maintenance, and Repair, (iv) Production, and (v) Transportation and Material Moving. The remaining occupations were consistent with expectations. One exception applies Healthcare Practitioners and Technical occupations, where one 2015 sample suggested that these occupations were underrepresented. However, the confidence intervals for the two 2015 samples overlapped, and so we suspect that this was a sampling error. Demographically, the labor force that is available through MTurk is younger, female-biased, less likely to contain African Americans, overrepresented by other ethnic groups (likely Hispanics), and more likely to contain highly educated members of the labor force (these trends mirror prior research on the MTurk population as a whole; see Keith et al., 2017). Nevertheless, our study thus contributes to the literature on the demographics of the MTurk population and on the generalizability of social science research that relies on MTurk (e.g., Paolacci et al., 2010; Henrich et al., 2010).

Given recent claims that the organizational science literature holds greater relevance for certain kinds of workers (e.g., core, salaried, managerial; Bergman & Jean, 2016) and that scholars have increasingly relied upon MTurk for reasons that include enhancing generalizability (Keith et al., 2017; Landers & Behrend, 2015), our findings regarding the occupational diversity of MTurk appear troubling. Overall, these findings suggest that, by increasingly relying upon MTurk, we may simply be more easily sampling from a population that we have historically overstudied (e.g., core, salaried, and managerial workers; Bergman & Jean, 2016). For instance, the most consistently sampled group of individuals worked in Management, Business, and Financial occupations, which are groups that we have long studied (Bergman & Jean, 2016). If so, then organizational scholars may be overstating conclusions generated using MTurk samples, particularly when the occupational characteristics matter (as they often do; see Johns, 2006; Newman et al., 2015). Therefore, we hope to caution future researchers against assuming that MTurk samples adequately sample from the broader labor force. For researchers using MTurk in their research, we have the following recommendations, thus contributing to the behavioral research methods literature on the proper use of MTurk (e.g., Cheung et al., 2017; Keith et al., 2017; Landers & Behrend, 2015)

Recommendations for Organizational Scholars Using MTurk for Research Purposes

Despite the concerns we have raised regarding the occupational diversity of the labor force that is accessible through MTurk, it is nevertheless striking how the distribution mirrors that of the broader labor force (albeit, imperfectly). Though MTurk overrepresents certain white-collar occupations and underrepresents blue collar occupations, the overall pattern does mirror the broader labor force. While this does echo the claim made by Michel et al. (2018), namely that the labor force that is accessible via MTurk is diverse, if researchers hope to generalize across the labor force while using MTurk, then other methodological choices must be made (e.g., stratified sampling, respondent weighting; for examples, see Simons & Chabris, 2012; Levay et al., 2016).

While scholars studying organizational phenomena may consider pre-screening for employment, this action could backfire. A recent observation has been made suggesting that roughly 20% of participants will lie to gain access to a study (Chandler & Paolacci, n.d.). It does not appear as though we encountered this problem because we informed our participants that, while we were interested in studying members of the labor force via MTurk, we would still allow non-labor force members to participate (they were given access to generic individual difference measures). We do not have experimental evidence demonstrating that this messaging had the desired effect, however. Future researchers drawing on MTurk may wish to test this. In the interim, scholars drawing on MTurk might consider adopting our messaging strategy to sample labor force participants through MTurk.

Another way for scholars studying organizational phenomena is to leverage what diversity the MTurk population does have to, for instance, form more precise predictions of organizational phenomena. Consider that, in testing a unified model of person-situation interactionism, Judge and Zapata (2015) predicted both the extent to which personality traits as a whole relate to performance (e.g., traits are more predictive in weak work situations, such as those granting decision making authority) and how strongly certain traits would relate to performance (e.g., conscientiousness is more predictive for occupations that require independence and strong attention-to-detail requirements). While they did not rely on MTurk (Judge and Zapata drew upon available data from the BLS as well as published meta-analyses), one could adapt their approach for studying the MTurk population by using multi-level modeling techniques (Kozlowski & Klein, 2000; Newman et al., 2015). By joining BLS data with data collected from MTurk, researchers might enhance their predictions of workplace outcomes, such as counterproductive behavior and citizenship behavior (Judge & LePine, 2007; Kish-Gephart, Harrison, & Treviño, 2010; O'Boyle, Forsyth, Banks, & McDaniel, 2012). Doing so can facilitate theory testing in ways that may not have been considered previously (e.g., comparing objective vs. subjective situations; see Tett & Burnett, 2003).

Limitations and Future Research

As all our samples were obtained via convenient sampling procedures, it remains uncertain the extent to which these samples adequately reflect the larger MTurk population. There are a variety of reasons why this could be the case (e.g., participants' use of message boards to identify our study; see Keith et al., 2017). Nevertheless, the consistency of our findings across our samples should be considered by organizational scholars hoping to draw on MTurk for their research with the hopes of generalizing to the larger labor force population. Additionally, as our studies were conducted in the U.S., it remains unclear how other MTurk populations might reflect (however imperfectly) their respective broader labor force. Future larger scale studies should also consider lower levels of BLS occupational classification to examine the extent to which narrower occupations might not be represented by MTurk.

Given that roughly half of MTurk workers might be Turking while working (Offermann, Coats, & Rauch, 2015), future research should examine the occupational features that motivate such behavior. For instance, in responding to the question “Does your job make a meaningful contribution to the world?”, roughly 40% of individuals answer no (see Graeber, 2018). Examining the occupational characteristics (e.g., job characteristics) that relate to questions via MTurk would be interesting for cultivating a deeper understanding about the nature of work for the labor force participants who are available via MTurk.

Given that our data were gathered and analyzed in 2015 and 2017, it might be insightful to have our procedures replicated annually to examine how the demographics of MTurk, specifically occupational characteristics, change over time, especially as access to technology becomes even easier and more commonplace. Thus, we encourage researchers to continue examining the occupational diversity of MTurk relative to the U.S. labor force. We also encourage researchers to consider novel ways of approximating the occupational diversity of both the U.S. labor force and MTurk as the currently existing taxonomies that we have used (i.e., BLS-based) probably do not capture all occupations. Lastly, our results raise a new and interesting research question: Why might MTurk possess the occupational diversity makeup that it does (i.e., why are some occupations overrepresented while others are under-represented)?

Similarly, what factors (e.g., access to technology, interest in psychology) might make it easier for workers coming from certain occupations to participate in OB studies hosted on MTurk? Answering these questions might unearth more constraints that impinge on the findings of occupational studies using MTurk.

References

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Wiley-Interscience.

Aguinis, H., & Lawal, S. O. (2012). Conducting field experiments using eLancing's natural environment. *Journal of Business Venturing*, *27*(4), 493–505. doi:10.1016/j.jbusvent.2012.01.002

Ahler, D. J., Roush, C. E., & Sood, G. (2018, November). *The Micro-Task Market for "Lemons": Collecting Data on Amazon's Mechanical Turk.*

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods*, *43*(3), 800. doi:10.3758/s13428-011-0081-0

Bergman, M. E., & Jean, V. A. (2016). Where Have All the "Workers" Gone? A Critical Analysis of the Unrepresentativeness of Our Samples Relative to the Labor Market in the IndustrialOrganizational Psychology Literature. *Industrial and Organizational Psychology*, *9*(1), 84–113. doi:10.1017/iop.2015.70

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*(03), 351–368. doi:10.1093/pan/mpr057

Berry, C. M., Ones, D. S., & Sackett, P. R. (2007). Interpersonal deviance, organizational deviance, and their common correlates: A review and meta-analysis. *Journal of Applied Psychology*, *92*(2), 410–424. doi:10.1037/0021-9010.92.2.410

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, *6*(1), 3–5. doi:10.1177/1745691610393980

Chandler, J., & Paolacci, G. (n.d.). Lie for a dime: When most prescreening responses are honest but most "Eligible" respondents are lies.

Cheung, J. H., Burns, D. K., Sinclair, R. R., & Sliter, M. (2017). Amazon Mechanical Turk in Organizational Psychology: An Evaluation and Practical Recommendations. *Journal of Business and Psychology*, *32*(4), 347–361. doi:10.1007/s10869-016-9458-5

Collins, J. M., & Gleaves, D. H. (1998). Race, Job Applicants, and the Five-Factor Model of Personality: Implications for Black Psychology, Industrial/Organizational Psychology, and the Five-Factor Theory. *Journal of Applied Psychology*, *83*(4), 531–544.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLOS ONE*, *8*(3), e57410. doi:10.1371/journal.pone.0057410

Dalal, R. S. (2005). A Meta-Analysis of the Relationship Between Organizational Citizenship Behavior and Counterproductive Work Behavior. *Journal of Applied Psychology*, *90*(6), 1241–1255. doi:10.1037/0021-9010.90.6.1241

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). Marina Del Rey, CA, USA: ACM Press. doi:10.1145/3159652.3159661

Eriksson, K., & Simpson, B. (2012). What do Americans know about inequality? It depends on how you ask them. *Judgment and Decision Making*, *7*(6), 5.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples: Data Collection in a Flat World. *Journal of Behavioral Decision Making*, *26*(3), 213–224. doi:10.1002/bdm.1753

Graeber, D. (2018). *Bullshit jobs*. New York, NY: Simon & Schuster.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2-3), 61–83; discussion 83–135. doi:10.1017/S0140525X0999152X

Highhouse, S., & Zhang, D. (2015). The New Fruit Fly for Applied Psychological Research. *Industrial and Organizational Psychology*, *8*(02), 179–183. doi:10.1017/iop.2015.22

Ilgen, D. R., & Bell, B. S. (2001). Conducting Industrial and Organizational Psychological Research: Institutional Review of Research in Work Organizations. *Ethics & Behavior*, *11*(4), 395–412. doi:10.1207/S15327019EB1104_03

Johns, G. (2006). The Essential Impact of Context on Organizational Behavior. *Academy of Management Review*, *31*(2), 386–408. doi:10.5465/amr.2006.20208687

Judge, T. A., & LePine. (2007). The bright and dark sides of personality: Implications for personnel selection in individual and team contexts. In J. Langan-Fox, C. Cooper, & R. Klimoski (Eds.), *Research Companion to the Dysfunctional Workplace*. Edward Elgar Publishing. doi:10.4337/9781847207081

Judge, T. A., & Zapata, C. P. (2015). The PersonSituation Debate Revisited: Effect of Situation Strength and Trait Activation on the Validity of the Big Five Personality Traits in Predicting Job Performance. *Academy of Management Journal*, *58*(4), 1149–1179. doi:10.5465/amj.2010.0837

Keith, M. G., Tay, L., & Harms, P. D. (2017). Systems Perspective of Amazon Mechanical Turk for Organizational Research: Review and Recommendations. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.01359

Kennedy, R., Clifford, S., Burleigh, T., Jewell, R., & Waggoner, P. (2018, October). The shape of an solutions to the MTurk quality crisis.

Kish-Gephart, J. J., Harrison, D. A., & Treviño, L. K. (2010). Bad apples, bad cases, and bad barrels: Meta-analytic evidence about sources of unethical decisions at work. *Journal of Applied Psychology*, *95*(1), 1–31. doi:10.1037/a0017103

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco, CA: Jossey-Bass.

Landers, R. N., & Behrend, T. S. (2015). An Inconvenient Truth: Arbitrary Distinctions Between Organizational, Mechanical Turk, and Other Convenience Samples. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *8*(2), 142–164. doi:10.1017/iop.2015.13

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The Demographic and Political Composition of Mechanical Turk Samples. *SAGE Open*, *6*(1), 215824401663643. doi:10.1177/2158244016636433

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

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. doi:10.1037/a0028085

Michel, J. S., O'Neill, S. K., Hartman, P., & Lorys, A. (2018). Amazon's Mechanical Turk as a Viable Source for Organizational and Occupational Health Research. *Occupational Health Science*, *2*(1), 83–98. doi:10.1007/s41542-017-0009-x

Newman, D. A., Joseph, D. L., & Feitosa, J. (2015). External validity and multi-organization samples: Levels-of-analysis implications of crowdsourcing and college student samples. *Industrial and Organizational Psychology*, *8*(2), 214–220. doi:10.1017/iop.2015.28

O'Boyle, E. H., Forsyth, D. R., Banks, G. C., & McDaniel, M. A. (2012). A meta-analysis of the Dark Triad and work behavior: A social exchange perspective. *The Journal of Applied Psychology*, *97*(3), 557–579. doi:10.1037/a0025679

Offermann, L. R., Coats, M. R., & Rauch, C. (2015). What do I owe my organization? MTurk at work. In *Annual Conference of the Society for Industrial and Organizational Psychology, Inc.* Philadelphia, PA.

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

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). *Running Experiments on Amazon Mechanical Turk* (SSRN Scholarly Paper No. ID 1626226). Rochester, NY: Social Science Research Network.

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023–1031. doi:10.3758/s13428-013-0434-y

Roulin, N. (2015). Don't Throw the Baby Out With the Bathwater: Comparing Data Quality of Crowdsourcing, Online Panels, and Student Samples. *Industrial and Organizational Psychology*, *8*(02), 190–196. doi:10.1017/iop.2015.24

Simons, D. J., & Chabris, C. F. (2012). Common (Mis)Beliefs about Memory: A Replication and Comparison of Telephone and Mechanical Turk Survey Methods. *PLOS ONE*, *7*(12), e51876. doi:10.1371/journal.pone.0051876

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 and Organizational Psychology*, *8*(02), 220–228. doi:10.1017/iop.2015.29

Stewart, N., Ungemach, C., Harris, A. J. L., Bartels, D. M., Newell, B. R., Paolacci, G., & Chandler, J. (2015). The Average Laboratory Samples a Population of 7,300 Amazon Mechanical Turk Workers. *Judgment and Decision Making*, 15.

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, *88*(3), 500–517. doi:10.1037/0021-9010.88.3.500

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*(02), 171–179. doi:10.1017/iop.2015.21

Table 1

Proportions of individuals sampled from each intermediate occupational category (2015).

Intermediate Aggregation Title	Expected	n1	LL1	Obs1	UL1	n2	LL2	Obs2	UL2
11-13 Management, Business, and Financial	0.10	129	0.15	0.18	0.21	240	0.21	0.24	0.26
15-19 Computer, Engineering, and Science	0.05	106	0.12	0.15	0.17	117	0.10	0.12	0.14
21-27 Education, Legal, Community Service, Arts, and Media	0.10	134	0.16	0.18	0.21	165	0.14	0.16	0.19
29 Healthcare Practitioners and Technical	0.06	46	0.05	0.06	0.08	36	0.02	0.04	0.05
31-39 Service	0.21	85	0.09	0.12	0.14	142	0.12	0.14	0.16
41 Sales and Related	0.10	58	0.06	0.08	0.10	92	0.07	0.09	0.11
43 Office and Administrative Support	0.16	125	0.14	0.17	0.20	165	0.14	0.16	0.19
45 Farming, Fishing, and Forestry	0.00	2	0.00	0.00	0.01	3	0.00	0.00	0.01
47 Construction and Extraction	0.04	10	0.01	0.01	0.02	11	0.00	0.01	0.02
49 Installation, Maintenance, and Repair	0.04	6	0.00	0.01	0.02	2	0.00	0.00	0.01
51 Production	0.07	10	0.01	0.01	0.02	25	0.02	0.02	0.04

53 Transportation and Material Moving	0.07	12	0.01	0.02	0.03	13	0.01	0.01	0.02
55 Military Specific	0.01	4	0.00	0.01	0.01	5	0.00	0.00	0.01

Note. LL = lower 95% confidence limit; Obs = observed proportion; UL = upper 95% confidence limit.

Table 2

Demographic distribution (gender, ethnicity, and education) of labor force sampled indirectly from MTurk (2015).

Characteristic	Expected	n1	LL1	Obs1	UL1	n2	LL2	Obs2	UL2
Gender									
Female	0.47	451	0.54	0.57	0.61	804	0.6	0.63	0.66
Male	0.53	334	0.39	0.43	0.46	474	0.34	0.37	0.4
Ethnicity									
White	0.79	635	0.78	0.81	0.83	1028	0.78	0.8	0.82
Black	0.12	41	0.04	0.05	0.07	55	0.03	0.04	0.06
Asian	0.06	37	0.03	0.05	0.06	59	0.04	0.05	0.06
Other	0.03	73	0.07	0.09	0.12	140	0.09	0.11	0.13
Education									
no high school	0.08					4	0	0	0.01
high school diploma/GED	0.26					108	0.09	0.1	0.12
some college	0.17					245	0.21	0.24	0.26
2 year degree	0.11					136	0.11	0.13	0.15
Bachelors or higher	0.38					549	0.5	0.53	0.56

Note. LL = lower 95% confidence limit; Obs = observed proportion; UL = upper 95% confidence limit.

Table 3

Proportions of individuals sampled from each intermediate occupational category (2017).

Intermediate Aggregation Title	Expected	n3	LL3	Obs3	UL3
11-13 Management, Business, and Financial	0.10	122	0.17	0.20	0.23
15-19 Computer, Engineering, and Science	0.06	85	0.11	0.14	0.17
21-27 Education, Legal, Community Service, Arts, and Media	0.10	102	0.14	0.17	0.20
29 Healthcare Practitioners and Technical	0.06	24	0.02	0.04	0.06
31-39 Service	0.21	74	0.10	0.12	0.15
41 Sales and Related	0.10	49	0.06	0.08	0.10
43 Office and Administrative Support	0.15	114	0.16	0.18	0.22
45 Farming, Fishing, and Forestry	0.00	4	0.00	0.01	0.02
47 Construction and Extraction	0.04	2	0.00	0.00	0.01
49 Installation, Maintenance, and Repair	0.04	15	0.01	0.02	0.04
51 Production	0.06	13	0.01	0.02	0.04
53 Transportation and Material Moving	0.07	10	0.01	0.02	0.03
55 Military Specific	0.01	2	0.00	0.00	0.01

Note. LL = lower 95% confidence limit; Obs = observed proportion; UL = upper 95% confidence limit.

Table 4

Demographic distribution (gender and ethnicity) of labor force sampled indirectly from MTurk (2017).

Characteristic	Expected	n3	LL3	Obs3	UL3
Gender					
Female	0.47	441	0.62	0.66	0.7
Male	0.53	226	0.3	0.34	0.38
Ethnicity					
African American	0.12	45	0.05	0.07	0.09
Asian	0.06	31	0.03	0.05	0.06
Caucasian	0.78	532	0.76	0.8	0.83
Other	0.03	59	0.07	0.09	0.11

Note. LL = lower 95% confidence limit; Obs = observed proportion; UL = upper 95% confidence limit.

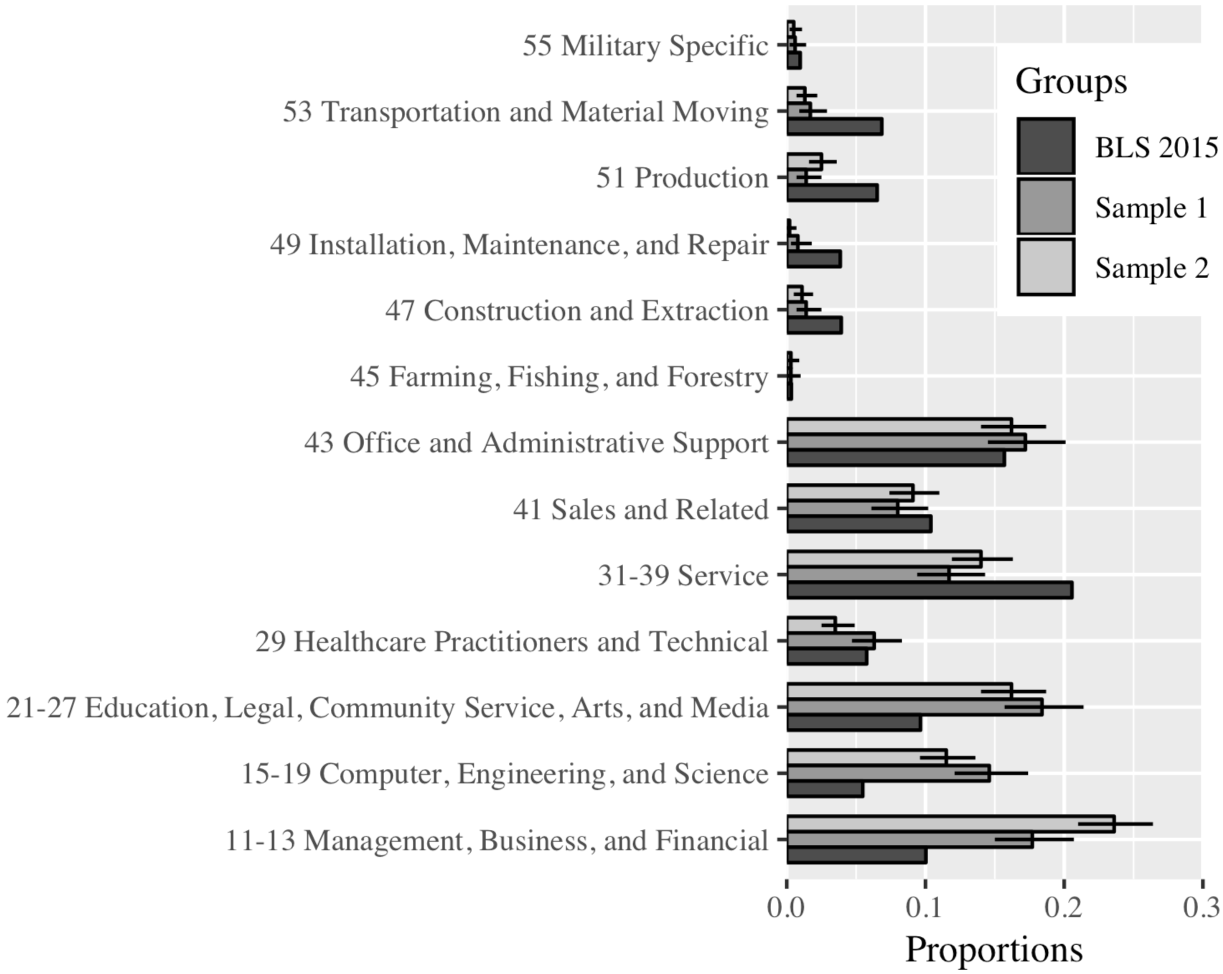

Figure 1. Comparing the occupational diversity of two MTurk samples (2015) with expectations based on the Bureau of Labor Statistics.

Figure 2. Comparing the occupational diversity of one MTurk sample (2017) with expectations based on the Bureau of Labor Statistics.