Comparing MTurk and the US Population's Occupational Diversity

Christopher M. Castille[1], Bodour H. Mahmoud[2], Rachel L. Williamson[3], John E. Buckner[4],

& James De Leôn[5]

[1] Nicholls State University

[2] Villanova University

[3] University of Georgia

[4] AlixPartners

[5] APT Metrics

Author Note

All correspondence should be sent to the lead author at chris_castille@mac.com.

Correspondence concerning this article should be addressed to Christopher M. Castille.

E-mail: chris_castille@mac.com

Abstract

We compare the occupational diversity of two convenient MTurk samples against the broader US economy from which they have been indirectly sampled. We focus on occupational diversity given a chief concern that findings from OB/HR investigations relying on MTurk might not generalize to the broader US economy . Additionally, as researchers may hope to target specific populations (e.g., sales professionals) understanding the extent to which these professionals are represented in the MTurk population would be helpful. Therefore, we compare the occupational diversity of MTurk against that described by the Bureau of Labor Statistics, testing the null hypothesis that these populations are equivalent in terms of occupational diversity. Across both samples, which were obtained from studies conducted in 2015, we found differences suggesting that the MTurk population is generally overrepresented by white collar professionals (though this is an imperfect trend) and generally underrepresented by blue collar workers.

Keywords: MTurk, Occupational Diversity

Word count: 3000

Comparing MTurk and the US Population's Occupational Diversity

Introduction

Amazon's Mechanical Turk (MTurk), an online labor market created by Amazon, has recently become popular among social scientists interested in collecting inexpensive, high-quality data from a diverse population (Aguinis and Lawal (2012) Behrend, Sharek, Meade, and Wiebe (2011) Bergman and Jean (2016) Paolacci and Chandler (2014) Shank (2016)). Leveraging the strengths of this tool, MTurk has been used by social scientists to replicate well-known phenomena (e.g., reaction times, priming, task switching; Crump, McDonnell, and Gureckis (2013); memory, Simons and Chabris (2012)) and by organizational behavior (OB) researchers to expand the OB literature (e.g., Yarkoni, Ashar, and Wager (2015)).

Yet, in spite of the growing enthusiasm surrounding the use of MTurk, it represents merely another convenient sampling methodology that comes with a few constraints, such as repeated participation or motivation due to compensation (Landers and Behrend (2015) Woo, Keith, and Thornton (2015)) and other yet-to-be-identified constraints. Here, we examine one potentially important constraint for using MTurk to conduct OB research: the occupational diversity of MTurk. Given that a fundamental goal of most OB research is that workers are sampled from an appropriate organizational setting (e.g., Pathak (2008)), understanding the occupational diversity of MTurk will improve our understanding of the boundary conditions or limits to the generalizability of MTurk research findings. By examining the occupational diversity of MTurk, our work addresses a call for research that describes the kinds of individuals who participate in MTurk studies (Griggs et al. (2016)).

Review of the Literature

What is MTurk?

MTurk is an internet marketplace where people seeking laborers have access to a population of workers willing to do tasks requiring a human intelligence for a small fee (Behrend et al. (2011)). Billed as a "marketplace for work that requires human intelligence", MTurk allows "requesters" to post Human Intelligence Tasks (HITs), which serve as self-contained jobs in which "workers" can participate. Via MTurk, researchers seeking participants often post survey studies for workers to voluntarily complete (Goodman, Cryder, and Cheema (2013)). Workers are given a short description about a project, including details about various tasks, estimated time investment, and payment, and may selectively choose which projects to complete. Here, researchers seeking participants often post survey studies for workers to voluntarily complete (Goodman et al. (2013)).

Why is MTurk Valuable for Researchers Studying Working Populations?

MTurk offers practical advantages that may not be available via other more traditional methods of data collection, such as sampling from specific organizations (Paolacci, Chandler, and Ipeirotis (2010)). The platform allows for easier subject pool access at an affordable cost and for subject prescreening (Mason and Suri (2012)). Additionally, it allows for subject anonymity, subject identifiability in longitudinal studies, and offers a simple and supportive infrastructure (Paolacci et al. (2010)). More importantly, its diversity is a key strength that sets it apart from other commonly studied populations (e.g., university students, community samples), making it an attractive subject pool for researchers to draw upon (Buhrmester, Kwang, and Gosling (2011); Erikson & Simpson, 2010; Paolacci et al. (2010)). For instance, Behrend et al. (2011) found that MTurk workers are more diverse on variables such as age, ethnicity, education, employment status, job experience, profession, nationality, and personality. The diversity of MTurk allows OB researchers to overcome concerns regarding the generalizability of their findings (e.g., Paolacci et al. (2010)), which has been a key

concern facing OB research (Chiao and Cheon (2010)).

However, while researchers have praised MTurk for its diversity, few organizational behavior researchers have considered how MTurk might be less diverse than expected in a key respect: occupational diversity. To our knowledge, no study has compared the occupational diversity of MTurk to the broader US population. While Behrend et al. (2011) work suggests that MTurk's occupational diversity is greater than the university population, our work expands on their research by comparing MTurk's occupational diversity to the broader US population. If the occupational diversity of the broader US population is indeed reflected in MTurk, findings from MTurk should better generalize across occupations. Alternatively, if MTurk's occupational diversity over- or under-represent certain occupations, generalizability of findings from MTurk should be more limited.

Research Question: Is the occupational diversity of the MTurk population representative of the broader US economy?

Methods

US Population Occupational Summary Statistics

Occupational Employment Statistics (OES) were obtained from the United States Bureau of Labor Statistics (BLS) (2015). Occupation diversity was measured by calculating the percentage of job holders falling under a particular occupational category. We adopted the Intermediate Aggregation and Standard Occupational Classification (SOC) used by the Bureau of Labor Statistics (BLS) for this study (see table 1). Percentages describe the percentage of workers in the US economy who work in occupations that fall under this broader occupational category.

- - - - - - - - - - - -

insert table 1 about here

- - - - - - - - - - - -

MTurk Occupational Summary Statistics

Participants' data were collected from two independently conducted studies ($n_1 = 779$; $n_2 = 1108$). In one study ($n_1 = 779$), occupational data was self-reported by participants who were asked to described their job title. They were then brought to ONET online and asked to retrieve the six-digit occupational classification code that most closely corresponded to their job title. In the second study ($n_2 = 1108$), participants again self-reported their job title. However, the six-digit occupational classification codes were obtained by training research assistants to match the self-reported job titles to those in the ONET database. Using this data, we described the occupational diversity of each sample using relative frequencies of occupational category representation as described by Table 1.

Analysis

A chi-square test of goodness-of-fit was used to test whether proportions from MTurk samples deviated from expected proportions as derived from the BLS. Specifically, we tested whether MTurk's occupational diversity deviated ($p < .05$) from that expected if we assume the BLS population parameters generalize to the MTurk population. This tests the null hypothesis that the populations are identical in terms of occupational diversity. To facilitate interpretation of our data, we also caluculated 95% confidence intervals for our sample proportions, which allows a visual inspection of whether or not the population value is within sampling error.

Results

Results shown in our table indicate that MTurk's occupational frequency distribution does not consistently align with the US economy for both sample 1 ($\chi^2(12) = 368.58, p < .001$) and sample 2 ($\chi^2(12) = 486.07, p < .001$). Examining the 95% confidence intervals across both samples, certain occupations were consistently over- or under-represented. More specifically individuals sampled from MTurk occupied generally more white-collar jobs (e.g., Management, Business, and Financial Occupations; Computer, Engineering, and Science Occupations) and generally fewer blue-collar jobs (e.g., Construction and Extraction Occupations; Installation, Maintenance, and Repair Occupations; Production Occupations) than what would be expected based on BLS data. For instance, 10.13% of workers in the US economy in 2015 worked in "Management, Business, and Financial Occupations." By contrast, approximately 17% and 23% of workers from the MTurk samples fell under this classification. These trends were generally consistent across both study 1 and 2.

- - - - - - - - - - - -

insert table 2 about here

- - - - - - - - - - - -

Discussion

What started our investigation was a simple research question: is the occupational diversity of the MTurk population representative of the broader US population? Our data suggest that the answer is not consistently so: certain occupations are over-represented, some fall within expectations, and still others are under-represented. Thus, OB researchers using convenient sampling procedures that draw on the MTurk population should consider whether occupational features might play a role in their investigation, which is point that we will elaborate upon in greater detail. Simply put, our results suggest that MTurk's (US)

occupational diversity does not align with the broader US economy.

Recommendations for Researchers Using MTurk

Given that OB researchers will likely wish to continue using MTurk as a means of gathering diverse, convenient, and low-cost survey sample data, researchers should consider how differences in the occupational diversity observed amongst MTurk workers relative to the US population might impinge upon the generalizability of their findings (Newman, Joseph, and Feitosa (2015)). Specifically, given the over-representation of workers from (1) Management, Business, and Financial, (2) Computer, Engineering, and Science, and (3) Education, Legal, Community Service, Arts, and Media occupations, we might safely assume that many of our findings from MTurk will generalize across jobs classified under these broader over-represented occupational categories. However, as many occupations were under-represented [e.g., (1) Service, (2) Construction and Extraction, (3) Installation, Maintenance, and Repair, (4) Production; and (5) Transportation and Material Moving], questions may arise as to whether findings will generalize to jobs falling under these broad classifications. To address this issue, researchers should consider capturing occupational data and modeling theoretically derived moderators. Person-environment (P-E) fit theories should help guide researchers in proposing how certain occupational groupings might influence individual-level relationships, lending to more nuanced and contextualized results (How, Meade, Behrend, and Lance (2009)). Further, O*NET might be useful in this regard as it provides researchers a means for quantifying key occupational differences (e.g., level of competition requirements, social skills requirements) and thereby testing P-E fit hypotheses pertaining to occupational group differences (see Judge and Zapata (2015)).

Limitations and Future Research

One key limitation is that we assumed that O*NET's classification system perfectly captured all occupations during 2015. Additionally, as both studies involved convenient sampling procedures, it remains uncertain the extent to which these samples adequately

reflect the larger MTurk population. Still, the consistency of our findings should be considered by future OB/HR researchers hoping to draw on MTurk for their research with the hopes of generalizing to the larger (or a specific) working US population. Additionally, as our studies were conducted in the US, it remains unclear how other MTurk populations might reflect (however imperfectly) their respective broader economy. 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. Lastly, examining diversity in other respects (age, racio-ethnic, income, education, etc.) should also be considered. Though MTurk offers a practically feasible method of attaining high quality data, future research will be needed to understand the limitations of this tool.

Given that our data were gathered and analyzed in 2015, 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 US economy. We also encourage researchers to consider novel ways of approximating the occupational diversity of both the US economy and MTurk because 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 over-represented 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 OB studies using MTurk.

Conclusion

We called into question the diversity of MTurk as a key strength and suggested one key area where diversity may be lacking: occupational diversity. Our results suggest that, compared to the broader US population, MTurk is not as diverse, both over- and underrepresenting certain occupations. Future research drawing on MTurk should consider how occupational differences might be theoretically relevant and incorporate these effects in analyses using MTurk sample data to help overcoming these shortcomings.

References

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

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.

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 industrial–organizational psychology literature. *Industrial and Organizational Psychology, 9*(01), 84–113.

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.

Chiao, J. Y., & Cheon, B. K. (2010). The weirdest brains in the world. *Behavioral and Brain Sciences, 33*(2-3), 88–90.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PloS One, 8*(3), e57410.

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*(3), 213–224.

Griggs, T. L., Eby, L. T., Maupin, C. K., Conley, K. M., Williamson, R. L., Griek, O. H. V., & Clauson, M. G. (2016). Who are these workers, anyway? *Industrial and Organizational Psychology, 9*(01), 114–121.

How, I., Meade, A. W., Behrend, T. S., & Lance, C. E. (2009). Dr. strangelove, or. *Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in the Organizational and Social Sciences*, 89.

Judge, T. A., & Zapata, C. P. (2015). The person–situation 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.

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*(02), 142–164.

Mason, W., & Suri, S. (2012). Conducting behavioral research on amazon's mechanical turk. *Behavior Research Methods*, *44*(1), 1–23.

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*(02), 214–220.

Paolacci, G., & Chandler, J. (2014). Inside the turk: Understanding mechanical turk as a participant pool. *Current Directions in Psychological Science*, *23*(3), 184–188.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk.

Pathak, R. P. (2008). *Methodology of educational research.* Atlantic Publishers & Dist.

Shank, D. B. (2016). Using crowdsourcing websites for sociological research: The case of amazon mechanical turk. *The American Sociologist*, *47*(1), 47–55.

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.

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.

Yarkoni, T., Ashar, Y. K., & Wager, T. D. (2015). Interactions between donor agreeableness and recipient characteristics in predicting charitable donation and positive social evaluation. *PeerJ*, *3*, e1089.

Table captions