Are We The Baddies? We May Get The Data We Deserve*

Thomas Jamieson
Victoria University of Wellington
tom.jamieson@vuw.ac.nz

December 1, 2025

Abstract

Much scholarship has debated the merits of samples from Amazon's Mechanical Turk (MTurk) and similar online marketplaces. Prior research has highlighted important, valid concerns about participant attention, 'cheating', fraudulent data, trolling and the use of VPNs which challenge the validity of work using these samples. However, while researchers must take these concerns seriously, we may be simply getting the data we deserve. In this paper, I argue that it is important for researchers to also focus on ourselves, and we should adopt more ethical practices that minimize data quality concerns, including signaling the value of the data to participants with adequate pay. Towards this end, I provide recommendations about how to address some ethical and practical implications of low compensation for participants.

Keywords: Crowdsourcing; experiments; MTurk; online research; survey research; research ethics
Word count: 2,909 (all inclusive)

*An early draft of this paper was presented in a research seminar at the University of Nebraska at Omaha. I thank participants in the seminar for their excellent comments and suggestions. Any errors that remain are my own responsibility.

1 Introduction

Over the past 10-15 years, Amazon's Mechanical Turk (MTurk) and similar online marketplaces such as Prolific and Lucid among others have revolutionized human subjects research, providing good-quality, low-cost data for researchers with limited research budgets. Much scholarship has debated the merits of these samples (Berinsky, Huber and Lenz, 2012; Casler, Bickel and Hackett, 2013; Huff and Tingley, 2015; Levay, Freese and Druckman, 2016; Mullinix et al., 2015), focusing on participant attention (Berinsky, Margolis and Sances, 2014; Hauser and Schwarz, 2016; Thomas and Clifford, 2017), cheating (Clifford and Jerit, 2016; Motta, Callaghan and Smith, 2017), fraudulent data (Kennedy et al., 2020), the use of VPNs complicating the interpretation of results (Loepp and Kelly, 2020), or trolling (Ahler, Roush and Sood, 2025). The situation has become so dire that (Ahler, Roush and Sood, 2025, 16) argued MTurk is becoming a ""market for lemons" in which bad actors eventually crowd out the good."

These are valid concerns that researchers must take seriously, but we may simply be getting the data we deserve.[1] In a recently published article, Ahler, Roush and Sood (2025) recommend a few steps researchers can take to improve data quality as much as possible, but the elephant in the room remains unmentioned — that inadequate compensation riddles crowdsourcing marketplaces and that researchers may (also) be bad actors. Online forums such as `Turkopticon` are rife with worker reports of under- or non-payment after survey completion, so it perhaps unsurprising that workers exploit loopholes and take liberties in completing surveys. This is particularly important because resource constraints mean many researchers have little choice options but to use these marketplaces to gather data from online convenience samples and make lemonade from lemons.

As questions around the validity of data from online marketplaces intensify, it is an opportune time to discuss whether researchers bear some responsibility for the crisis in crowdsourced data quality. In this paper, I argue that it is important for us to turn the focus on ourselves as well as on participants, and adopt research practices that minimize data quality concerns, including signaling the value of the data to participants with adequate pay.

[1] In this piece I use personal and collective pronouns to describe the situation to explicitly acknowledge that I am also part of the research community that frequently uses crowdsourced marketplaces for research due to their affordability and accessibility, so any criticisms made here are also applicable to myself.

2 The Data We Deserve?

Minimal norms and expectations around research ethics for low-risk research include protections related to human subjects' privacy and confidentiality (Acquisti, Brandimarte and Loewenstein, 2015; Jamieson and Salinas, 2018), the obligation to receive informed consent from participants (Singer, 1978), and avoiding deception where possible (Sharpe, Adair and Roese, 1992). However, it is worth considering these principles as a bare minimum and insufficient if participants are not treated with dignity in other respects.

As such, while it is vital to examine our data and understand the quality of responses, it is equally important to consider our own contribution to the data quality crisis. In most research settings, resources are scarce, so samples derived from online marketplaces can be especially attractive as a low-cost alternatives to more expensive samples. However, systematic under-compensation of labor likely have both ethical and practical implications for data quality.

2.1 Ethical Issues with Inadequate Compensation

Arguably, current research practices are unethical. Researchers often pay participants significantly less than the minimum wage.[2] For example, Ahler, Roush and Sood (2025) paid participants 60 cents for a HIT (human intelligence task) that was expected to take ten minutes to complete — a rate of just $3.60 an hour. In the end, participants received a slightly better rate than expected with the median time taken being nine minutes and 33 seconds, resulting in an effective hourly rate of $3.77 (Ahler, Roush and Sood, 2025).

Of course Ahler, Roush and Sood are by no means alone in setting low rates for crowdsourced work, and I have also paid relatively low rates for crowd-sourced labor, including an effective hourly rate of $6.00 in Jamieson and Van Belle (2018, 2022). It is difficult to imagine other situations in which it would be considered appropriate to pay significantly less than the minimum wage, yet researchers routinely expect to collect data from participants while paying as little as possible.

Other articles also highlight the value we place on crowd-sourced labor compared to other labor. For instance Clifford and Jerit (2016) paid $0.50 to MTurk participants and $25 for university staff

[2] Previous research has illustrated how median wages for MTurk workers are just $1-$3 an hour (Auer et al., 2021; Hara et al., 2018; Newman et al., 2021; Saito et al., 2019), and only four percent of MTurk workers earned more than the U.S. federal minimum wage of $7.25 an hour (Hara et al., 2018).

to complete the same survey, explicitly highlighting the different value we place on data from online marketplaces compared to that from our colleagues.[3] While an exhaustive evaluation of compensation for crowdsourced labor across the discipline is beyond the scope of this comment, there are reasons to expect these rates are broadly representative of what is considered a "fair wage" on MTurk (Ahler, Roush and Sood, 2025, 17).[4] Despite these disciplinary norms, our research is ethically compromised if we do not adequately compensate participants for their time.

2.2 Practical Issues with Inadequate Compensation

Beyond ethics, there are also practical issues with inadequate compensation. As discussed earlier, prior research demonstrates myriad reasons for researchers to distrust the validity and honesty of workers' responses (Ahler, Roush and Sood, 2025; Berinsky, Margolis and Sances, 2014; Clifford and Jerit, 2016; Hauser and Schwarz, 2016; Kennedy et al., 2020; Loepp and Kelly, 2020; Motta, Callaghan and Smith, 2017; Thomas and Clifford, 2017). However, less attention has been paid to the ways in which we may have contributed to this distrustful environment.

Previous research suggests workers withdraw their labor if they perceive their wages as being unfair (Akerlof and Yellen, 1990), with empirical research suggesting a relationship between compensation and response quality on MTurk (Litman, Robinson and Rosenzweig, 2015). If participants believe they are not being paid fairly, pools of participants may be contaminated by distrust, which may erode motivation to provide truthful and thorough responses to survey questions and prompts. On the other hand, previous research demonstrates that financial incentives improve the quality of MTurk workers' responses improves (Jamieson and Weller, 2020).

In the absence of adequate pay, it is possible that While workers may have various intrinsic motivations (Kaufmann, Schulze and Veit, 2011), our low rates of pay ensure it *only* makes sense for workers to participate for intrinsic reasons, which researchers have less ability to control. This may explain the trolling illustrated in Ahler, Roush and Sood (2025), which could be the result of workers seeking to find additional utility from their work given they do not receive adequate financial compensation.

Further, if researchers seek to pay more than their peers, Ahler, Roush and Sood (2025, 17)

[3] This comparison refers to the MTurk Study 3 and Campus Staff studies in Clifford and Jerit (2016).

[4] This assumes payment occurs at all, and previous studies report workers' experience of unfair treatment, researcher refusal to pay or rejection of completed tasks (Berg, 2016; Hara et al., 2018).

argue they risk distorting the market and creating "large incentives for foreign workers to try to game the system despite being outside the sample frame." In short, researchers may be getting what we deserve - our low compensation rates have arguably contributed to poor quality data. While scrutiny on workers is important, it is also important for us to look in the mirror.

3 Improving Our Practices

To improve research practices, it is worth considering what researchers' obligations are to their participants.

Although the situation is difficult, there are some practices that we can adopt to address at least some of these ethical and practical implications. While it may be too late to solve problems with marketplaces already synonymous with low compensation, it is also possible for researchers to ensure emerging platforms steer clear of these issues.

Ahler, Roush and Sood (2025) provided a series of well-founded recommendations to ensure workers act in good faith. My recommendations aim to ensure we meet our side of the bargain so that researchers also act in good faith, broadly ordered by their degrees of difficulty.

- First, we can pay participants a reasonable wage for their time. As a bare minimum, we should endeavor to pay rates that equate to the minimum wage where the study is fielded (e.g. $7.25 for the United States). Beyond setting the compensation rate, we can also be conservative when estimating how long it will take survey participants to complete the study and share this information when recruiting participants. If we are concerned about higher rates of compensation affecting the labor market and encouraging deceptive practices among workers, higher rates of pay can be folded into studies as performance-based bonuses that reward the kinds of participation we want to encourage in our studies.

- Second, at a micro-level, err on the side of caution when rejecting HITS. Our tools of detecting fraudulent activity are rapidly improving, and we should use these wherever possible (Ahler, Roush and Sood, 2025). However, if there remains inconclusive evidence, it may be prudent to accept the HIT and avoid false positives given the precarious nature of at least some workers (Berg, 2016), and the potential for spillover effects on other studies if workers become

demotivated after erroneous rejection of their work.

- Third, as researchers and reviewers, we can create a culture of adequate compensation. Just as norms have evolved and standards have emerged for pre-registration, data availability, and replication, it is possible for the discipline to coalesce around compensation for human subjects across all platforms. We should report compensation including an effective hourly rate in manuscripts and articles. However, beyond simply reporting rates, reviewers and editors could critique scholars paying rates below the minimum wage or if they do not report compensation rates, much as we would expect comment if the study did not satisfy other disciplinary standards regarding the treatment of human subjects.

- Fourth, beyond compensation, we can seek to reset the relationship between requesters and workers so that we treat participants as participants in the true sense of the word. This could involve debriefing participants at the end of surveys about the study's objectives, provide links to where outputs will be posted when completed, and actively share and disseminate knowledge. Some third-party vendors like CloudResearch provide researchers with the ability to contact workers directly, so outputs can be shared with participants after publication. By promoting a more transparent and cooperative relationship between researchers and participants, this may help mitigate the cycle of distrust that currently pervades online marketplaces.

4 Discussion

In short, all too often our analysis of data quality focuses exclusively on the good faith of participants while ignoring our own contribution to the data quality crisis. Rather than focusing exclusively on workers, it is essential that we adopt practices to help ensure the long-term health of online marketplaces.

While scholars may not directly or knowingly contribute to an environment of distrust among participants, until now there has been an acceptance of compensation rates that do not indicate we actually value the data or the time of the participants providing it. Researchers should pay reasonable rates, avoid rejecting HITS where possible, cultivate a culture of adequate compensation in our discipline, and treat participants like participants. Ultimately, these (relatively) small changes

may help ensure data quality, but also address the ethical and practical implications of prevailing research practices.

References

Acquisti, Alessandro, Laura Brandimarte and George Loewenstein. 2015. "Privacy and human behavior in the age of information." *Science* 347(6221):509–514.

Ahler, Douglas J., Carolyn E. Roush and Gaurav Sood. 2025. "The micro-task market for lemons: data quality on Amazon's Mechanical Turk." *Political Science Research and Methods* 13(1):1–20.

Akerlof, George A. and Janet L. Yellen. 1990. "The Fair Wage-Effort Hypothesis and Unemployment." *The Quarterly Journal of Economics* 105(2):255–283.

Auer, Elena M., Tara S. Behrend, Andrew B. Collmus, Richard N. Landers and Ahleah F. Miles. 2021. "Pay for performance, satisfaction and retention in longitudinal crowdsourced research." *PLOS ONE* 16(1):e0245460.

Berg, Janine. 2016. "Income Security in the On-Demand Economy: Findings and Policy Lessons from a Survey of Crowdworkers.".
URL: *https://papers.ssrn.com/abstract=2740940*

Berinsky, Adam J., Gregory A. Huber and Gabriel S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20(3):351–368.

Berinsky, Adam J., Michele F. Margolis and Michael W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58(3):739–753.

Casler, Krista, Lydia Bickel and Elizabeth Hackett. 2013. "Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing." *Computers in Human Behavior* 29(6):2156–2160.

Clifford, Scott and Jennifer Jerit. 2016. "Cheating on Political Knowledge Questions in Online Surveys: An Assessment of the Problem and Solutions." *Public Opinion Quarterly* 80(4):858–887.

Hara, Kotaro, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch and Jeffrey P. Bigham. 2018. A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal, QC: pp. 1–14.

Hauser, David J. and Norbert Schwarz. 2016. "Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants." *Behavior Research Methods* 48(1):400–407.

Huff, Connor and Dustin Tingley. 2015. ""Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents." *Research & Politics* 2(3):2053168015604648.

Jamieson, Thomas and Douglas A. Van Belle. 2018. "Agenda setting, localisation and the third-person effect: an experimental study of when news content will directly influence public demands for policy change." *Political Science* 70(1):58–91.

Jamieson, Thomas and Douglas A. Van Belle. 2022. *That Could Be Us: News Media, Politics, and the Necessary Conditions for Disaster Risk Reduction.* Ann Arbor, MI: University of Michigan Press.

Jamieson, Thomas and Güez Salinas. 2018. "Protecting Human Subjects in the Digital Age: Issues and Best Practices of Data Protection." *Survey Practice* 11(2):4405.

Jamieson, Thomas and Nicholas Weller. 2020. "The Effects of Certain and Uncertain Incentives on Effort and Knowledge Accuracy." *Journal of Experimental Political Science* 7(3):218–231.

Kaufmann, Nicolas, Thimo Schulze and Daniel Veit. 2011. "More than fun and money: Worker Motivation in Crowdsourcing – A Study on Mechanical Turk." *AMCIS 2011 Proceedings - All Submissions* .
URL: *https://aisel.aisnet.org/amcis2011*$_s$*ubmissions*/340

Kennedy, Ryan, Scott Clifford, Tyler Burleigh, Philip D. Waggoner, Ryan Jewell and Nicholas J. G. Winter. 2020. "The shape of and solutions to the MTurk quality crisis." *Political Science Research and Methods* 8(4):614–629.

Levay, Kevin E., Jeremy Freese and James N. Druckman. 2016. "The Demographic and Political Composition of Mechanical Turk Samples." *SAGE Open* 6(1):2158244016636433.

Litman, Leib, Jonathan Robinson and Cheskie Rosenzweig. 2015. "The relationship between motivation, monetary compensation, and data quality among US- and India-based workers on Mechanical Turk." *Behavior Research Methods* 47(2):519–528.

Loepp, Eric and Jarrod T. Kelly. 2020. "Distinction without a difference? An assessment of MTurk Worker types." *Research & Politics* 7(1):2053168019901185.

Motta, Matthew P, Timothy H Callaghan and Brianna Smith. 2017. "Looking for Answers: Identifying Search Behavior and Improving Knowledge-Based Data Quality in Online Surveys." *International Journal of Public Opinion Research* 29(4):575–603.

Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman and Jeremy Freese. 2015. "The Generalizability of Survey Experiments." *Journal of Experimental Political Science* 2(2):109–138.

Newman, Alexander, Yuen Lam Bavik, Matthew Mount and Bo Shao. 2021. "Data Collection via Online Platforms: Challenges and Recommendations for Future Research." *Applied Psychology* 70(3):1380–1402.

Saito, Susumu, Chun-Wei Chiang, Saiph Savage, Teppei Nakano, Tetsunori Kobayashi and Jeffrey P. Bigham. 2019. TurkScanner: Predicting the Hourly Wage of Microtasks. In *The World Wide Web Conference*. New York, NY: Association for Computing Machinery pp. 3187–3193.

Sharpe, Donald, John G. Adair and Neal J. Roese. 1992. "Twenty Years of Deception Research: A Decline in Subjects' Trust?" *Personality and Social Psychology Bulletin* 18(5):585–590.

Singer, Eleanor. 1978. "Informed Consent: Consequences for Response Rate and Response Quality in Social Surveys." *American Sociological Review* 43(2):144–162.

Thomas, Kyle A. and Scott Clifford. 2017. "Validity and Mechanical Turk: An assessment of exclusion methods and interactive experiments." *Computers in Human Behavior* 77:184–197.