

# Crowdsourcing as a Tool for Research

by Hanzelle Kleeman



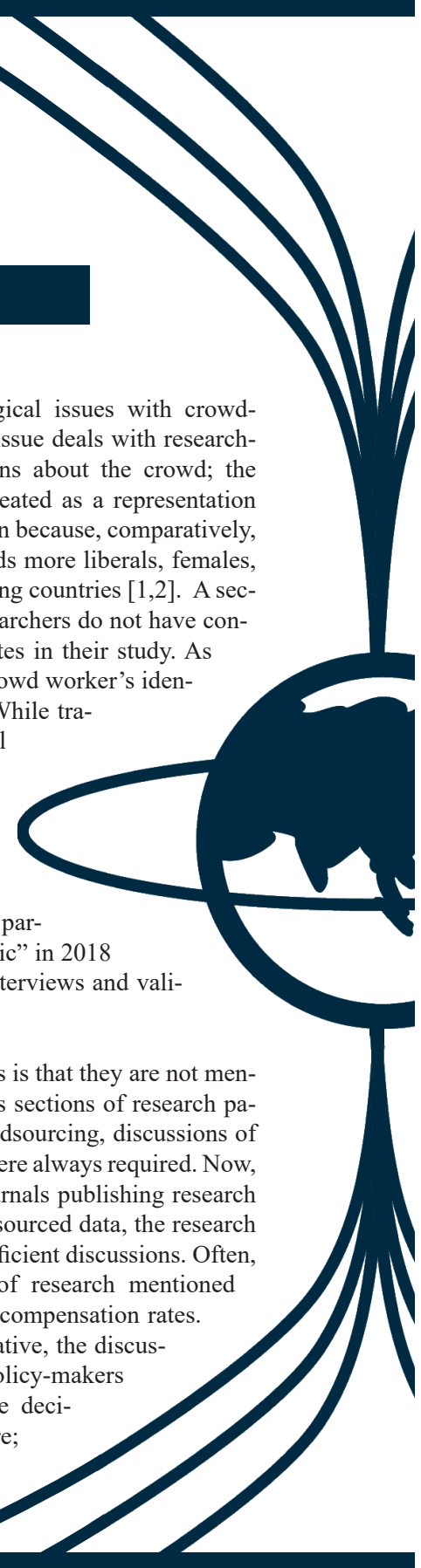
Crowdsourcing platforms are becoming a common tool for researchers around the world. Advocates of these platforms boast that the virtual data collection has many strengths when compared to old-school methods of surveying. Such proponents focus on aspects of convenience and diversity. Unfortunately, serious concerns identified by critical computing scholars remain unaddressed. To better understand these concerns, the research team has completed a comprehensive literature review on the ethics of crowdsourcing, an analysis of published research papers that utilized crowdsourced data, and interviews with crowd workers. These processes are all a part of the initial research phase. The next step will be to solve the identified issues by reimagining and reengineering crowdsourcing platforms. Such efforts are valuable because the concerns raised for crowdsourcing are relevant to the many issues seen within the larger scope of emerging technologies. In this introduction of the project, a glimpse of the identified concerns is shared in three dimensions: methodological, political, and ethical.


## Methodological Concerns

When compared to traditional research tools, crowdsourcing can have many strengths. The ability to reach a larger and more diversified population via crowdsourcing platforms is valuable because it provides statistically significant data sets in a cost and time efficient manner, allows academics to move away from their traditional undergraduate population, and enables researchers to select participants based on certain demographics (e.g., only people who live in California). Unfortunately, scholars have

seen many methodological issues with crowdsourced data. The first issue deals with researchers' implicit assumptions about the crowd; the crowd should not be treated as a representation of the general population because, comparatively, the crowd skews towards more liberals, females, and citizens of developing countries [1,2]. A second concern is that researchers do not have control over who participates in their study. As of now, validation of crowd worker's identities is not a priority. While traditional surveying is still vulnerable to inaccurate or falsified data, the issue is magnified when participation is strictly virtual. A related issue is "bot" participation. The "bot panic" in 2018 is repeatedly cited in interviews and validates such a concern.

A note for all limitations is that they are not mentioned in many methods sections of research papers. Since before crowdsourcing, discussions of limitations and biases were always required. Now, reading through top journals publishing research that is reliant on crowdsourced data, the research team is encountering deficient discussions. Often, the only components of research mentioned are attention checks or compensation rates. While these are informative, the discussion is lacking. The policy-makers using the data to make decisions need to know more;





the reported demographics, the likeliness that the reported demographics are correct, whether that demographic is appropriate, the quality of responses, the responders' familiarity with a topic, the biases associated with the given compensation rates, the probability of “bot” participation, and so on. Researchers familiar with crowdsourcing platforms may think these arguments are obvious or unnecessary, but they need to remember that most people are not yet familiar with crowdsourced data. Ultimately, these themes of low-quality data, security, and widespread education of new technologies are present in all discussions surrounding the governance of emerging technologies.

### Political Concerns

The political issues seen in crowdsourcing are primarily driven by the unbalanced power dynamic between requesters and crowd workers. The basic design of platforms is a major contributor to this issue because it hinders communication between requesters and workers, causing feedback requests and questions to go unanswered. More so, the design gives an opportunity for requesters to rate workers, without the opportunity for workers to rate requesters.

Additional factors contributing to this unbalanced power dynamic is the non-regulation of the digital workplace, the dehumanization of workers, and the lack of citizenship status available in a digital community. The unregulated environment of crowdsourcing platforms causes issues because it leaves crowd workers without the legal protections or bene-

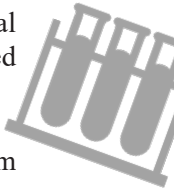
fits experienced by traditional employees. If crowd workers view platforms as their place of employment, platforms should be regulated as a workplace. The dehumanization of crowd workers is a result of this non-workplace, work-place environment; risk of dehumanization is amplified by increased use of machine learning (e.g., leads humans to overlook humans completing micro tasks) and the fact that employee-employer relationships have a short time commitment. The final contributor, lack of a digital community, means that crowd workers have limited opportunities to organize themselves and advocate for employment standards. In completing the literature review, the research team came across an interview conducted by Bivins in which a crowd worker shared, “if I had to summarize my one-week experience as an MTurk worker in one phrase, it would be a sustained sense of powerlessness.” [3].

Possible solutions for decreasing this imbalance would be to create mechanisms for workers to communicate among themselves and with requesters, to provide workers with equitable access to information, and to recognize crowd workers for their contributions to larger goals. In reality, these solutions are similar to the ones used to empower traditional, “real” employees; the difficult part will be convincing society that virtual employees are “real” too.

### Ethical Concerns

As alluded, one of the major ethical concerns with using crowd sourcing as a tool for research is the tendency for workers to be de-valued. Unfortunately, many researchers see crowdsourcing as “a tremendous amount of human computation power [that] is available for accomplishing jobs almost for free” [4]. As a result, crowd workers are subject to incompatible tasks, privacy violations, and labor exploitation. Such descriptions have inspired the phrase “digital sweatshop”.

The most commonly cited complaint among interviewees is unethical payment. This issue is escalated by bad time estimates associated with each type



of task; this is caused by the standardization of tasks that are by nature, not standard. A second contributor to unfair payment is requesters' lack of experience with crowdsourcing platforms. For example, when a requestor publishes a task incorrectly, they may choose to reject the work done by crowd workers. As a result, crowd workers are not paid for that work and receive bad ratings. This has further ethical consequences because low ratings can prevent a worker from being eligible to participate in future requests. A third contributor to low payment is the lack of interest from the I.R.B. (Institutional Review Board) to create guidelines with regards to crowdsourcing as a tool for research. If the I.R.B., as well as research journals, committed to the creation and enforcement of appropriate payment levels, any burdens distributed by researchers could be mitigated.

Another ethical concern is the potential lack of consent and comprehension experienced by crowdsourced research participants. In many cases, tasks will present a warning message if the requester marked the content as sensitive. For crowd workers, this may provide little information as to the degree or subject of sensitivity. This concern is particularly important to the issue of digital research participation, because crowd workers are often persons of diminished autonomy. This pairs well with the discussion of low payment; researchers need to minimize harm to research participants.

Overall, ethical discussions of crowdsourcing pose society with many hard questions (e.g., what is eth-

ical payment?). The problem is that, as computing power continues to increase, and new technology continues to emerge, these questions will multiply.

## Conclusion

Crowdsourced data has become the norm for academia. The aspects that make the tool so valuable – the speed and scale at which research is conducted – are the same aspects that make addressing the methodological, political, and ethical concerns so urgent. In solving the problems introduced here, engineers and social scientists need to collaborate to re-imagine and redesign crowdsourcing platforms. Similar collaboration is required to solve the inevitable challenges that accompany future technologies.

## References

- [1] Barbosa, N. M. Chen, M., "Rehumanized Crowdsourcing: A Labeling Framework," Glasgow, Scotland, UK: CHI Conference, 2019
- [2] I. Stamm, L. Ecklund, "With great power comes great responsibility: crowdsourcing raises methodological and ethical questions for academia," LSE Impact of Social Sciences blog, September 2019
- [3] T. Bivins, "Crowds, Community, and Jeremy Bentham," Journal of Media Ethics, October 2016
- [4] D. Schlagwein, D. Cezec-Kecmanovic, and B. Hanckel, "Ethical norms and issues in crowdsourcing practices: A Habermasian analysis," Wiley, 2018



Hanzelle Kleeman is a third year student in the Mechanical Engineering, Public Affairs, and Engineering Grand Challenges programs. Her work deals with the ethics of emerging technologies and strives to improve the impact of technology by encouraging human-centered design.