25th June 2021

Crowdsourcing for Research in Systematic Cognitive Biases in Search Decision-Making and Evaluations

By Pragya Chatterjee

Abstract

Today, we have numerous crowdsourcing platforms – that outsource available to us as a tool for research as well as commercial ordeals. There is speculation regarding this method due to the uncertainty that it encompasses given that not once can the researcher connect with their human research subject, who participate in this endeavor over the internet. This short paper discusses the positives and negatives of crowdsourcing as a research tool, as well as of the two widely used crowdsourcing platforms that exist to support research and other ventures.

Crowdsourcing as a Tool

Authors of *Crowdsourcing as a Tool for Research: Implications of Uncertainty* define the term 'crowdsourcing' as "efforts that engage large numbers of people over the Web to help collect and process data" (Law et al , 2017, p. 1545), who also report that not many researchers readily use crowdsourcing platforms. In a study conducted by Riesch et al (2014), they interviewed 30 UK scientists and studied their attitude towards volunteer-fueled research – study reports suggest these researchers' negative perceptions were driven by two factors: (1) unrealistic expectations of crowdsourcing; (2) need to persuade the scientific community/colleagues of the effectiveness of crowdsourcing (Riesch et al, 2014) (Law et al, 2017). Riesch's subjects identified ethical issues, like reliance on unpaid individuals, public access to raw data, and concerns regarding risks for junior scientists, as sources of other scientists' skepticism towards crowd-driven research (Law et al, 2017).

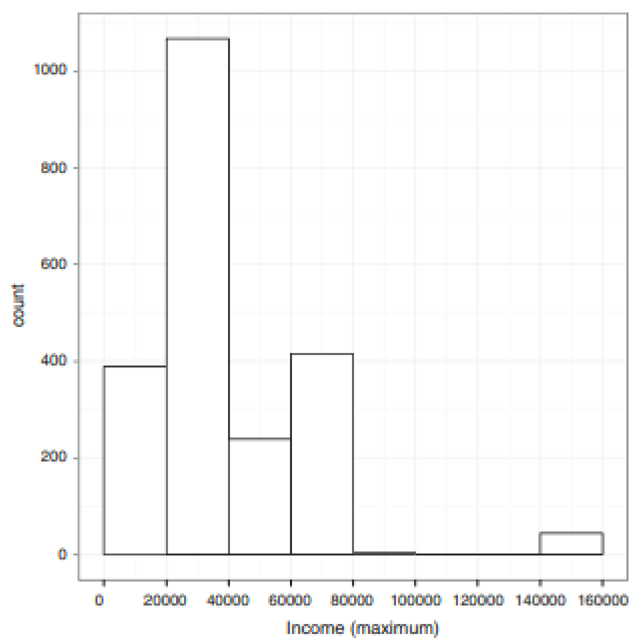
Focus groups of 28 business scholars organized by Schlagwein et al., agreed that it was unethical, as well as uneasy to rely on volunteers, while suggesting that it was imperative to add an integrated payment system (Schlagwein et al, 2014). However, psychology research has shown that offering payments can also reduce the quantity and quality of intrinsically motivated contributions (Law et al, 2014)(Bruno et al, 2001). Another one of Law et al.'s study subjects included a scientist conducting narrative research, concluding that research more aligned with public interests and social action than with publication was more likely to benefit from crowdsourcing (Law et al, 2014). In qualitative research cases, where individuals inspect the same data, crowdsourcing could be a useful tool in potentially contributing novel insights.

Amazon MT

The two most widely used crowdsourcing marketplaces are Amazon Mechanical Turk (AMT) and CrowdFlower (Alonso, 2011). While primarily created for Amazon, AMT has been employed as a platform that connects a wide range of Requestors with up to 500,0000 Providers (Martin et al 2014), while majority of these workers are from USA and India as the only two currency of payment are the US dollar and Indian rupees (Mason et al, 2012). Most studies reported that money is the primary motivator for Providers (also known as Turkers) on this platform (Martin et al 2014). According to Antin and Shaw's study (2014), Social desirability bias is a larger component for US Turkers compared to Indian Turkers (Antin et al, 2014).

Referring to Mason and Suri's *Conducting behavioral research on Amazon's Mechanical Turk* (2012), AMT's unique benefits include: "(1) subject pool access, (2) subject pool diversity, and (3) low cost" (Mason et al, 2012, p. 2). Some other benefits of AMT include: subject availability is fairly stable – with an observed variability due to number of jobs in the market; AMT facilitates international and cross-cultural research with a low cost, built-in payment mechanism; experiments can be created and put on AMT easily and rapidly, which drastically reduces time spent on experimental cycles and execution (Mason et al, 2012). Demographic studies collectively report that majority of the US workers on AMT are females, while the average age is approx. 32 (Mason et al, 2012).

According to a study conducted in 2010, 12% US workers and 27% Indian workers reported that AMT is their primary source of income (Mason et al, 2012). The figure below from Mason et al (2012), shows us the distribution of maximum of the income interval in USD as self-reported by AMT workers (Mason et al, 2012, p.5)



Distribution of the maximum of the income (in U.S. dollars) interval self-reported by workers

Most income-related studies on AMT workers suggest that majority of the workers are not using AMT to make a living. Huang et al. suggests that an abundance of workers in found between Tuesdays and Saturdays. While faster completion times range between 6 AM and 3 PM GMT (Huang et al, 2010). Considering the pandemic economic downturn, as well as the lockdowns in various parts of the world (including India), it is possible that the abundance of workers on AMT is around its peak. Requesters can decide how much they would like to pay the participants, while AMT platform fees are as follows: 20% on reward/bonus you (the requester) pay the workers, minimum fee is $0.01 per assignment or bonus payment.

To ensure quality results, multiple submissions by the same worker can be blocked by using external Human Intelligence Tasks (or HITs), while another important policy forbids workers from having bots to automatically do their work for them – infringement of these policies are rare, but there exist some legitimate workers who could be described as “spammers” (Mason et al, 2012). To keep spammers in check, AMT has a mechanism that allows the requester to reject a worker’s submission, which then goes on their record (Mason et al, 2012).

To run studies on AMT, one must sign up as a requester – two or three accounts are required to register as a requester depending on your use of the interface, a requester account, an Amazon Payments Account, and, if needed, an Amazon Web Services account. To sign up for a requestor account, visit the following web address: Amazon Requestor. The email used to set up the account is the one that will be used to communicate to the researcher. And to set up an Amazon Payments account, access the following web address: Amazon Payments. For funding, either a US credit card or US bank account can be used, including a US billing address as AMT stopped supporting requesters from outside the US since October 2007, although international workers are accepted. To interact with the Amazon API, follow the following web address: AWS. If reported by workers, Amazon has the ability to ban the respective requesters. (Mason et al, 2012)

Appen (CrowdFlower)

Appen, previously known as CrowdFlower, then FigureEight, is an aggregator platform that delegates tasks to multiple partner channels through recruited users. More than one billion tasks have been completed through Appen as of 2015, while work worth 5 human years is being complete on this platform every single day (Winter et al., 2015). Crowdsourcing platforms other than AMT and Appen seem to be focusing on outsourcing jobs such as writing, editing, transcribing, tagging, internet searching, etcetera. In 2014, the number of studies using Appen for research has increased, and as of 2015, it has been used in behavioral, psychological, and linguist experiments, i.e., to investigate public perceptions (Winter et al., 2015).

More recently though, Appen has narrowed its sight to focus mainly on collecting labelled training data for AI, with a global crowd of over 1 million contractors who speak over 235 languages, in over 170 countries. To attract workers, Appen poses the following examples as work projects: search media evaluation, social media evaluation, translation services, transcription services, survey and data collection, linguistic specialties, lexicon annotation, speech evaluation, image annotation and/or transcription, video annotation, and sensor data annotation. This platform utilizes Payoneer or PayPal to pay their workers. With no indication of any other service except AI data modeling on their official website, it might be safe to assume it is not the most ideal platform to be connected to human research subjects.

Conclusion

Another crowdsourcing marketplace, used mainly for research purposes, is called Prolific, that costs a minimum of $6.50/hour per participant. Other crowdsourcing platforms such as Toloka and InCognitoMatch are fairly new with very little research done about it. They are not widely used for research purposes, especially Toloka, which, similarly to Appen, collected training models for Artificial Intelligence projects. For research in Systematic Cognitive Biases in Search Decision-Making and Evaluations, AMTurk might be our best bet due to its wide range of usability, worker diversity, easy use, low cost, worker availability, and the fact that it has been previously used in quite a few research projects including behavioral and psychology research.

References

Alonso, Omar, & Lease, Matthew. (2011). Crowdsourcing for information retrieval. *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1299-1300.

Edith Law, Krzysztof Z. Gajos, Andrea Wiggins, Mary Gray & Alex Williams. (2017). Crowdsourcing as a Tool for Research: Implications of Uncertainty. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1544- 1561.

Hauke Riesch and Clive Potter. 2014. Citizen science as seen by scientists: Methodological, epistemological, and ethical dimensions. Public Understanding of Science 23, 1 (2014), 107–120. DOI: http://dx.doi.org/10.1177/0963662513497324

Schlagwein Daniel and Daneshgar Farhad. 2014. User requirements of a crowdsourcing platform for researchers : findings from a series of focus groups. In Proceedings of the 2014 Pacific Asia Conference on Information Systems. 1–11.

Bruno S. Frey and Reto Jegen. 2001. Motivation Crowding Theory. Journal of Economic Surveys 15, 5 (2001), 589–611. DOI: http://dx.doi.org/10.1111/1467-6419.00150

Martin, David, Hanrahan, Benjamin, O'Neill, Jacki, & Gupta, Neha. (2014). Being a Turker. Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, 224-235.

Antin, J. and Shaw, A. (2012). Social desirability bias and self-reports of motivation: A study of amazon mechanical turk in the us and india. *In Proc. of CHI '12, ACM press (2012),* 2925–2934.

Mason, Winter, & Suri, Siddharth. (2012). Conducting behavioral research on Amazon's Mechanical Turk. Behavior Research Methods, 44(1), 1-23.

Huang, E., Zhang, H., Parkes, D. C., Gajos, K. Z., & Chen, Y. (2010). Toward automatic task design: A progress report. *In Proceedings of the ACM SIGKDD Workshop on Human Computation* (pp. 77–85). New York: ACM.

Winter, Joost de, Kyriakidis, M., Dodou, Dmitra, & Happee, Riender (2015). Using CrowdFlower to study the relationship between self-reported violations and traffic accidents. *6th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences, AHFE 2015.* 2518- 2525.