

Response Paper: The Generalizability of Online Experiments Conducted During the COVID-19 Pandemic

Andreas Peintner

andreas.peintner@uibk.ac.at

March 26, 2023

1 Summary

The work in [5] investigates the generalizability of COVID era survey experiments, which may be challenging due to the changes in how participants respond to treatments. In their study it is shown that pre-pandemic experiments are replicable, but show reduced magnitudes. Evidence supports the claim, that the increased share of inattentive participants on online survey platforms may have contributed to this effect.

The COVID-19 pandemic led to a boom of online research activity which raises concerns about external validity of experiments and their generalizability. Work in [2, 4] described the importance of the time dimension in the setting and that outcomes may change over time (temporal validity). The pandemic imposed two factors to online survey experiments: First, participants experienced massive disruption in their daily lives and may respond differently on treatments. Secondly, the composition of markets for online survey respondents changed. Providers were struggling to meet the increased demand and therefore had to rely on relatively low-quality respondents.

To demonstrate that experiments conducted during the COVID era are generalizable or not, the authors decided to produce 33 replications of 12 previously published survey experiments. Experiments replicated in terms

of sign and significance, but at reduced magnitudes. This leads to the conclusion that the pandemic does not pose a threat to the generalizability of results from online experiments during this period. However, it may imposed changes to the online survey respondent pool. To proof this claim, the authors used a variety of measures, including attention check questions (ACQs), subject meta-data (user agent string) and external estimates of attentiveness (interview with provider). It can be concluded, that declining attentiveness across respondents can explain the decrease in the replication estimates.

2 Discussion

The authors of this study claimed that they relied on 1,000 US-based participants to replicate the experiments. However, it is not stated if the pre-pandemic studies relied on US-based respondents or not. Using only US-based participants can introduce a bias into the studies itself. Also it would be interesting to see if people in a global context would show similar treatment effects. The challenges introduced by the pandemic to e. g., European citizens could lead to different states of anxiety and stress due to contrasting political and social situations.

In this paper it is also shown, that marketplaces for survey respondents boomed in terms of requests of academic clients as well as new survey respondents on the platforms. In my opinion the reasons for this can be manifold: First, the pandemic induced economic insecurity forced people to look out for potential additional income sources to secure their financial status. Further, the pandemic related restrictions and social distancing possibly had an increasing effect on the usage of mobile phones/games. Both factors explain the increased inattention of respondents, since participants are more motivated to finish the survey as fast as possible, without really reading the questions and therefore reduce their compliance with experimental stimuli.

To tackle the issue of inattention in online samples the authors recommend to include pretreatment checks [3], since posttreatment checks can induce bias [1]. However, the authors analyzed in the paper that participants coming from web application spent approximately 7 min less time completing surveys than those from web browsers. Reading questions and answers carefully takes time and therefore in my opinion survey completion time is a good indicator for attentiveness of participants. Additionally, the attention-based filtering on respondents can be performed on a question level, where

the question completion time is below a certain threshold (e. g., time to read the question and answers).

References

- [1] A. J. Berinsky, M. F. Margolis, and M. W. Sances. Separating the shirkers from the workers? making sure respondents pay attention on self-administered surveys. *American journal of political science*, 58(3):739–753, 2014.
- [2] M. G. Findley, K. Kikuta, and M. Denly. External validity. *Annual Review of Political Science*, 24:365–393, 2021.
- [3] J. M. Montgomery, B. Nyhan, and M. Torres. How conditioning on post-treatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, 62(3):760–775, 2018.
- [4] K. Munger. Knowledge decays: Temporal validity and social science in a changing world. *Unpublished manuscript*. <https://osf.io/4utsk>, 2019.
- [5] K. Peyton, G. A. Huber, and A. Coppock. The generalizability of on-line experiments conducted during the covid-19 pandemic. *Journal of Experimental Political Science*, 9(3):379–394, 2022.