Discussion on Berinsky, Druckman, and Yamamoto, "Why Replications Do Not Fix the Reproducibility Crisis: A Model and Evidence from a Large-Scale Vignette Experiment"

Byung-Jae Lee

Social Science Data Innovation Center
Yonsei University

January 12, 2018

Replication Crisis

- Replicability is one of the important criteria for science.
- Replication crisis in social science.
- Two kinds of publication bias
  - *File Drawer Bias* a positive test result is more likely to be published than a negative test result, ceteris paribus.
  - *"Gotcha" Bias* a positive test result is more likely to be published when there exits a prior study that tests the same hypothesis and had a negative result than when a prior study shows a positive result (and vice versa).
- Mechanism behind these mechanisms are proclivity toward novelty and sensationalism.

Two Metrics

- Actual False Positive Rate (AFPR) in published replication studies

$$\tilde{\alpha}_2 = \Pr(\textit{replication test significant/replication published, the null is true})$$

- Reproducibility

$$R = \Pr(\textit{replication test significant} \mid \textit{original test significant and published, replication published})$$

Vignette Experiment

- Political Science Faculty in Ph.D. granting institutions in the US
- Five vignettes each
- Author
- Reviewer
- Editor

Summary Results

- (Extant studies) While respondents, on average, indicated 67.1% chance of submitting, recommending, or supporting a paper with a significant test result, they gave only 45.2% chance of doing the same for a paper with a non-significant finding.
- (Replication studies) On average, 62.5% chance of moving a significant test result in a replication study toward publication, whereas they only gave 44.1% chance for a non-significant replication test result.
- 49.1% chance of submitting/recommending/supporting publication of insignificant test result when the study fails to replicate an earlier test result, compared to 39.1% when it successfully replicates a previously non-significant finding.
- 63.9% chance of making a decision in favor of publishing a replication test result when that replication finds a significant effect which runs contrary to a previously significant replication result that successfully reproduces an earlier significant findings.

Estimating AFPR and Reproducibility

- If original study rejects the null at 0.01 level, AFPR is 0.063. If the original study rejects the null at 0.05, AFPR is 0.060.
- A positive replication result that is significant at the 0.05 level is estimated to have an AFPR 0.079 if the result is contrary to an original null result. (publicatoin bias)
- Replications are not an elixir.
- Publication bias would improve reproducibility by statistically significant marings across all possible values of statistical power and the pre-study odds.
- Publication bias in replication studies might actually have a positive impact on the standard metric of reproducibility.

Questions

- "Garden-of-forking-Paths" and Multiple Comparisons (and p-value hacking)
- Minuscule Intervention? (e.g., Power pose)
- Time-reverse Heuristic
- Are pre-registration and better statistics solutions to replication crisis? Or is the solution in better measurement and design?
- Direct and Indirect Effect of Replication
- Transparency and replicability