Commentary on the Attempt to Replicate the Effect of the American Flag on Increased Republican Attitudes

Melissa J. Ferguson,1 Travis J. Carter,2 and Ran R. Hassin2
1Cornell University, Ithaca, NY, USA, 2Colby College, Waterville, ME, USA, 3Hebrew University, Jerusalem, Israel

Abstract. In this commentary, we reflect on what we have learned from the experience of being part of the ManyLabs replication, both in terms of the phenomenon being studied, and in terms of the mechanics of such an ambitious replication project. Replication attempts like this one are clearly valuable and will continue to inform our science. We discuss a number of lessons we have taken from the process related to the divide between direct and conceptual replication, and whether the data can inform our current theory regarding the original effect. In discussing these issues, we advocate for transparent flexibility in data analyses and the involvement of the original authors at every stage of the process.

Keywords: replication, priming, flag priming

We begin with two points. First, we thank the authors of the ManyLabs project (Klein et al., 2014) and believe that this foray in large-scale replication is important. Second, we do not dispute the conclusion that the result of Carter, Ferguson, and Hassin’s (2011) Study 2 was not replicated.

Many in the field are identifying practices to improve our science, and an increasing willingness to conduct and publish replications can only help. This is a learning process, however, and we view the ManyLabs project not only as a replication of experiments, but also as an experiment in replication, priming, flag priming, conceptual replications difficult (McGuire, 2013).

Consider also differences in samples. The ManyLabs authors do this by including contextual variables (e.g., lab vs. online) in their analyses. Although we did not have any a priori reason to expect in-lab versus online differences, there are significant differences between online and in-lab samples on nearly all the variables related to the replication of our study. This points to the necessity of including potential interactions with sample characteristics in plans and analyses to account for such differences.

Lastly, whereas our original study was a stand-alone experiment, participants in the ManyLabs study engaged in many experiments, many of which contained direct references to the US (i.e., 2, 3, 6, 7, and 10). This introduces variability in the manipulation of whether participants were primed (or not) by the US that cannot be accounted for by controlling for position. Although one could test those who completed the flag study first (which Klein et al. did), this results in a smaller sample size, making a test of our current theoretical model with moderators included (see below) underpowered (McClelland & Judd, 1993).

A different political atmosphere, different subject pools, and different states of mind separate the original and the replication attempt. For these reasons, we view this as a conceptual, and not a direct, replication. We can learn from its failure, just as we can learn from other recent studies that identify moderators to the original phenomenon (e.g., Kalmoe & Gross, 2013).

What About Theory?

Since the original study, we developed a theory to account for the dynamic nature of primes that depend on shifting cultural knowledge. We are testing how to identify those who
possess the relevant implicit associations, and are therefore most likely to be influenced by a flag prime (and how). The ManyLabs team graciously included our proposed moderators. Unfortunately, a test of our model is not possible because the contextual variables cannot be included in the analyses given the effects of those variables on our factors, as noted. The original effect was not replicated, but these data do little to confirm or disconfirm our current model.

The point is that replications are theory-laden. It may take time to develop theories that can fully account for the conditions required to observe a phenomenon (see Cesario, 2014). Hence, understanding the theory behind the effects – and contacting authors for their latest theoretical developments – is an important step in the process.

Data Analyses

In order to create a confirmatory design, the ManyLabs authors preregistered the analyses. This prevents post hoc hypotheses and befits replications. However, data can sometimes surprise us in ways that render the original plan insufficient.

For example, the use of hierarchical regression without including lower-order interaction terms can lead to misleading results. The ManyLabs analyses show that the two 3-way interactions we predicted based on our current model are at $p = .05$, and $p = .07$. However, when the predictors are first centered/standardized, then these interactions are significant at $p < .001$. To be clear, we conducted that these $p$ values are misleading, and do not in fact reflect support for those interactions; when influential lower-order interactions are not included in a model, the higher-order interactions can be difficult to interpret. (We note that although Klein et al. conduct analyses to test our proposed moderators while including all lower order terms, these analyses still do not include the contextual variable (i.e., lab/online) and do not account for whether the sample could have been contaminated by previous ManyLabs studies mentioning the U.S.).

So, although pre-approved plans have their advantages, they should leave room for flexibility. There must be transparent ways to conduct additional analyses when warranted by the data themselves.

Author Contact and Peer Review

Beyond contacting the original authors for materials, procedures, and theoretical updates, it is important that replications are peer-reviewed in a way that will allow the discovery of unintentional flaws. Including the original authors as reviewers serves this goal, and an impartial editor can adjudicate the legitimate concerns while minimizing any motivated cognition. If fact, this is what they do daily.

For example, the figure of effect sizes that has become the symbol of this research includes international samples. Our original study tested the influence of an American flag on Americans’ support for American political policies. There is nothing in the paper to suggest that this effect could be transplanted to other countries.

The bottom line? We believe the failure to conceptually replicate our original study will be informative. We also believe that this experiment in replicating teaches us about the do’s and do not’s of our future science. We thank again the ManyLabs authors.

Acknowledgments

The authors declare no conflict-of-interest with the content of this article. Author contributions: Analyzed data: M.F. and T.C.; Wrote paper: M.F., T.C., and R.H.

References


Published online May 30, 2014

Melissa J. Ferguson

Dept. of Psychology
230 Uris Hall
Cornell University
Ithaca, NY 14853-7601
USA
E-mail MJF44@cornell.edu