Reflections on the Commitment–Forgiveness Registered Replication Report

Eli J. Finkel

Department of Psychology and the Kellogg School of Management, Northwestern University

This Registered Replication Report (RRR) represents the second high-profile replication attempt of a study of mine. The first, a replication of Eastwick and Finkel (2008), was included among the 100 studies in the Reproducibility Project (Open Science Collaboration, 2015), and the results replicated almost identically (Selterman, Chagnon, & Mackinnon, 2015). The second—Study 1 from Finkel, Rusbult, Kumashiro, and Hannon (2002)—was the subject of Cheung et al.'s RRR (2016, this issue), and the results were quite different: The experimental manipulation that produced a sizable difference in subjective commitment in the original study produced a negligible difference between conditions in the RRR—a manipulation check failure. And given that the manipulation failed to induce the relevant psychological state, it also, not surprisingly, failed to influence the forgiveness measures that were hypothesized to be affected by changes in that psychological state.

I am surprised that this particular effect failed to replicate, and not only because it had worked in 2002. This was not a subtle manipulation: I expected that participants' relationship commitment would be higher moments after writing about how their life had become linked to their partner than after writing about how they are independent of their partner.

In Praise of Direct Replication

The present RRR serves as a potent reminder that there is often no substitute for direct replication, even though the process is resource-intensive and inherently imperfect. None of the new norms or standards designed to reduce false-positive rates would have raised red flags about the original study (which comes as no surprise to me: I happen to know that there are no file-drawered variants of this study). Indeed, the original paper fully disclosed imperfect findings. We had predicted four forgiveness effects—higher levels of exit and neglect and lower levels of voice and loyalty—in response to commitment priming, but we observed statistically significant effects for only two of these effects; the paper explicitly acknowledged that the other two effects represented failures to support our hypotheses. The p values for the two statistically significant effects were small (p = .0005 for exit and p = .0100 for neglect). If we had p hacked anything in this study—and I am fairly certain that we did not (no optional stopping, no post-hoc outlier elimination, etc.)—we presumably would have tried to p hack the voice effect to significance, as that p value was ~.15.[1] Even by 2016 standards, our sample size (N = 89) for this two-cell design was not especially small. The effect sizes were not implausibly large: The absolute values of the βs across the four hypothesized effects ranged from .05 to .42 and averaged .24. In short, the study does not stand out as an obvious candidate for problems—and yet, it failed to replicate.

Methodological Implications

Why did the manipulation that successfully influenced commitment in 2002 fail to do so in the RRR? I don't know. The initial procedure was entirely paper-and-pencil, whereas the RRR used computerized procedures for the manipulation check and forgiveness measures. But I had approved that procedural change in advance and continue to have no compelling hypothesis for why the medium (paper-and-pencil vs. computer) would make a difference here.

The RRR used the first version of a manipulation that has been refined in subsequent work. Although I believe that the original manipulation is reasonable, I no longer use it in my own work. For example, I have become concerned that the "low commitment" prime includes some

Corresponding Author:
Eli J. Finkel, Department of Psychology, Northwestern University, Swift Hall 102, 2029 Sheridan Road, Evanston, IL 60208-2710
E-mail: finkel@northwestern.edu

potentially commitment-enhancing elements (e.g., "What is one trait that your partner will develop as he/she grows older?"). As such, my collaborators and I have replaced the original 5-item primes with refined 3-item primes (Hui, Finkel, Fitzsimons, Kumashiro, & Hofmann, 2014). I have greater confidence in this updated manipulation than in the original 2002 manipulation. Indeed, when I first learned that the 2002 study would be the target of an RRR—and before I understood precisely how the RRR mechanism works—I had assumed that it would use this updated manipulation. That said, I would feel a whole lot more confident in the refined manipulation if it were to be supported in a preregistered investigation.

Theoretical Implications

When comparing results from direct replications to those from an initial study, it is important to distinguish apples-to-apples from apples-to-oranges comparisons. The measures and procedures are virtually identical, of course, but that does not necessary mean that all results can be interpreted in the same way. In the RRR, the manipulation check affords an apples-to-apples comparison, and a clear replication failure: The manipulation had the intended effect on subjective commitment in the original study but failed to do so in the RRR.

But the rest of the comparisons are apples-to-oranges—the results from the RRR can't be interpreted in the same way as the results from the original study. The goal of the original study was to test whether the construct of relationship commitment causally affected the construct of forgiveness. The fact that the priming manipulation did not affect subjective commitment in the RRR suggests that the effects of priming on the outcome measures in the RRR are not testing the causal effect of subjective commitment on forgiveness. Consequently, the effect of the priming manipulation on the forgiveness measures does not test the same theoretical prediction in the RRR as it did in the original study. This logic extends to the analyses in which self-reported commitment is the predictor variable because that variable had been experimentally altered in the original study, but not in the RRR.

The manipulation failure in the RRR does have a silver lining, however: It allows for a relatively clean test of the hypothesis that subjective commitment (the intended manipulation check) is positively correlated with forgiveness in a large-scale preregistered study with no publication bias. A recent meta-analysis (Fehr, Gelfand, & Nag, 2010) reported an association between commitment and forgiveness of $r = ~.20$. The meta-analytic correlations reported in the RRR are consistent with such a relationship. To produce a more direct comparison to Fehr et al.'s (2010) meta-analytic effect size, I created a composite forgiveness measure from the RRR data by averaging across voice, loyalty, exit, and neglect—after reverse coding exit and neglect. A multilevel modeling analysis accounting for the nesting of participants within lab (see Appendix for SAS code) revealed a similar association between commitment and forgiveness, $\beta = .24$, $t(2266) = 11.82$, $p < .001$. Given the failure of the experimental manipulation to affect subjective commitment, the RRR results can provide no definitive evidence for or against a causal relationship between commitment and forgiveness, but this correlational analysis provides additional evidence for a robust association between relationship commitment and various relationship maintenance mechanisms, including forgiveness (Rusbult, Olsen, Davis, & Hannon, 2001).

Conclusion

Conducting high-quality, laboratory-based direct replications is a painstaking, resource-intensive enterprise, but doing so is the gold standard for assessing the robustness of a given finding. Based on the results of this RRR, the experimental findings from Study 1 of Finkel et al. (2002) are not robust. Consequently, I will no longer cite that study as providing causal evidence for the effect of commitment on forgiveness, and I discourage others from doing so. Although I am surprised by the failure of the manipulation check and disappointed that the results of the RRR did not confirm the causal effects my colleagues and I originally reported, I deeply respect the process. It has left us with major unanswered questions, but it has also left us wiser than we were before.

Appendix

SAS code for the analysis predicting forgiveness from self-reported commitment.

```
PROC MIXED METHOD=REML COVTEST;
    CLASS Lab;
    MODEL ForgZ=ComMCheckZ / S;
    RANDOM Int / TYPE=UN SUB=lab;
RUN;
```

Note. "Lab" is a categorical variable (coded 1–16) representing the lab in which the participants were run. "ForgZ" is a forgiveness measure calculated by averaging Voice Forgiveness, Loyalty Forgiveness, Exit Forgiveness (reverse-scored), and Neglect Forgiveness (reverse-scored), and then standardizing. "ComMCheckZ" is a standardized version of the commitment manipulation check scale score. An auxiliary analysis in which slopes were also permitted to vary randomly (by adding "ComMCheckZ" to the RANDOM statement) revealed results that were virtually identical to those reported in the main text, albeit with degrees of freedom values linked to the number of participating labs rather than to the number of participants: $\beta = .24$, $t(15) = 11.44$, $p < .001$.

Acknowledgments

Declaration of Conflicting Interests

Notes

1. Our primary analyses controlled for participant sex and its interaction with the commitment manipulation. The inclusion of covariates without also reporting the "uncontrolled" results has been flagged as a potential indicator of p hacking (Simmons, Nelson, & Simonsohn, 2011). I do not recall why we included sex in the final statistical models (perhaps we did so to placate reviewers), but all hypothesis tests yield identical conclusions when omitting sex from the models.

References

Cheung, I., Campbell, L., LeBel, E., . . . Yong, J. C. (2016). Registered Replication Report: Study 1 from Finkel, Rusbult, Kumashiro, & Hannon (2002). *Perspectives on Psychological Science*, *11*, 750–764.

Eastwick, P. W., & Finkel, E. J. (2008). Sex differences in mate preferences revisited: Do people know what they initially desire in a romantic partner? *Journal of Personality and Social Psychology*, *94*, 245–264.

Fehr, R., Gelfand, M. J., & Nag, M. (2010). The road to forgiveness: A meta-analytic synthesis of its situational and dispositional correlates. *Psychological Bulletin*, *136*, 894–914.

Finkel, E. J., Rusbult, C. E., Kumashiro, M., & Hannon, P. A. (2002). Dealing with betrayal in close relationships: Does commitment promote forgiveness? *Journal of Personality and Social Psychology*, *82*, 956–974.

Hui, C. M., Finkel, E. J., Fitzsimons, G. M., Kumashiro, M., & Hofmann, W. (2014). The Manhattan effect: When relationship commitment fails to promote support for partner interests. *Journal of Personality and Social Psychology*, *106*, 546–570. doi:10.1037/a0035493

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). doi:10.1126/science.aac4716

Rusbult, C. E., Olsen, N., Davis, J. L., & Hannon, P. (2001). Commitment and relationship maintenance mechanisms. In J. H. Harvey & A. Wenzel (Eds.), *Close romantic relationships: Maintenance and enhancement* (pp. 87–113). Mahwah, NJ: Erlbaum.

Selterman, D., Chagnon, E., & Mackinnon, S. (2015). Do men and women exhibit different preferences for mates? A replication of Eastwick & Finkel (2008). *SAGE Open*, *5*(3). doi:10.1177/2158244015605160

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366.