

Many Labs 5: Registered Replication of Albarracín et al. (2008), Experiment 5



Christopher R. Chartier¹, **Jack D. Arnal²**, **Holly Arrow³**,
Nicholas G. Bloxson¹, **Diane B. V. Bonfiglio¹**,
Claudia C. Brumbaugh⁴, **Katherine S. Corker⁵**,
Charles R. Ebersole⁶, **Alexander Garinther³**, **Steffen R. Giessner⁷**,
Sean Hughes⁸, **Michael Inzlicht⁹**, **Hause Lin⁹**, **Brett Mercier¹⁰**,
Mitchell Metzger¹, **Derek Rangel¹**, **Blair Saunders¹¹**,
Kathleen Schmidt¹², **Daniel Storage¹³**, and **Carly Tocco⁴**

¹Department of Psychology, Ashland University; ²Psychology Department, McDaniel College;
³Department of Psychology, University of Oregon; ⁴Department of Psychology, Queens College, City
University of New York; ⁵Department of Psychology, Grand Valley State University; ⁶Department
of Psychology, University of Virginia; ⁷Rotterdam School of Management, Erasmus University;
⁸Department of Experimental Clinical and Health Psychology, Ghent University; ⁹Department of
Psychology, University of Toronto; ¹⁰Department of Psychological Science, University of California,
Irvine; ¹¹School of Social Sciences, University of Dundee; ¹²School of Psychological and Behavioral
Sciences, Southern Illinois University Carbondale; and ¹³Department of Psychology, University of Denver

Abstract

In Experiment 5 of Albarracín et al. (2008), participants primed with words associated with action performed better on a subsequent cognitive task than did participants primed with words associated with inaction. A direct replication attempt by Frank, Kim, and Lee (2016) as part of the Reproducibility Project: Psychology (RP:P) failed to find evidence for this effect. In this article, we discuss several potential explanations for these discrepant findings: the source of participants (Amazon's Mechanical Turk vs. traditional undergraduate-student pool), the setting of participation (online vs. in lab), and the possible moderating role of affect. We tested Albarracín et al.'s original hypothesis in two new samples: For the first sample, we followed the protocol developed by Frank et al. and recruited participants via Amazon's Mechanical Turk ($n = 580$). For the second sample, we used a revised protocol incorporating feedback from the original authors and recruited participants from eight universities ($n = 884$). We did not detect moderation by protocol; patterns in the revised protocol resembled those in our implementation of the RP:P protocol, but the estimate of the focal effect size was smaller than that found originally by Albarracín et al. and larger than that found in Frank et al.'s replication attempt. We discuss these findings and possible explanations.

Keywords

replication, priming, open data, open materials, preregistered

Received 7/4/18; Revision accepted 6/24/20

Albarracín et al. (2008) conducted a series of experiments examining whether the activation of general action versus inaction goals influenced subsequent levels of motor and cognitive output. In their fifth experiment, undergraduate students completed a sentence-unscrambling task that included words related to action or inaction before completing a series of SAT-style test questions assessing their verbal and math ability. Results were consistent with the authors' hypotheses: Participants primed

with action-related words demonstrated more cognitive output by solving more problems ($M = 12.83$, $SD = 1.86$) than their counterparts who were primed with

Corresponding Author:

Christopher R. Chartier, 146, SCOE, Ashland University, 401 College Ave.,
Ashland, OH 44805
E-mail: cchartie@ashland.edu

inaction-related words ($M = 10.78$, $SD = 3.15$), $F(1, 34) = 5.68$, $p = .02$.

Frank, Kim, and Lee (2016) attempted to replicate this finding as part of the Reproducibility Project: Psychology (RP:P; Open Science Collaboration, 2015). However, they found no evidence that the priming manipulation influenced subsequent task performance in a sample of participants recruited via Amazon's Mechanical Turk (MTurk), $F(1, 86) = 0.08$, $p = .778$. The average number of correctly completed problems was not significantly different between the action-priming condition ($M = 11.33$, $SD = 3.84$) and the inaction-priming condition ($M = 11.57$, $SD = 3.86$).

Several explanations could account for these discrepant results. Two noteworthy differences between the original study and replication effort are the source of the participants and the experimental setting. In Albarracín et al.'s (2008) original study, undergraduate students completed the procedure in the laboratory, whereas in Frank et al.'s (2016) replication, MTurk workers were recruited online and completed the study in whatever setting they chose. Undergraduates may be more susceptible to influence than are nonstudents (Sears, 1986) and therefore more likely to be affected by the priming manipulation than MTurk workers are. Also, Frank et al.'s sample had a mean age of 31.25 years and had relatively diverse educational backgrounds, ranging from no high school education to completed advanced degrees. Further, subtle manipulations, such as the scrambled-sentence task, may have decreased efficacy in environments without tight control because of increased unsystematic variability. Both the source of the participants and the experimental setting may have strengthened the effect in the original study or weakened it in the replication study.

The original authors also offered another explanation for the failed replication based on research published after their original study. They suggested that participants' affect might moderate the impact of action/ inaction priming on cognitive output, as was found in a follow-up study (Albarracín & Hart, 2011). In that study, in addition to being primed with action- or inaction-related words, participants were asked to write a letter to a friend about a personal experience that made them very happy (positive-affect condition), an experience that made them very frustrated and angry (negative-affect condition), or a typical day in their life (neutral-affect condition). Participants then completed the same set of SAT-type problems as in Albarracín et al. (2008, Experiment 5). In the positive-affect and neutral-affect conditions, participants primed with action-related words solved more problems than did those primed with inaction-related words. However, this effect was significantly reversed in the negative-affect condition. Participants in Frank et al.'s (2016) replication may have

varied more in their affective states than those in the original study, and this could have reduced the effect size. Indeed, the increased variance of performance on the cognitive task in the replication relative to the original study is consistent with such a moderation effect.

We set out to investigate whether these potential explanations account for the differences between the findings of Albarracín et al.'s (2008) Experiment 5 and the replication conducted by Frank et al. (2016). We collected two new sets of data. First, using the protocol developed by Frank et al. (2016) for the RP:P, we collected data from a sample of MTurk workers. Second, using a revised protocol (see the Method section), we collected data from samples of undergraduate participants at eight universities (see the Method section). In both protocols, we also included the same two-item affect measure used by Albarracín and Hart (2011). Thus, we were able to directly examine whether the discrepant findings were due to changes in sample and setting, variations in affect, or both. Our affect results should be interpreted with caution, however, because the affect measure followed the cognitive task. If affect changed because of the intervening time or the intervening task, the affect measure may not have accurately captured the affective state of our participants during the cognitive task.

Disclosures

Preregistration

Our design and confirmatory analyses were preregistered on the Open Science Framework (<https://osf.io/6qn4t>).

Data, materials, and online resources

All materials, data, and code are available on the Open Science Framework (<https://osf.io/6qn4t>).

Reporting

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Ethical approval

Data were collected in accordance with the Declaration of Helsinki guidelines, and all sites had approval from their institutional review board prior to collecting data.

Method

Participants

For the revised protocol, participants were recruited from eight universities in Europe and North America.

Table 1. Location and Sample Size for Each Study Site

Institution	Country	Total sample (<i>n</i>)	Participants excluded (<i>n</i>)	Participants included in analysis (<i>n</i>)
Ashland University	United States	81	0	81
Erasmus University	Netherlands	161	15	146
Ghent University	Belgium	89	3	86
Queens College	United States	100	0	100
University of Illinois	United States	86	5	81
University of Oregon	United States	178	4	174
University of Toronto	Canada	123	2	121
Wesleyan University	United States	97	2	95
Mechanical Turk	International	581	1	580
Overall	International	1,496	32	1,464

Undergraduates in introductory-psychology courses at these institutions participated in exchange for course credit. We replicated Frank et al.'s (2016) power analysis, using the effect size for the focal effect from the original study ($d = 0.79$) and $\alpha = .05$. This analysis indicated that we would need 86 participants to achieve 95% power at each site, and we aimed to collect data from at least 86 participants from at least three collection sites. The final sample included 884 participants. For the RP:P protocol, we collected data from an MTurk sample. MTurk participation was limited to the United States, as in the RP:P study. Given the ease and convenience of data collection with MTurk samples relative to in-lab samples, and to maximize power to detect the focal replication effect, we aimed to collect data from at least 500 MTurk participants. Our final MTurk sample consisted of 580 participants. Table 1 provides a summary of sample sizes by location.

Materials

The original study materials provided by the original authors were computerized and presented using Qualtrics (<https://www.qualtrics.com>). The wording and page order followed the materials provided exactly.

Priming task. A scrambled-sentence task was used to prime action or inaction. Each of the 12 items consisted of five words (e.g., “shoes,” “feet,” “green,” “cover,” “your”), four of which participants could use to create a sentence (e.g., “Shoes cover your feet.”). Depending on the condition, 8 of the items contained action-related words (e.g., “book,” “action,” “is,” “the,” “fictional”) or inaction-related words (e.g., “the,” “tells,” “inaction,” “watch,” “time”). The other 4 items were filler items.¹

Dependent measures

SAT questions. Twenty-one questions similar to those found on the SAT² were used to measure cognitive output.

These questions assessed verbal ability (e.g., antonyms, sentence completion, and analogies) and quantitative ability (e.g., word problems and algebraic equations). The number of correct responses was the a measure of cognitive output.

Affect measures. To test whether affect accounted for differences in cognitive output between conditions, we assessed affect using two exploratory items. Participants reported the extent to which they were happy and angry using scales from 1 (*not at all*) to 9 (*extremely*), as in Albarracín and Hart (2011).

Manipulation check. The word-fragment-completion task used in our validation study (reported at the end of the Method section) also served as a manipulation check in the revised protocol. Participants were asked to finish 40 word stems that could be completed with action-related, neutral, or inaction-related words. This manipulation check was not part of Albarracín et al.'s (2008) original Experiment 5.

Procedure

For our revised protocol, we followed the original experimental procedure of Albarracín et al.'s (2008) Experiment 5 with minor variations. Participants completed the study at individual workstations in a traditional lab setting. They were provided with a pencil and scrap paper that they could use to work through the various problems. Participants were first informed that the study was a pilot test of an instrument designed to assess how people form sentences. They were then given unlimited time to complete the computer-based priming task. Next, they were asked to complete the 21 SAT-type questions assessing their verbal and quantitative abilities. Following this task, participants indicated their current levels of happiness and anger and then performed the word-fragment-completion task (the

manipulation check). Finally, they were debriefed and thanked.

The procedure in the RP:P protocol was similar but had several key differences. Participants completed the study on MTurk, were not provided with scrap paper or instructed to use any, and did not complete the manipulation check.

Validation study

A separate validation study assessed the efficacy of the priming manipulation in a sample of 428 undergraduates and 297 MTurk participants. First, these participants completed the scrambled-sentence task used in the original study and the RP:P replication attempt. An error resulted in an inaction-related word (*pause*) being presented to participants in the neutral-priming condition. Thus, the items in this condition had 11 neutral words and 1 inaction-related word but should have had 12 neutral words. Direct comparisons between the neutral-priming condition and either the action-priming or the inaction-priming condition should be made with caution, although comparisons between the action-priming and inaction-priming conditions are unaffected by this error. Activation of the concepts of action and inaction was measured using a word-fragment-completion task. The number of action-word-derived stems completed as a word associated with action (as opposed to a neutral, inaction-related, or ambiguous word) and the number of inaction-word-derived stems completed as a word associated with inaction (as opposed to a neutral, action-related, or ambiguous word) were counted for each participant.

A three-condition one-way analysis of variance (ANOVA) found a significant effect of priming condition on action-related completions, $F(2, 722) = 3.76, p = .024, r = .10$. Participants in the action-priming condition generated the most action-related completions ($M = 9.37, SD = 2.87$). However, Tukey-corrected post hoc pairwise tests (Tukey's honestly significant differences) showed that the number of action-related completions differed significantly only between the action-priming condition and the neutral-priming condition ($M = 8.68, SD = 2.87, p = .021$). There was no significant difference between the action-priming condition and the inaction-priming condition ($M = 8.88, SD = 2.84, p = .142$), or between the neutral- and inaction-priming conditions, $p = .723$.

A three-condition one-way ANOVA also found a significant effect of priming condition on inaction-related completions, $F(2, 722) = 19.55, p < .001, r = .23$. Participants in the inaction-priming condition generated the most inaction-related completions ($M = 6.63, SD = 2.10$). Tukey-corrected post hoc pairwise tests showed that the number of inaction-related completions differed

significantly between the inaction-priming condition and the neutral-priming condition ($M = 5.73, SD = 2.00, p < .001$), as well as between the inaction-priming condition and the action-priming condition ($M = 5.51, SD = 2.11, p < .001$), but did not differ significantly between the neutral- and action-priming conditions, $p = .472$. Taken together, these findings suggest that the scrambled-sentence task was generally—although not universally—effective in eliciting the desired priming effects.

Results

The original authors described no data-exclusion rules. We excluded only participants who did not complete either the priming or the SAT-problem task ($n = 32$). This left a final sample of 1,464 participants. To ensure that the SAT problems were within a reasonable difficulty range, relative to that in the original study, we established an a priori “acceptable” range. The upper bound was set as 1 *SD* above the mean for the action-priming condition in the original study, that is, 14.69 correct items out of 21. The lower bound was set as 1 *SD* below the mean for the inaction-priming condition in the original study, that is, 7.63 correct items out of 21. None of the sites' averages fell outside of this range, and as a result, data from all the sites were analyzed (see Fig. 1).

Confirmatory analyses

Our confirmatory analyses of the performance data were similar to the original analysis (i.e., we used simple ANOVAs comparing the number of correctly solved problems in the two experimental conditions).

Revised protocol. To account for multisite data collection, we fitted two-level multilevel regression models with participants nested within site. The first model predicted the primary outcome (number of SAT problems answered correctly) from just a fixed intercept, a random intercept, and a random slope for experimental condition nested within site. This model failed to converge, so we removed the random slope for condition. We then added the fixed effect of experimental condition to the model. Adding this fixed effect improved the model, $\chi^2(1, N = 884) = 5.51, p = .019, \text{pseudo-}R^2 = .008$ (see Fig. 1 for individual sites' results, converted to *r* values.). This analysis revealed that participants primed with action solved more problems ($M = 10.20, SD = 3.40$) than did those primed with inaction ($M = 9.85, SD = 3.41, d = 0.10, 95\% \text{ confidence interval (CI)} = [-0.03, 0.24]$). This effect-size estimate was smaller than that obtained by Albarracín et al. (2008), $d = 0.79, 95\% \text{ CI} = [0.11, 1.47]$, and larger than that obtained by Frank et al. (2016), $d = -0.06, 95\%$

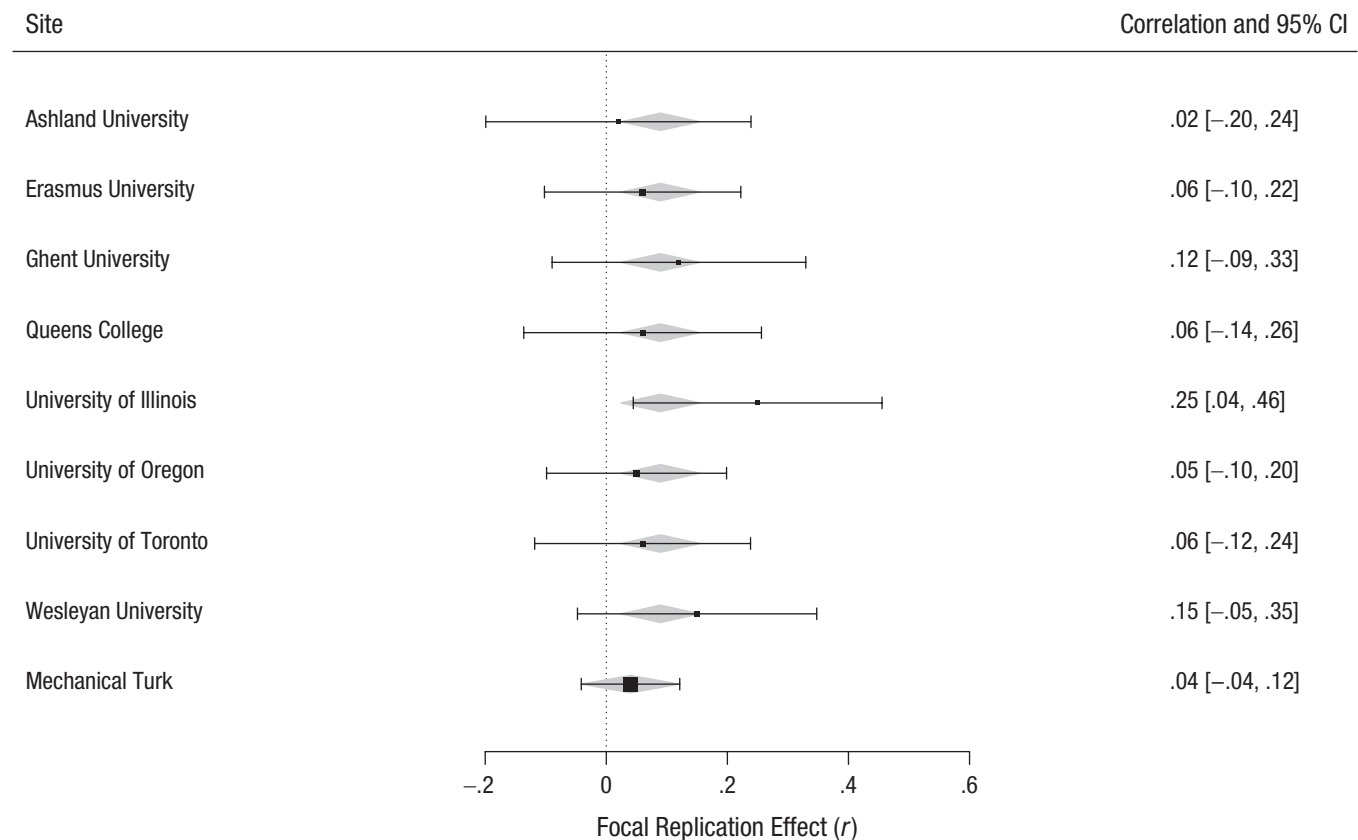


Fig. 1. Replication results by collection site. Each row summarizes the effect size (presented as r values) for the focal replication effect at the indicated site. Positive effect sizes indicate effects consistent with the direction of the effect in the original study. The sizes of the plotted points are an inverse function of the model weights, and the error bars indicate the 95% confidence intervals (CIs). The gray diamonds represent the meta-analytic aggregate effect-size estimates within the Reproducibility Project: Psychology protocol (Mechanical Turk sample) and the revised protocol (all other samples).

CI = [-0.48, 0.36], as its 95% confidence interval did not contain the estimates from those studies.

RP:P protocol. For our comparison MTurk sample, the ANOVA comparing the number of correctly solved problems in the two experimental conditions indicated that participants primed with action did not solve more problems ($M = 10.96$, $SD = 3.93$) than did those primed with inaction ($M = 10.65$, $SD = 3.72$), $F(1, 578) = 0.92$, $p = .337$, $d = 0.08$, 95% CI = [-0.08, 0.24] (see Fig. 1). This effect-size estimate was similar to that of Frank et al. (2016), as its 95% confidence interval contained the estimate from that study. In addition, this effect-size estimate was similar to that of the revised protocol, as its 95% confidence interval contained the estimate from the revised protocol. It is important to note that the overall pattern of results and the magnitude of the focal effect were quite similar in the revised protocol and the RP:P protocol. Although the effect met the traditional criterion for statistical significance in the revised protocol but not the RP:P protocol, such a difference in significance does not necessarily signal a meaningful difference between the protocols.³

Exploratory analyses

To more closely mirror the original study, we also conducted the main analyses with only U.S. sites included. We made no a priori predictions regarding differences in results between the analyses with all sites included (reported in the previous section) and the analyses with only U.S. sites included (reported in this section). We repeated the process of fitting two-level multilevel regression models with participants nested within site but this time include data from only those sites within the United States. The first model predicted the primary outcome (number of SAT problems answered correctly) from just a fixed intercept, a random intercept, and a random slope for experimental condition nested within site. This model failed to converge, so we removed the random slope for condition. We then added the fixed effect of experimental condition to the model, retaining just the random intercept, and the effect of experimental condition remained nonsignificant. Adding the effect of condition did not improve the model, $\chi^2(1, N = 531) = 3.20$, $p = .074$, pseudo- $R^2 = .007$. Overall, participants

primed with action did not solve more problems ($M = 10.55$, $SD = 3.56$) compared with those primed with inaction ($M = 10.15$, $SD = 3.44$), $d = 0.11$, 95% CI = $[-0.06, 0.28]$.

For our manipulation check, we used a one-tailed independent-samples t test to examine possible between-conditions differences in the number of action-word-derived word stems completed as a word associated with action (as opposed to a neutral, inaction-related, or ambiguous word) conditions. Participants in the action-priming condition completed these stems with significantly more words associated with action ($M = 8.38$, $SD = 2.98$) than did participants in the inaction-priming condition ($M = 7.94$, $SD = 2.81$), $t(878) = 2.23$, $p = .013$, $d = 0.15$, 95% CI = $[0.02, 0.28]$. Notably, however, the effect size was small.

Similarly, a one-tailed independent-samples t test was used to evaluate possible between-conditions differences in the number of inaction-word-derived word stems completed with a word associated with inaction. Participants in the inaction-priming condition completed these stems with significantly more words associated with inaction ($M = 6.79$, $SD = 2.26$) than did participants in the action-priming condition ($M = 6.05$, $SD = 2.16$), $t(880) = -4.94$, $p < .001$, $d = -0.33$, 95% CI = $[-0.47, -0.20]$. These results suggest that the action- and inaction-priming manipulations were successful, though the effect on inaction-related completions was somewhat larger than the effect on action-related completions.

In our final exploratory analyses, we examined affect as a possible moderator of the focal replication effect in both the revised protocol and the RP:P protocol. For the revised protocol, because of an unidentified coding error in establishing the data-collection links in Qualtrics, the affect data were successfully recorded for the revised protocol at only four sites: Wesleyan University, the University of Illinois, the University of Oregon, and the University of Toronto. Adding the effect of happiness improved the model, $\chi^2(1, N = 469) = 5.48$, $p = .019$, pseudo- $R^2 = .009$. Adding the effect of anger did not improve the model, $\chi^2(1, N = 469) = 1.33$, $p = .249$, pseudo- $R^2 = .002$. To supplement these results, we fitted a third model that included and compared both the interaction between priming condition and happiness and the interaction between priming condition and anger, as well as the main effects from the two models just reported. We found that neither the condition-happiness interaction ($p = .544$) nor the condition-anger interaction ($p = .660$) was significant. Together, these results suggest that neither happiness nor anger moderated the effect of action versus inaction priming on cognitive output. These results do not provide evidence that differences in participants' affect explain the

discrepant results of the original study and the RP:P replication.

In our analysis of the data for the RP:P protocol, a model with the effects of priming condition and happiness revealed a nonsignificant effect of happiness on cognitive output, $F(1, 574) = 1.73$, $p = .189$, $r = .05$, whereas a model with the effects of priming condition and anger revealed a significant effect of anger on cognitive output, $F(1, 574) = 10.11$, $p = .002$, $r = .13$. To supplement these results, we fitted a third model that included and compared both the interaction between priming condition and happiness and the interaction between priming condition and anger, as well as the main effects from the two models just reported. We found that neither the condition-happiness interaction ($p = .836$) nor the condition-anger interaction ($p = .860$) was significant. These results again suggest that neither happiness nor anger moderated the effect of action versus inaction priming on cognitive output. They do not provide evidence that differences in participants' affect explain the discrepant results of the original study and the RP:P replication.

Discussion

Our goal in the current study was to resolve the discrepancy between the findings of Albarracín et al.'s (2008) Experiment 5 and Frank et al.'s (2016) replication attempt concerning the effect of action and inaction priming on cognitive output. We examined three potential moderating factors: participant type, setting, and affect. Results from a large international sample suggest that the effect originally described by Albarracín et al. is detectable in a large sample, but that the effect size is likely smaller than originally reported.

There appeared to be no meaningful difference in this estimated effect size between our U.S. and non-U.S. samples. It thus seems unlikely that the changes between the original study and Frank et al.'s (2016) replication attempt—specifically, the source of participants (Amazon's MTurk vs. traditional undergraduate student pool) and the setting of participation (online vs. in lab)—led to meaningful differences in the estimated effect size. Rather, our large and international study suggests that the best effect-size estimate lies between the estimates from the original study and the RP:P replication attempt. This conclusion is bolstered by the fact that the manipulation check, as well as the validation study's results, suggest that the scrambled-sentence primes were effective in eliciting the desired cognitive states. However, the manipulation check in the main study is limited by the fact that it did not immediately follow the priming task, and its results are therefore difficult to interpret, although it did immediately follow

the priming task in the validation study. Finally, the focal effect did not appear to be moderated by participants' affect.

As is the case with any specific investigation of a psychological phenomenon of interest, our results should be interpreted with caution because they are necessarily constrained by the specific protocols used and samples tested. The present data do, however, improve understanding of the target effect because they were drawn from a broader and larger sample than used in the previous studies and also were obtained using multiple implementations of the protocol.

Conclusion

Albarracín et al. (2008) found that participants primed with concepts related to action demonstrated more cognitive output on a subsequent cognitive test than did those primed with concepts related to inaction. Frank et al. (2016) attempted to replicate this effect and did not find supportive evidence. The current crowdsourced replication project was an attempt to resolve the discrepancy between these two sets of results. Overall, we found that the best estimated effect size for increased cognitive output following action primes versus inaction primes lies between the effect-size estimates of Albarracín et al. and Frank et al. We hope that productive communication and collaboration between original authors and replication teams, as experienced in the present study, will lead to a better understanding of specific psychological phenomena, as well as to a more robust and productive psychological science.

Transparency

Action Editor: Daniel J. Simons

Editor: Daniel J. Simons

Author Contributions

We categorized contributions using the Contributor Roles Taxonomy (CRediT; <https://casrai.org/credit>). C. R. Chartier contributed to curating the data, administering the project, supervising the investigation, writing the original draft of the manuscript, and reviewing and editing the final manuscript. D. B. V. Bonfiglio, D. Rangel, N. G. Bloxson, and S. R. Giessner contributed to data curation. K. S. Corker contributed to designing the methodology and reviewing and editing the final manuscript. S. Hughes, A. Garinther, and K. Schmidt contributed to investigation, writing the original draft, and reviewing and editing the final manuscript. M. Inzlicht contributed to investigation, supervision, and reviewing and editing the final manuscript. H. Lin, D. Storage, C. Tocco, and H. Arrow contributed to investigation and reviewing and editing the final manuscript. M. Metzger contributed to supervising the investigation and reviewing and editing the final manuscript. B. Mercier contributed to investigation. B. Saunders contributed to investigation, preparing the documents for the institutional

review board, supervising the investigation, and editing the manuscript. C. C. Brumbaugh, J. D. Arnal, and C. R. Ebersole contributed to writing the original draft and reviewing and editing the final manuscript.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This project was supported by a grant from the Association for Psychological Science and from Arnold Ventures.

Open Practices

Open Data: <https://osf.io/6qn4t>

Open Materials: <https://osf.io/6qn4t>


Preregistration: <https://osf.io/6qn4t>


All data and materials have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/6qn4t>. The design and analysis plans were preregistered at the Open Science Framework and can be accessed at <https://osf.io/6qn4t>. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/2515245920945963>. This article has received badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.





ORCID iDs


Christopher R. Chartier  <https://orcid.org/0000-0002-4568-4827>

Katherine S. Corker  <https://orcid.org/0000-0002-7971-1678>

Charles R. Ebersole  <https://orcid.org/0000-0002-8607-2579>

Michael Inzlicht  <https://orcid.org/0000-0001-9297-6497>

Hause Lin  <https://orcid.org/0000-0003-4590-7039>

Kathleen Schmidt  <https://orcid.org/0000-0002-9946-5953>

Acknowledgments

We thank David C. Cicero (University of Hawaii at Manoa) for assisting with data collection for the validation study.

Notes

1. One of the 12 items was not solvable, to mimic a mistake in the original study.
2. In what follows, we refer to these as "SAT questions," but do not mean to imply that they were taken directly from SAT tests.
3. The two preceding sentences are an addition subsequent to the provisionally accepted manuscript. They are intended to avoid confusion about the implications of the difference in significance between the two protocols.

References

- Albarracín, D., Handley, I. M., Noguchi, K., McCulloch, K. C., Li, H., Leeper, J., . . . Hart, W. P. (2008). Increasing and decreasing motor and cognitive output: A model of general action and inaction goals. *Journal of Personality and Social Psychology*, 95, 510–523.

- Albarracín, D., & Hart, W. (2011). Positive mood + action = negative mood + inaction: Effects of general action and inaction concepts on decisions and performance as a function of affect. *Emotion*, 11, 951–957. doi:10.1037/a0024130
- Frank, M. C., Kim, S. K., & Lee, K. J. (2016). *Replication of Albarracín et al. (2008, JPSP)*. Retrieved from <https://osf.io/vy1bc/>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349, Article aac4716. doi:10.1126/science.aac4716
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, 51, 515–530.