**Peer Review and Communication History**

**MS Title**: Leading Us unto Temptation? No Evidence for an Asymmetry in Automatic Associations between Goals and Temptations

**Author Names**: Zoë Francis, Aravinth Jebanesan, Michael Inzlicht

**Submitted:** Jun 7, 2021

**Editor First Decision**: Revise & Resubmit

Jul 24, 2021

Dear Dr. Francis,

I have now received reviews of your manuscript, "Leading Us Unto Temptation? No Evidence for an Asymmetry in Automatic Associations Between Goals and Temptations” from two researchers who have published original work on counteractive control theory. I also independently read the manuscript before consulting these reviews, and once again just after reading them and reading Fishbach et al. 2003.

The reviewers were quite enthusiastic about the manuscript. So was I. I was impressed by how closely you hewed to the original paradigm, while also making thoughtful amendments to it when you believed it would increase evidential value. I enthusiastically encourage you to submit a revised version for further consideration.

The reviewers did an outstanding job in their reviews, and I thank them for drawing attention to issues that would have eluded my attention. Rather than rehearse what the reviewers articulated so well, below I list some points that they did not make. In your resubmission, please include a document with a point-by-point response to both the points I list here and the reviewers’ comments, outlining each change made in your manuscript or providing a suitable rebuttal.

My reactions:

* Do you have syntax for how you got from the raw data file to the critical file csv that you analyzed? If so, can you upload it? I won’t have the software knowledge to verify that no mistakes in data cleaning were made, but perhaps future readers will.
* Upon running mod1 (and all of the other models), I get a convergence failure: “Warning message:In checkConv(attr(opt, “derivs”), optpar,ctrl=controlcheckConv, :Model failed to converge with max|grad| = 0.00204309 (tol = 0.002, component 1).” Did you get this warning message as well? If so, it should be reported, and an alternative estimation strategy should be pursued and reported as unplanned. I reran mod1 using the allFit function, which tries a bunch of different optimizers. The optimizer “bobyqa” is the only one that converged without issues.

Of course, possibly this is a problem with my computer, the version of R I used (please report the version you used), or because there was some intermediate step I needed to take before running the models that is not documented. (Speaking of which, it would be helpful if there was a README file on OSF so that readers know exactly what each file is, what each variable name means, etc.). If you aren’t running into these issues, please double-check that your R settings aren’t suppressing warning messages, and then re-run your analyses on a computer other than the original computer you used and maybe a different version of R as well to see if software specifications could be the issue. In each model I got slightly different estimates than what you report, but perhaps this is because I got convergence errors and you did not. I am not the savviest at debugging software problems, so these issues could definitely be my fault, but your readers might not be R savvy either so it is important that the code will run without issue for them.

* An additional robustness check to consider would be to exclude no data whatsoever from analysis. All exclusions have the potential of inducing selection bias, violating random assignment. To handle extreme values, a robust estimator could be employed (though I am not an expert on handling influential data points so take this recommendation with a grain of salt), <https://search.r-project.org/CRAN/refmans/robustlmm/html/rlmer.html>.
* I know what summary() does, but not summaryh(), which did not run for me. Is this a typo?
* Where is the R code replicating the original Fishbach et al. (2003) analytic strategy?
* Page 7: “A sensitivity power analysis conducted with simr (Green & Macleod, 2016) finds that our sample had over 95% power to detect an interaction effect of 30ms (r = .06); this effect size is one quarter of the asymmetry effect size found in Study 1 of Fishbach et al., 2003 and approximately half the asymmetry effect size from Study 3. Additionally, our sample had over 80% power to detect a correlation between the degree of priming asymmetry and self-regulatory success of r = .19 (from Study 3 of Fishbach et al., 2003).” It not clear to me how you extracted these figures from the simr code you posted. Can you spell it out in the syntax file?
* Fishbach et al. (2003) used one-tailed tests in their experiments, at time reporting p-values that would not have met the conventional two-tailed alpha of .05 (e.g., study 2, p. 300). Support for the moderating effect of self-regulation strength appears to be marginal even when using (what I think are) two-tailed p-values (Study 3, p. 302). You might consider noting this as a source of uncertainty in whether the original findings really provided strong support in the first place, even by the crude standards of p <.05.
* Fishbach et al. 2003, p.302: “These measures of perceived success in pursuing academic goals were designed to capture the ease/difficulty of overcoming temptation (in addition to general academic ability). The items were highly correlated (r = .83, p <.001) and were therefore combined into a unitary index of perceived self-regulatory success.” You report a smaller correlation between these items, was it significantly smaller? Also, one might argue that this measure much more strongly taps cognitive ability than self-regulation strength. The weakness of this measure might be seen as an additional justification for the inclusion of the trait self-control measure you used.
* What did exploratory analysis of the funnel debrief reveal?
* You use an inverse gaussian model which is mentioned several times in the pre-registration, save that the “Sample Size” section of the pre-registration calls for an ex-gaussian model. Please clarify if this a typo, a holdover from an earlier plan, or something else.
* As an aside, I don’t think it matters much here but in future papers you may want to consider not grand-mean centering (see Antonakis, J., Bastardoz, N., & Rönkkö, M. (2021). On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. Organizational Research Methods, 24(2), 443-483.)
* Page 22: “Although our student body is ethnically and culturally diverse, the sample is not representative of the population at large, and these results cannot necessarily be generalized to people from other age groups, cultures, or educational backgrounds.” I appreciate you mentioning this constraint on generalizability. However, it could also be worth mentioning that this is cold comfort for Fishbach et al. (2003) unless there is some principled explanation for why the results would not hold in your sampled population. (Maybe their average self-regulation abilities are too low?) Given that they tested the theory on a similar population, I don’t see much room for an appeal to selection bias.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at [editorialoffice@collabra.org](mailto:editorialoffice@collabra.org).

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,  
William H.B. McAuliffe

# Reviewer 1

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

This manuscript reports the results of an attempted replication of a seminal finding re: a key tenet of counteractive control theory: namely, the asymmetric cognitive activation of goals and temptation-related concepts upon priming. The manuscript is very well written, the study clearly and thoroughly reported. The study is well-powered, methods are sound and well-placed to robustly test this aspect of a widely cited theory (and extend the original study in an important way, with the addition of irrelevant prime words and increased statistical power). The manuscript is a great fit for Collabra, and authors provide links to pre-registered study protocol, open data and code. I have only several very minor suggestions for the authors.

* Page 5 – a brief definition/description of construal theory/abstract construal would be useful when first mentioned. Shortly after on page 6, reporting the direction of the moderation by temptation strength would be useful.
* Related to these moderators, it would be useful to see treatment of the distinction/overlap and potential relationship between state-level moderators and stable trait-like moderators in the introduction. E.g., briefly discuss the treatment / testing of these moderators in past studies (experimental conditions - manipulations of state-based constructs versus testing individual differences) and justify the use of individual difference measures. I note that this is discussed briefly in the discussion section, but including a brief treatment of this issue in introduction would enhance understanding of the existing research / situate the findings re: moderators.
* Additional details of pilot study are needed – e.g., how many people recruited and how/where recruited?

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

# Reviewer 2

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

Leading Us Unto Temptation? No Evidence for an Asymmetry in Automatic Associations Between Goals and Temptations

Review

This paper presents a carefully executed direct replication of a key study (Study 3) in a highly influential paper on the cognitive foundations of self-control (Fishbach et al., 2003); the authors argue that there are surprisingly few published replications of the central, counterintuitive result—asymmetric automatic associations between goals and temptations—thus making a compelling case for their paper.

The paper comes across as clear, complete, and carefully crafted. It is methodologically sound, and the analyses are thorough, extended in depth and rigor beyond those of the target paper. The exposition is sharp, and the argumentation effective.

This paper makes a very valuable contribution to the literature, and I would recommend that it gets published, subject to revision. Specific comments and requests follow below:

1. The authors argue at multiple points that counteractive control theory predicts the targeted effect, asymmetry in associations between goals and temptations (e.g., in the first sentence of the abstract). I don’t think this is actually correct; counteractive control theory is indeed consistent with the hypothesized pattern, and the pattern could represent one mechanism behind counteractive self-control, but counteractive control theory, as far as I understand it, doesn’t require a mechanism at the nonconscious level. In fact, this would have been the main contribution of the target paper—to probe deeper into the mechanisms behind counteractive control. Moreover, the target paper doesn’t explicitly derive its predictions from counteractive control theory—see, for example, the sections on “The Mental Structure of Personal Goals” and “Asymmetrical Activation Patterns” (Fishbach et al., 2003). And, finally, the authors essentially concede this point at the very end of the Discussion, in which they acknowledge that their results don’t speak to all potential asymmetries between goals and temptations. Therefore, in the spirt of not throwing the baby out with the bathwater, I would recommend that the authors carefully look over their manuscript and revise statements about counteractive control theory, where these are incorrect or imprecise—unless, of course, the authors can cite specific interpretations of counteractive control theory that indeed make such claims.
2. The authors present analyses that differ from those presented by the target paper; it would be helpful for readers to learn why they gave preference to different analytical techniques. Would there be any particular advantages? (The authors also present analyses with techniques that mirror those used by the target paper, so there is no concern about differences in analyses driving the different findings.) Related to this point, it would also be helpful if the authors could expand on the rationale/intuition behind the specification of the inverse Gaussian distribution and the glmer function in R (p. 14).
3. While I realize that the paper is not intended as a ‘commentary’ per se, I would encourage the authors to interrogate whether the target paper’s hypothesis was plausible in the first place. Extraordinary claims require extraordinary evidence, and it would seem that the target paper is making extraordinary claims. How is it that nonconscious cognitive processes could be capable of classifying a ‘goal’—which must be protected and pursued—and, nearly simultaneously, identify that an unexpected stimulus encountered is a ‘threat’, which must be inhibited? How could this possibly develop from ‘over-learning’, as the authors of the target paper argue? It would seem to require sentience of some form. I am thus concerned that the target paper has been on thin theoretical ice from the very beginning; it feels like a homunculus argument—we have a ‘little mind’ within, which makes intelligent judgments automatically and below our threshold of conscious awareness.
4. I would also encourage the authors to interpret their findings not just against the target hypothesis, but also against the wider theoretical frameworks discussed in the target paper, in particular in the section, “Alternative Explanations.” The target paper claims that “temptation–goal associations are very different in nature from semantic associations” (p. 306), and then it argues that the pattern of asymmetric associations is inconsistent with predictions of spreading activations models for hierarchical associative networks. The latter would suggest reduced activation by higher order concepts for the lower order concepts, due to dilution from more connections (as the target paper puts it, “…robin activates bird more than bird activates robin”; p. 306). In the case of the present paper, it is the goal that appears to activate temptation, more so than the other way around. Would this mean that the goals potentially are lower in the hierarchy? In the presence of a goal, temptations (or obstacles) would always be at issue. But a ‘temptation’ could be tied to pleasant hedonic experiences, presumably, without evoking a competing goal. In other words, perhaps the temptation concept is inherently more ‘diluted’ than is the goal concept. Wouldn’t this be the essence of many self-control problems, that people sometimes struggle to see that tempting behavior conflicts with more important goals (as in, e.g., Myrseth and Fishbach, 2009)?

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  |  | ✔ |

**Author Response**  
Sep 12, 2021

Dear Dr. William McAuliffe, Editor of *Collabra: Psychology*

Please find submitted our revised manuscript, “Leading us unto temptation: No evidence for an asymmetry in automatic associations between goals and temptations”.

Thank you for the opportunity to further improve and clarify our paper. We found the comments from yourself and the two external reviewers to be very helpful, and have revised the manuscript accordingly. Attached is point-by-point description of how we responded to and addressed each reviewer comment. We would also be happy to make further revisions to the manuscript if needed.

Thank you for the opportunity to resubmit our manuscript, and for your consideration of our submission.

Sincerely,

Dr. Zoë Francis, on behalf of all authors

**Response to Editor and Reviewers**

* Do you have syntax for how you got from the raw data file to the critical file csv that you analyzed? If so, can you upload it? I won’t have the software knowledge to verify that no mistakes in data cleaning were made, but perhaps future readers will.   
    
  Absolutely! We have updated the data wrangling R syntax, along with all of the raw data files, to a new subfolder on the OSF project page.   
    
  Because some of the data was collected via DirectRT in-lab and the remainder of the data was collected online via JSPsych, there are two sets of raw data and two partially separate pipelines (that eventually merge the two data sources together). The data wrangling process was thus a bit complicated, but hopefully we’ve added sufficient comments in the syntax to help interested future readers verify our work.
* Upon running mod1 (and all of the other models), I get a convergence failure: “Warning message:In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :Model failed to converge with max|grad| = 0.00204309 (tol = 0.002, component 1).” Did you get this warning message as well? If so, it should be reported, and an alternative estimation strategy should be pursued and reported as unplanned. I reran mod1 using the allFit function, which tries a bunch of different optimizers. The optimizer “bobyqa” is the only one that converged without issues.

Of course, possibly this is a problem with my computer, the version of R I used (please report the version you used), or because there was some intermediate step I needed to take before running the models that is not documented. (Speaking of which, it would be helpful if there was a README file on OSF so that readers know exactly what each file is, what each variable name means, etc.). If you aren’t running into these issues, please double-check that your R settings aren’t suppressing warning messages, and then re-run your analyses on a computer other than the original computer you used and maybe a different version of R as well to see if software specifications could be the issue. In each model I got slightly different estimates than what you report, but perhaps this is because I got convergence errors and you did not. I am not the savviest at debugging software problems, so these issues could definitely be my fault, but your readers might not be R savvy either so it is important that the code will run without issue for them.  
  
We had also inconsistently experienced some convergence issues. In addition to updating the OSF analysis code to specify the use of a bobyqa optimizer, we have added the following footnote to the methods section:   
  
(Page 15). “The preregistered glmer code often resulted in convergence warnings (e.g., for main analysis of asymmetry effect, max|grad| = .00205, tol = .002), so we altered our analysis plan to use a bobyqa optimizer. These analyses resulted in nearly identical parameter estimates as those estimated with the default optimizers (as we preregistered), but without convergence warnings. Additionally, for some models it was necessary to predict reaction time in seconds (rather than in ms) to address convergence problems from models being nearly unidentifiable. In these cases, coefficients were subsequently transformed so that all parameters are reported in ms.”  
  
We also went through each analysis reported in the results section and ensured that all of the reported statistics were from converged models with no warnings (that match the R code posted on OSF).  
  
We have also added the version of R and version of the glmer package to the methods section:   
  
(page 15). “…using the glmer function from the lme4 version 1.1-27.1 package (Bates et al., 2015) with R version 4.0.4 (R Core Team, 2021).

Finally, we have added a READ.ME file to OSF, which explains the files and variable names.

* An additional robustness check to consider would be to exclude no data whatsoever from analysis. All exclusions have the potential of inducing selection bias, violating random assignment. To handle extreme values, a robust estimator could be employed (though I am not an expert on handling influential data points so take this recommendation with a grain of salt), [https://search.r-project.org/CRAN/refmans/robustlmm/html/rlmer.html](https://linkprotect.cudasvc.com/url?a=https%3a%2f%2fsearch.r-project.org%2fCRAN%2frefmans%2frobustlmm%2fhtml%2frlmer.html.&c=E,1,pwayoAHMX14_n2hQGvM_dQV5hUwB3DCe6c2FPU0w2U0DgYwL6zrCBaTxODbDXrvqbaYLUlY6cJ-P5Ffy9ZG5-foaJdWvt3lzFb5SQeiJwvI,&typo=1&ancr_add=1).

We confirmed that the results stay relatively consistent with the inclusion of the 21 participants who had poor accuracy (less than 80%) on the lexical decision task. In this expanded dataset, there was still a significant main effect of priming and no effect of target type (goal-related or temptation-related. There was also no interaction between priming and target type (p = .40), again failing to find the hypothesized pattern of moderation.   
  
In our analysis that directly replicates the analysis plan of Fishbach et al., 2003 (using ANOVA, etc.), we realized that we had unintentionally used the same participant exclusion criteria (excluding those 21 participants). Because Fishbach et al., 2003 did not have any stated exclusion criteria, the direct replication analysis now includes these 21 participants. Results are very consistent across the two different sets of outlier criteria and exclusions.

* I know what summary() does, but not summaryh(), which did not run for me. Is this a typo?   
    
  Summaryh is a function from the custom package hausekeep (available on github), which presents the same information as summary(), but automatically formatted into APA style.   
    
  We have added the installation source of the hausekeep package to the R code on OSF, with a note stating that it is the source of the summaryh() function, and explaining what it does; identical statistical results can be retrieved using summary() instead, if someone would rather not download the package.
* Where is the R code replicating the original Fishbach et al. (2003) analytic strategy?  
    
  Apologies! We had added that analysis to the paper later, and had overlooked updating the OSF R file. The R code that follows the Fishbach et al (2003) analysis strategy has now been added.   
    
  When adding this code (in combination with considering the exclusion criteria discussed above), we also noticed that we had excluded the 21 participants with low overall accuracy from the analyses replicating the original Fishbach analytic strategy. Because Fishbach et al. do not mention any exclusion criteria, we have updated the statistics in the “Alternative analysis: Comparison to Fishbach et al” section to include these 21 participants. Some of the results slightly differ (i.e., the unexpected interaction term is no longer significant), but the main results – the failures to replicate – remained.   
    
  (page 19). “The 21 participants who had been excluded from the above preregistered analyses (due to having lower than 80% accuracy on the lexical decision task) were re-included for these analyses.”
* Page 7: “A sensitivity power analysis conducted with simr (Green & Macleod, 2016) finds that our sample had over 95% power to detect an interaction effect of 30ms (r = .06); this effect size is one quarter of the asymmetry effect size found in Study 1 of Fishbach et al., 2003 and approximately half the asymmetry effect size from Study 3. Additionally, our sample had over 80% power to detect a correlation between the degree of priming asymmetry and self-regulatory success of r = .19 (from Study 3 of Fishbach et al., 2003).” It not clear to me how you extracted these figures from the simr code you posted. Can you spell it out in the syntax file?  
    
  We have added two additional lines of code and additional comments to the syntax file to more clearly walk through how we found the 95% power for an interaction effect of 30ms (and why 30ms is equivalent to an effect size of *r* = .06).   
    
  The second power analysis, for the correlation, was determined separately using G\*Power. We have clarified that in the manuscript text:   
    
  “Additionally, our sample had over 80% power to detect a correlation between the degree of priming asymmetry and self-regulatory success of r = .19 (from Study 3 of Fishbach et al., 2003) according to G\*Power version 3.1.9 (Faul et al., 2009)”.
* Fishbach et al. (2003) used one-tailed tests in their experiments, at time reporting p-values that would not have met the conventional two-tailed alpha of .05 (e.g., study 2, p. 300). Support for the moderating effect of self-regulation strength appears to be marginal even when using (what I think are) two-tailed p-values (Study 3, p. 302). You might consider noting this as a source of uncertainty in whether the original findings really provided strong support in the first place, even by the crude standards of p <.05.   
    
    
  Thank you, that is an important point. We have added the following to our discussion of moderations:  
    
  (Page 22). “…one could argue that our study was underpowered to find moderation by these individual differences, if the moderations are small. However, the original evidence for these moderations is relatively weak, and does not meet current standards for statistical evidence. For example, the moderation by self-regulatory success used one-tailed tests and were not all significance (Fishbach et al., 2003) and a p-curve analysis of Fujita & Sasota (2011) finds inconclusive evidentiary value for moderation by construal level3 (Simonsohn et al., 2014). Instead, the asymmetry between temptations and goals may not exist at the most automatic level of processing, even for those individuals high in self-regulatory success.  
    
  Footnote (page 22):   
  “For a p-curve analysis of the moderation by construal-level, we included the simple effects of construal level for high-goal value conditions, for each of the three studies from Fujita & Sasota: *t*(122)=2.09, *t*(158)=2.71, and *t*(157)=2.02. The continuous test for inadequate evidential value was *z* = -1.46, *p* = .07, while the continuous test for evidential value was *z* = -0.58, *p* = .72.”
* Fishbach et al. 2003, p.302: “These measures of perceived success in pursuing academic goals were designed to capture the ease/difficulty of overcoming temptation (in addition to general academic ability). The items were highly correlated (r = .83, p <.001) and were therefore combined into a unitary index of perceived self-regulatory success.” You report a smaller correlation between these items, was it significantly smaller? Also, one might argue that this measure much more strongly taps cognitive ability than self-regulation strength. The weakness of this measure might be seen as an additional justification for the inclusion of the trait self-control measure you used.  
    
  Thank you for pointing this out. Yes, we do find a significantly smaller correlation between the two items, and have added that observation to the manuscript.   
    
  (page 13) “This correlation between the two items was significantly lower than the correlation found in Fishbach et al., 2003 (Study 3, *r* = .83; Fisher comparison *z* = 3.64, *p* < .001)”.   
    
  (Limitations section, page 24). “Additionally, the two-item measure of self-regulatory success had significantly lower internal reliability in our sample compared to in the original Fishbach and colleagues (2003) study. Because this measure has only two items and has not been externally validated, it may not have been a reliable or valid measure of self-regulatory success.”  
    
  In regards to the relative strength of the trait self-control measure (relative to the self-regulatory success measure), we have added the following to the introduction:   
    
  (Page 6). “Because trait self-control is conceptually similar to self-regulatory success, and the trait self-control measure has been validated and is more commonly used (de Ridder et al., 2012), we also examine whether trait self-control moderates the predicted asymmetry”.
* What did exploratory analysis of the funnel debrief reveal?  
    
  Many of the participants did expect that the lexical decision task was somehow associated with the “second study” (the individual difference measures and ratings of goal importance), but they did not specifically know that the lexical decision task was a measure of priming. Participants most often thought that the first task was measuring their reaction time or vocabulary in general (e.g., as a measure of cognitive ability). We did not code participants’ written answers from the funnel debrief, and so did not conduct any quantitative analyses.   
    
  The participants overall also did recall seeing the goal and temptation-related words more often than the neutral words (which was accurate, given that the goal and temptation words did each appear more often than any individual neutral/irrelevant word).
* You use an inverse gaussian model which is mentioned several times in the pre-registration, save that the “Sample Size” section of the pre-registration calls for an ex-gaussian model. Please clarify if this a typo, a holdover from an earlier plan, or something else.  
    
  Thank you for your detailed reading! I now see that we accidentally stated “ex-gaussian” in the “sample size” section of the pre-registration, which was indeed an error. We always intended to use the inverse-gaussian model, as stated in the Analysis section of the preregistration and specified in the preregistered R analysis code (argument “family = inverse.gaussian(link = "identity"))”.
* As an aside, I don’t think it matters much here but in future papers you may want to consider not grand-mean centering (see Antonakis, J., Bastardoz, N., & Rönkkö, M. (2021). On ignoring the random effects assumption in multilevel models: Review, critique, and recommendations. Organizational Research Methods, 24(2), 443-483.)  
    
  Thank you for the reference! We will consider this for our future research.
* Page 22: “Although our student body is ethnically and culturally diverse, the sample is not representative of the population at large, and these results cannot necessarily be generalized to people from other age groups, cultures, or educational backgrounds.” I appreciate you mentioning this constraint on generalizability. However, it could also be worth mentioning that this is cold comfort for Fishbach et al. (2003) unless there is some principled explanation for why the results would not hold in your sampled population. (Maybe their average self-regulation abilities are too low?) Given that they tested the theory on a similar population, I don’t see much room for an appeal to selection bias.  
    
  We agree that our sample is similar in composition to the sample used in Fishbach et al., 2003, and did not mean to imply that a difference in sample could be used to explain the difference in results. We have rephrased this paragraph in the limitations section to the following:  
    
  (Page 24) “The current study was also conducted with a sample of undergraduate students who were enrolled in a Canadian university. This was an appropriate sample for studying academic goals and was similar to the undergraduate participants in Fishbach et al., (2003), with our sample also rating their academic goals as highly important. While a difference in sample composition cannot easily explain the difference in results between the current study and the original study, neither study used a representative sample. Thus, these results cannot necessarily be generalized to people from other age groups, cultures, or educational backgrounds.   
    
  We have also added the following to the Participants section of the methods, where we directly compare the importance ratings of academic goals across both samples:  
    
  (Page 9) “The sample of students rated their academic goals as highly important (on a scale from 1-6, M = 5.34, SD = 0.73, mode = 6). The importance of academic goals for these participants was not significantly different from the importance of academic goals among the students who participated in Fishbach et al. (2003)’s Study 3 (converted to 6-point scale, M = 5.17, SD = 0.84, mode = 6; comparison t(296) = 1.69, p = .092).”

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.

REVIEWER 1 COMMENTS

This manuscript reports the results of an attempted replication of a seminal finding re: a key tenet of counteractive control theory: namely, the asymmetric cognitive activation of goals and temptation-related concepts upon priming. The manuscript is very well written, the study clearly and thoroughly reported. The study is well-powered, methods are sound and well-placed to robustly test this aspect of a widely cited theory (and extend the original study in an important way, with the addition of irrelevant prime words and increased statistical power). The manuscript is a great fit for Collabra, and authors provide links to pre-registered study protocol, open data and code. I have only several very minor suggestions for the authors.

* Page 5 – a brief definition/description of construal theory/abstract construal would be useful when first mentioned. Shortly after on page 6, reporting the direction of the moderation by temptation strength would be useful.  
    
  Thank you for pointing out these oversights. We now provide a brief description of construal level, and provide more description of the moderation by temptation strength.   
    
  (Page 6). “Construal level theory describes how people can think about the same event at different levels of abstraction – for example, one might focus either on low-level, concrete details of an event (e.g., “holding a pen in my hand”) or on higher-level, abstract representations of the event (e.g., “expressing my thoughts”).”  
    
  (Page 6). “Participants more readily identified dieting goal-related words after being primed with a weak temptation (e.g., a picture of a somewhat attractive cake) and were slower to identify dieting goal-related words after being exposed to a stronger temptation (e.g., a picture of a highly attractive cake).”
* Related to these moderators, it would be useful to see treatment of the distinction/overlap and potential relationship between state-level moderators and stable trait-like moderators in the introduction. E.g., briefly discuss the treatment / testing of these moderators in past studies (experimental conditions - manipulations of state-based constructs versus testing individual differences) and justify the use of individual difference measures. I note that this is discussed briefly in the discussion section, but including a brief treatment of this issue in introduction would enhance understanding of the existing research / situate the findings re: moderators.  
    
  We agree that noting these methodology differences in the introduction is worthwhile, to clarify that we are examining potential moderations by construal level and temptation strength in a different way than previously done. We have changed that section of the introduction to the following:   
    
  (Page 6) “We thus investigated four potential moderators: self-regulatory success, construal level, temptation strength, and trait self-control. Note that construal level and temptation strength were previously examined as experimentally manipulated, state-level moderators, while self-regulatory success and trait self-control were examined as measured, trait-level moderators. Because the primary purpose of this research was to replicate Study 3 of Fishbach et al., 2003, we did not include any additional manipulations in our experimental paradigm. Instead, we examine all four potential moderators as measured individual difference variables, at the trait-level (see Discussion), as preregistered secondary analyses.”
* Additional details of pilot study are needed – e.g., how many people recruited and how/where recruited?  
    
  We have added the additional details from the pilot study:  
    
  (Page 9) “(*N =* 130 undergraduates from a different large Canadian university participant pool)”

REVIEWER 2 COMMENTS

This paper presents a carefully executed direct replication of a key study (Study 3) in a highly influential paper on the cognitive foundations of self-control (Fishbach et al., 2003); the authors argue that there are surprisingly few published replications of the central, counterintuitive result—asymmetric automatic associations between goals and temptations—thus making a compelling case for their paper.

The paper comes across as clear, complete, and carefully crafted. It is methodologically sound, and the analyses are thorough, extended in depth and rigor beyond those of the target paper. The exposition is sharp, and the argumentation effective.

This paper makes a very valuable contribution to the literature, and I would recommend that it gets published, subject to revision. Specific comments and requests follow below:

1. The authors argue at multiple points that counteractive control theory predicts the targeted effect, asymmetry in associations between goals and temptations (e.g., in the first sentence of the abstract). I don’t think this is actually correct; counteractive control theory is indeed *consistent* with the hypothesized pattern, and the pattern could represent *one* mechanism behind counteractive self-control, but counteractive control theory, as far as I understand it, doesn’t require a mechanism at the nonconscious level. In fact, this would have been the main contribution of the target paper—to probe deeper into the mechanisms behind counteractive control. Moreover, the target paper doesn’t explicitly derive its predictions from counteractive control theory—see, for example, the sections on “The Mental Structure of Personal Goals” and “Asymmetrical Activation Patterns” (Fishbach et al., 2003). And, finally, the authors essentially concede this point at the very end of the Discussion, in which they acknowledge that their results don’t speak to all potential asymmetries between goals and temptations. Therefore, in the spirt of not throwing the baby out with the bathwater, I would recommend that the authors carefully look over their manuscript and revise statements about counteractive control theory, where these are incorrect or imprecise—unless, of course, the authors can cite *specific* interpretations of counteractive control theory that indeed make such claims.  
     
   Thank you for this providing this conceptual clarification; we certainly agree that the lack of an automatic, non-conscious asymmetry does not dispute all of counteractive control theory, and does not even directly speak to asymmetries at the level of behaviour. We have edited multiple places in the manuscript (including the abstract, introduction, and discussion) to specify that this study is specifically testing the claim that the asymmetry occurs at the automatic level. In places where we previously referred to claims made by “counteractive control theory”, we now refer to “the asymmetry hypothesis”, and further clarify in multiple places that this study is focused specifically on the asymmetry in automatic associations. Below are some examples of the changes that we have made to increase the precision of our writing:

(Abstract, Page 2): “The asymmetry hypothesis of counteractive control theory suggests that – at least for successful self-regulators – exposure to temptations facilitates the accessibility of goal-related cognitive constructs, whereas exposure to goals inhibits the accessibility of temptation-related cognitive constructs.. Using a lexical decision task, Fishbach et al., 2003 (Study 3) found that this asymmetry even existed at an automatic level of processing.”   
  
(Introduction, page 4): “However, the *asymmetry hypothesis* of counteractive control theory posits exposure to a temptation activates the concept of one’s goal, but exposure to one’s goal does *not* activate the thought of temptations (Fishbach et al., 2003, 2010; Fishbach & Converse, 2010). Fishbach et al., (2003) suggest that this asymmetry, which cannot be explained by semantic relatedness or co-occurrence alone, plays a role in self-regulation processes.”  
  
(Discussion, page 21) “opposite to what the asymmetry hypothesis of counteractive control theory predicts.”

1. The authors present analyses that differ from those presented by the target paper; it would be helpful for readers to learn why they gave preference to different analytical techniques. Would there be any particular advantages? (The authors also present analyses with techniques that mirror those used by the target paper, so there is no concern about differences in analyses driving the different findings.) Related to this point, it would also be helpful if the authors could expand on the rationale/intuition behind the specification of the inverse Gaussian distribution and the glmer function in R (p. 14).  
     
   We have added the following explanation in the analysis section, including references that further discuss the benefits of trial-level analysis. As the reviewer mentioned, the particulars of the analytic technique did not matter in this case, but we are happy to direct readers to other people who have made more eloquent arguments for the advantages and limitations of trial-level analysis.   
     
   (page 15) “Analyzing trial-level data is recommended where possible, as it does not ignore the variability that exists in human behaviour across different trials, and the resulting statistics better match the theoretical claims that we aim to make (Lo & Andrews, 2015; Speelman & McGann, 2013; Whelan, 2008). Furthermore, only by analyzing trial-level data can we model random stimuli for each target word, allowing us to more readily generalize our results to target words beyond the particular words included in the study (Yarkoni, 2020). Because raw reaction time was the dependent variable, and reaction times are nearly always right-skewed, we specified an inverse Gaussian distribution (following recommendations by Lo & Andrews, 2015) using the glmer function in R (Bates et al., 2015).”
2. While I realize that the paper is not intended as a ‘commentary’ per se, I would encourage the authors to interrogate whether the target paper’s hypothesis was plausible in the first place. Extraordinary claims require extraordinary evidence, and it would seem that the target paper is making extraordinary claims. How is it that nonconscious cognitive processes could be capable of classifying a ‘goal’—which must be protected and pursued—*and*, nearly simultaneously, identify that an unexpected stimulus encountered is a ‘threat’, which must be inhibited? How could this possibly develop from ‘over-learning’, as the authors of the target paper argue? It would seem to require sentience of some form. I am thus concerned that the target paper has been on thin theoretical ice from the very beginning; it feels like a homunculus argument—we have a ‘little mind’ within, which makes intelligent judgments automatically and below our threshold of conscious awareness.  
     
     
   We agree. We found the original theoretical explanation of the proposed asymmetry hypothesis (e.g., via overlearning) to be underdeveloped in the original manuscript (perhaps due to its “thin theoretical ice”). Because we do not want to misrepresent the target paper’s intended theoretical explanation, we have only included a brief statement of the proposed explanation for an automatic asymmetry. We also allude to the implausibility of the asymmetry in the discussion section.   
     
   (Page 4). “Fishbach et al., (2003) suggest that this asymmetry, which they state cannot be explained by semantic relatedness or co-occurrence alone, is learned due to its self-regulatory functionality.”  
     
   (Discussion, Page 22): “The hypothesis of an asymmetric association between temptations and goals at the subconscious, automatic level (the level tested here) seems to be the most implausible of the potential mechanisms that could explain asymmetries observed at a behavioural level and counteractive control more broadly.”
3. I would also encourage the authors to interpret their findings not just against the target hypothesis, but also against the wider theoretical frameworks discussed in the target paper, in particular in the section, “Alternative Explanations.” The target paper claims that “temptation–goal associations are very different in nature from semantic associations” (p. 306), and then it argues that the pattern of asymmetric associations is inconsistent with predictions of spreading activations models for hierarchical associative networks. The latter would suggest reduced activation by higher order concepts for the lower order concepts, due to dilution from more connections (as the target paper puts it, “…robin activates bird more than bird activates robin”; p. 306). In the case of the present paper, it is the goal that appears to activate temptation, more so than the other way around. Would this mean that the goals potentially are lower in the hierarchy? In the presence of a goal, temptations (or obstacles) would always be at issue. But a ‘temptation’ could be tied to pleasant hedonic experiences, presumably, without evoking a competing goal. In other words, perhaps the temptation concept is inherently more ‘diluted’ than is the goal concept. Wouldn’t this be the essence of many self-control problems, that people sometimes struggle to see that tempting behavior conflicts with more important goals (as in, e.g., Myrseth and Fishbach, 2009)?

Thank you for this contribution, that’s an excellent point. We have added the following paragraph to the discussion:  
  
(Discussion, page 21) “The small interaction found instead (Figure 1), unlike the original findings, can be explained via a hierarchical semantic network, where lower-level concepts activate higher-level concepts more readily than higher-level concepts activate lower-level concepts (Anderson, 1983; Collins & Loftus, 1975) – for example, a lower-level concept “poodle” activates the higher-level concept “dog” more than the reverse. Temptation-related concepts, like procrastination and distraction, do not only apply to academic-goals, but also are relevant in other contexts (e.g., one can procrastinate instead of exercising or going to sleep; Bernecker & Job, 2020; Brown, 2019). Temptations like procrastination may thus be a higher-level concept, with more disparate connections to multiple lower-level concepts (academics, exercise, and sleep). On the other hand, academic-goal concepts (like “homework”) may be lower in the hierarchical semantic network, with fewer connections to other concepts, thus more readily priming the related temptation.”

Thank you again to both reviewers and the editor for your helpful comments and suggestions. We believe that our manuscript has been greatly strengthened by the reviewing process, and hope that the revisions have addressed your questions and concerns.

**Editor Final Decision: Revise & Resubmit**

Sep 19, 2021

Dear Zoë Francis,

I have now read your revised manuscript. I appreciate your careful attention to the concerns the reviewers and I raised. I am happy to provisionally accept your manuscript for submission. However, I found a few small things I would like you to address.

First, please add the bobyqa optimizer as an argument to the glmer models; Version 7 of your syntax I see posted, while updated in the other ways you described in your revision cover letter, still appears to use the default optimizer and still produces convergence warnings. When I add the bobyqa optimizer, I get the same point estimates as your report, but different standard errors, test statistics, and p-values (see example below). I checked and am using the same versions of the R and the lme4 package as you report having used. Please double-check the numbers you report (if your models are printing the same numbers you reported, then we probably just won’t get to the bottom of why I am getting slightly different ones, which is OK).

mod1 <- glmer(rt\_new ~ Prime\_RelorNoC\* TargetGoalTemptC

* + (1|stimulus) + (1|run\_id), data = Critical, control=glmerControl(optimizer="bobyqa",optCtrl=list(maxfun=100000)), family = inverse.gaussian(link = "identity"))

summaryh(mod1)  
term results  
1: (Intercept) b = 633.70, SE = 7.66, z = 82.71, p < .001, r = 1.00  
2: Prime\_RelorNoC b = -8.87, SE = 1.49, z = -5.95, p < .001, r = -0.93  
3: TargetGoalTemptC b = -5.15, SE = 6.79, z = -0.76, p = .448, r = -0.82  
4: Prime\_RelorNoC:TargetGoalTemptC b = 5.41, SE = 2.82, z = 1.92, p = .055, r = 0.83

Second, are you certain that z is the test statistic you should be reporting, not t? This appears to be a discrepancy between summary() and summaryh().

Third, your revision cover letter and manuscript say that you did not exclude participants for the Fishbach 2003-style anova analyses. But your Version 7 R syntax uses the dataset with the exclusions. Perhaps your reported analyses came from a Version 8 that you have yet to post?

Fourth, where is the input and output of the p-curve analysis documented? Can you post this and update the README accordingly?

Fifth, you can de-anonymize the OSF links now.

Last, could you add the information you provided me regarding the funnel debrief as a footnote in the manuscript?

I look forward to receiving your final revision and accepting it for publication in Collabra: Psychology.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact the editorial office at [editorialoffice@collabra.org](mailto:editorialoffice@collabra.org).

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

William McAuliffe

**Author Response**  
Nov 3, 2021

Dear Dr. William McAuliffe, Editor of *Collabra: Psychology*

Please find submitted our re-revised manuscript, “Leading us unto temptation: No evidence for an asymmetry in automatic associations between goals and temptations”.

Thank you for the provisional acceptance of this paper and for your additional comments.

We have addressed each of the remaining points as follows:

* We have (i) changed the OSF links to be non-blinded, (ii) added the information about the funnel debrief to the manuscript in a footnote (now on page 14), and (iii) replaced the z-values with t-values. Thank you for noting that error/discrepancy between the summary and summaryh functions. We have also added to the methods section that the p-values were calculated using the lmerTest function.
* Additionally, we have updated the R script in OSF to correctly include the bobyqa optimizer and refer to the correct dataset for the secondary direct-replication analysis – when I made these changes to the OSF file in the earlier revision, I had not pressed “save” and so the publicly visible file was still incorrect! The revised R file is now posted.
* The models posted from this version of the R file have been consistently converging (with consistent results) when using the bobyqa optimizer on my computer, although they do take some time to compute. After re-downloading the data file from OSF and using the exact same R code as you did, I still get slightly different standard errors/p-values from the ones that you found/included in your letter… so it might unfortunately remain a mystery. At least none of this unexplained variation affects the overall interpretation of the results.
* All of the inputted values (three t-values) and the output from the p-curve analysis are listed in a footnote on page 23. Because all the input and output values are already in the manuscript, we have not posted additional information in OSF.

Thank you greatly for your support and helpful, detailed-oriented contributions throughout this process!

Sincerely,

Dr. Zoë Francis,

on behalf of coauthors

Aravinth Jebanesan

Dr. Michael Inzlicht

**Editor Final Decision: Accept**

Nov 16, 2021

Dear Dr. Francis,

I have now had a chance to read over the revision of your manuscript “Leading Us Unto Temptation? No Evidence for an Asymmetry in Automatic Associations Between Goals and Temptations.” I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work, it was a pleasure to have been involved in its peer-review process. I hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point. Our managing editor will contact you in case there are any pre-prodution file related questions. You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely,  
William McAuliffe