

Author response to reviews of

Manuscript BUL-2023-1221

Reporting bias, not external focus: A robust Bayesian meta-analysis of the attentional focus literature

submitted to *Psychological Bulletin*

[RC] Reviewer comment

Manuscript text

Dear Dr. Editor,

Thank you overseeing the review of our manuscript for publication at the *Psychological Bulletin*. We appreciate the positive and helpful comments we received from this review process. We believe these comments have helped to strengthen our manuscript.

Based on comments from both Reviewers, the associate editor, and yourself, we have made some changes to the manuscript. Below we provide our point-by-point response (in normal font) to the Reviewers' comments (in **bold**). All substantial changes in the revised manuscript are included in this response letter (in a text box) and appear in blue text in the revised manuscript. Thank you again for considering our work and overseeing the review process.

UPDATE THIS AFTER ALL RESPONSES ARE FINISHED

Assignment: Mike

1. Reviewer 1

[RC 1.1.] “Reporting bias, not external focus: A robust Bayesian meta-analysis of the attentional focus literature” is a very well-researched, informed, rigor, and well-written meta-analysis of this important area of motivational psychology. I enjoyed reading it, I found the analyses and methods compelling, the application rigorously conducted, and the interpretation of the statistical results important, correct, and remarkably nuanced. Because the data and codes are shared anyone can reproduce these findings. Furthermore, the results are important, novel, and have wide ranging practical consequences from the way elite athletes train to how musicians are taught. The topic is appropriate for *Psychological Bulletin*, and it makes a notable contribution to both theory and practice. It is also an early application of a new meta-analysis Bayesian model-averaging across different models of heterogeneity, and the leading meta-analysis methods (both fixed and random effects, as well as several of the better models of publication bias) in a robust way—RoBMA-PSMA (Bartos et al., 2023a). I believe that this method will soon be the default way to test and correct for potential selection for statistical significance. The findings here are also in line with the recent findings of a broad application of RoBMA-PSMA to psychology, in general—Bartoš et al. (2023b): Meta-Analyses in Psychology Often Overestimate Evidence for and Size of Effects. Royal Society: Open Science. <https://doi.org/10.1098/rsos.230224>.

Like this paper, Bartoš et al. (2023b) also finds fault with meta-analysis methods as the source of these ubiquitous false positive meta-analysis results. Thus, the application is timely and can serve as an exemplar of how to employ and interpret Bayesian modeling averaging for the readers of Psychological Bulletin and for future meta-analyses in psychology.

This paper is so well written, and its reasoning is so sound that I have no substantive comments about how the paper might be improved. This is the first meta-analysis study that I have reviewed in years, where I cannot think of any way to improve the analyses. Nonetheless here are a few comments that the authors might consider if they are asked to revise this paper.

Response here

Assignment: Mike

[RC 1.2.] The abstract is very nicely written.

Response here

Assignment: Mike

[RC 1.3.] I appreciate the authors' focus on the magnitude of effect sizes that can be regarded as scientifically meaningful rather than negligible. I have long argued that we need to focus on whether meta-analysis estimates are of a scientifically meaningful size (nonnull) as opposed to being merely statistically significant. This inappropriate focus, which is much like the critical debate surrounding null hypothesis testing in psychological research (especially p-values), is the source of much of the bias and false positives in meta-analyses of psychology.

Response here

Assignment: Mike

[RC 1.4.] On page 6, the authors state that: "Importantly, a random effects model assumes no reporting bias." While this is true, this does not reflect an extensive based that corroborates how important this assumption is. Please add something like: "and has been shown to be quite biased in the presence of selective reporting for statistical significance (Stanley, Doucouliagos and Ioannidis, 2017 &2022; Carter et al., 201; Kvaren et al., 2000; Bartoš et al, 2023, Bom and Rachinger, 2019; Carter et al., 2015 & 2018; Kvaren et al., 2000; Stanley, Doucouliagos and Ioannidis, 2017 &2022)." Unlike external focus, there is a lot of evidence, of every kind, that random-effects is very biased in the presence of selection for statistical significance.

Response here

Assignment: Mike

[RC 1.5.] On page 8, the authors state that: "we addressed the following questions: . . .(b) How sensitive are random effects estimates to the assumption that reporting bias is absent?" I agree that this is an important hypothesis to test and discuss. However, I do not think that the authors have fully addressed this. Please add a sentence or two in the 'Discussion' about how random effects estimates are quite

biased in these applications and how random effects estimates manufacture an estimate of the mean (and its statistical significance) that notably changes what we know about this research literature(s).

Response here

Assignment: **Mike**

[RC 1.6.] I also highly approve of the authors' use of leverage statistics. This is an important step to follow in all meta-analyses of psychology. More meta-analyses should adopt this practice as the default. Perhaps add this point to the 'recommendations' for research going forward?

Response here

Assignment: **Mike**

[RC 1.7.] The authors were generous and quite professional in the discussion of their findings relative to previous meta-analyses, especially Chua and colleagues (2021). This is appropriate and well written with the correct caveats. However, please add a couple of sentences about just how large (substantially so) the differences between their preferred estimate of mean effect and Chua and colleagues' (2021). This could be wrapped into the expanded discussion I mention in 4, above.

Response here

Assignment: **Mike**

[RC 1.8.] Lastly, more a philosophy of science consideration. As always, the authors do an excellent job of qualifying their findings with proper limitations. I ask for no changes to any of this. However, in my view of science, I believe that the authors are a little too generous to OPTIMAL theory. Yes, this lack of an effect might be explained by the limitations of our research methods and measures. Yes, this is a possible and logical inference to draw. But so is the possibility that OPTIMAL theory is just 'wrong.' This too is consistent with the evidence. I do not believe that we can sort out which is the 'true' inference to draw here, because all tests are tests of the theory combined with embedded limitations of our empirical/experimental methods to detect such an effect. I merely ask the authors to consider adding a sentence or a footnote that mentions this alternative explanation of their findings. In my view, if a mature theory that has had decades of empirical testing cannot produce corroborating evidence, then it should be abandoned and, in its place, adopt some new theory which can marshal genuine supporting evidence. If our methods remain too crude for any theory to be genuinely corroborated by the evidence, then we should admit this and stop calling these theories 'science.' Instead, such psychological theories are philosophy, which is where the discipline began.

Response here

Assignment: **Mike**

[RC 1.9.] Thank you for asking me to review this important study. It was a pleasure reading.

Response here

2. Reviewer 2

[RC 2.1.] This manuscript reports a Bayesian meta-analysis as a direct response to another meta-analysis published in Psychological Bulletin in 2021: “Chua, L.-K., Jimenez-Diaz, J., Lewthwaite, R., Kim, T., & Wulf, G. (2021). Superiority of external attentional focus for motor performance and learning: Systematic reviews and meta-analyses. Psychological Bulletin, 147, 618-645.” The authors of the current submission, applying a different type of analysis that allowed for the detection of publication bias, report findings that seriously question the validity of the conclusions of the previous work by Chua et al.

I am no expert on Bayesian meta-analyses and recommend to include another reviewer assessing/verifying the correctness of the reported analyses in this submission. If the reported analyses prove to be correct, then I think Psychological Bulletin should have a keen interest if not feel the responsibility to publish this manuscript as they decided to report the previous meta-analysis by Chua et al., which - if the analyses in this submission are correct - turns out to be flawed and reporting unevidenced conclusions.

In general, the manuscript is clearly structured, well written and doing a good job in summarizing the literature. I have a couple of relatively minor issues the authors may wish to consider.

This is our response

Assignment: Mike

[RC 2.2.] This paper first and foremost questions the validity of an empirical finding, namely that adopting an external focus is beneficial for motor learning. This is then linked to certain theories, including common coding theories or constraints-based approaches. What I miss is a broader discussion what the outcome of this meta-analysis (namely, no evidence for the benefit of an external focus) would in turn mean for the veracity of those theories, the original findings seemed to confirm. In the final discussion, the authors do this a bit for the OPTIMAL theory (which in my view, is not a theory proper, but at best a framework), but they do not discuss this for the more fundamental perception-action theories.

This is our response

Assignment: Mike

[RC 2.3.] On p. 5 (ll. 63-67) the authors argue that there is “consensus in the motor learning community that adopting an external focus of attention can improve motor performance [...]”. There are quite some empirical studies that have provided counterevidence that this must be true, for instance, showing distinct effects for experienced versus novice learners. Therefore, rather than saying that there is “consensus” I suggest to rewrite to read “a majority of researchers argue that [...]”. In addition, the relevant empirical work providing counterevidence should be presented, too.

For example (not this is not a comprehensive list and that there are several other papers that would need be included):

Perkins-Ceccato, N., Passmore, S. R., & Lee, T. D. (2003). Effects of focus of attention depend on golfers' skill. *Journal of Sports Sciences*, 21, 593-600.

Singh, H., & Wulf, G. (2020). The distance effect and level of expertise: Is the optimal external focus different for low-skilled and high-skilled performers? *Human Movement Science*, 73, 102663.

This is our response

Assignment: **Mike**

3. Reviewer 3

[RC 3.1.] The present manuscript reports a Bayesian meta-analysis of the effects of an external versus internal focus of attention, which has been a major topic of interest in the motor learning literature for the past two decades or so. As the authors note, several meta-analyses have been published of this literature. The authors re-analyzed the data of a set of comprehensive meta-analyses conducted by Chua et al. and published in this journal two years ago [Chua, L.-K., Jimenez-Diaz, J., Lewthwaite, R., Kim, T., & Wulf, G., (2021). Superiority of external attentional focus for motor performance and learning: Systematic reviews and meta-analyses. *Psychological Bulletin*, 147, 618-645]. According to the authors, their re-analysis showed evidence of publication bias in Chua et al.'s meta-analysis and only "negligible" effect sizes in favor of an external focus after correcting for publication bias.

General comments

The influence of a performer's focus of attention has been of significant interest to researchers in motor learning since Wulf, Höß, and Prinz (1998) first reported learning benefits of an external focus relative to an internal focus of attention (and no instructed focus). That interest has quickly spread to related fields, including physical therapy, music education, athletic training, medical education, etc. The popularity of this topic is, among other things, likely due to the fact that the advantages of an external focus have been replicated numerous times - perhaps more so than any other effect in the field of motor learning. Chua et al.'s (2021) meta-analyses confirmed the superiority of an external relative to an internal focus for both motor performance and learning. The current authors conducted a focal evaluation of publication bias in the literature reviewed by Chua et al. Based on their results, the authors questioned the "existing consensus that an external focus is always more effective than an internal focus" (abstract). I have a number of significant concerns with the paper that I detail below.

This is our response

Assignment: **Brad**

Major issues

[RC 3.2.] Why were "neutral default" priors used for Bayesian modeling in the current study when there is ample prior knowledge of the likely effect sizes from at least seven previous meta-analytic studies synthesizing more than two decades of experimental results? For example, do the current authors truly believe that there is equal chance of finding a beneficial effect of external focus of attention over internal focus of attention versus no such advantage, equal chance of significant between-studies

variability versus the presence of homogenous methodology in the extant literature on attentional focus, and equal chance of the presence of publication bias versus the absence of publication bias? Would the use of these unrealistic probabilities not be a case of demonstrating confirmation bias so that they could favorably set out to find what they want to find anyway by tailoring the prior and/or model to a suit a predetermined conclusion before the data can be observed? For example, using a less diffuse prior than what actually represents the phenomenon under study can falsely favor the null hypothesis and/or over-deflate the point estimates during bias correction. Assigning an equal prior probability to the null hypothesis and the other hypothesis is unrealistic and runs contrary to a reasonably informed, or knowledge-based, use of myriad empirical evidence. The RoBMA framework allows the use of different prior distributions so that “researchers with substantial prior knowledge can test more specific hypotheses than those specified with the default model ensemble or incorporate prior knowledge about the research environment” (Bartoš et al., 2023). Were attempts made to explore other prior distributions in effect, heterogeneity, and publication bias? If not, explain the decision not to. Provide the sensitivity analysis results for checking the impact of using different probability values and prior definitions. Also discuss how any prior misspecification of the alternative hypothesis may render the values of the Bayes factors inaccurate since they do not provide an objective assessment, but can be influenced by subjective beliefs of the analyst in defining the alternative hypotheses. What criteria were used to pick the “best” performing model(s)? The results of model comparisons should be reported. In addition to model comparison, was model checking done to ensure no misspecification (especially for the “best” performing models) and to facilitate model expansion (e.g., using a parametric of higher order than that used in the PEESE method) if the current class of models do not fit the data? Please elaborate on the approach adopted (e.g., visual inspection of plots on model and data, posterior predictive check) and report the results (e.g., non-existent impact of misspecification, profound impact of misspecification) thus obtained.

This is our response

Assignment: Brad

- [RC 3.3.] Several discrepancies stand out in the abstract. The Bayes factor indicating the posterior odds favoring the null hypothesis should range from 1.33 to 5.88 (but reported as 1.3 and 5.74, respectively, in the abstract of the current manuscript version) for all studies including outliers, and from 1.37 to 3.85 for studies excluding outliers, as reported in the results section (p. 13-16). Why were the former set of numbers selected for reporting in the abstract despite efforts to identify extreme studies and remove them from the sensitivity analysis? Were the data of these outlier studies also included in the selection moderator analysis as well as Bayesian model fitting or averaging? Four research questions were clearly stated (p. 8 line 162 to p. 9 line 169), but only findings pertaining to the first three were reported in the abstract. Did the current authors choose to omit the reporting of the overall result of the analysis made to address their fourth objective because it was a non-significant finding?

This is our response

Assignment: Brad

- [RC 3.4.] Bayesian analyses are typically used when there is limited knowledge because there are only a few primary studies or empirical data are known to be unreliable. These conditions do not appear to be the case here. During the literature search process, for example, previous meta-analyses found large numbers of studies that are suitable for full-text review prior to final eligibility consideration

for quantitative synthesis. Kim et al. (2017) identified 42 articles for full-text review from an initial pool of 790 search results on attentional focus studies evaluating balance learning alone (without studies evaluating the performance of balancing skills). Makaruk et al. (2020) reviewed the full text of 47 articles from an initial pool of 380 search results on attentional focus studies evaluating jump performance alone (without studies evaluating the learning of jumping skills). Grgic et al. (2021) found 28 articles for full-text review from an initial pool of 47 search results on attentional focus studies evaluating acute and long-term effects on gaining muscular strength alone (without studies evaluating the learning of skills). From an initial pool of 144 search results on attentional focus studies evaluating acute effects on gaining muscular endurance alone (without studies evaluating the learning of skills), 20 articles were identified for full-text review in Grgic and Mikulic (2022). Nicklas et al. (2022) found 225 articles for full-text review from an initial pool of 2,382 de-duplicated search results on attentional focus studies evaluating performance alone of various motor skills (without studies evaluating the learning of skills). Eleven full-text articles were reviewed from an initial pool of 78 search results in Li et al. (2022) on attentional focus studies evaluating sprint performance alone (without studies evaluating the learning of sprinting skills). Notably as reported in the article elected by the current authors for re-analysis, the literature search originally produced 262 articles on attentional focus studies evaluating performance and/or learning of various motor skills that are suitable for full-text review from an initial pool of 2,680 search results after de-duplication (Chua et al., 2021). Please comment on the choice of using a Bayesian method given that this literature is well-fleshed out and relatively large.

This is our response

Assignment: **Brad**

[RC 3.5.] It is unclear if the current authors' use of "reporting bias" refers to the same intended use of "publication bias" by the authors of references cited in the current study (e.g., Chua et al., 2021; Kim et al., 2017; Lohse et al., 2016; McKay et al., 2022). Conceptually, publication bias (i.e., bias as a function of whether a study is published or not) can be differentiated from reporting bias (i.e., within-study outcome or selective reporting bias) in that the former represents the preferential publication of some research findings over others in terms of the direction and/or strength of the evidence (e.g., positive and significantly significant based on a predetermined p-value) whereas the latter describes the preferential reporting of a subset of outcomes in a published study (Dwan et al., 2013). The former is more associated with the hidden elephant of missing studies as compared to the latter being related to the missing-data problem, requiring different methodological approaches for assessment and treatment. According to the current authors' analysis, there was not a significant difference between the one-primary outcome measure effect and the other-outcomes measure effect synthesizing one, two, or three outcomes not selected by Chua et al. (2021) from each of their included studies (i.e., no outcome-selection bias).

This is our response

Assignment: **Brad**

[RC 3.6.] In the Chua et al. (2021) paper, as with most systematic reviews and meta-analyses, a number of potential dimensions to the motor learning literature on external and internal attentional focus were examined, including risk of bias and potential sources of heterogeneity. Importantly, Chua et al. used the approach of Mathur & VanderWeele (2020) in sensitivity analyses to evaluate the risk of publication bias. The Mathur & VanderWeele method includes "a worst-case meta-analytic point estimate for maximal publication bias" that is obtained by conducting a standard meta-analysis of only the negative

and ‘non-significant’ studies. It was developed in part through systematic review of multiple meta-analyses in several fields, including experimental psychology, to assess the amount of publication bias that would be required to attenuate the point estimate or its confidence interval to any non-null value. The current authors give short shrift to this comprehensive approach of Mathur & VanderWeele (2020) that was used in the original Chua et al. meta-analyses. Indeed, there is no acknowledgement that the phrasing developed by Mathur & VanderWeele to communicate the level of publication bias came from them: ‘For publication bias to shift the observed point estimate to the null, ‘significant’ results would need to be at least [30 fold] more likely to be published than negative or ‘non-significant’ results.’ Instead, the current authors note that “it is plausible” (p. 7, line 125) or “it is also possible” (p. 7, line 127) that nearly significant point estimates were censored less than null results or that positive results in favor of an internal focus were least preferred. Given the comprehensiveness of the Mathur and VanderWeele (2020) method, these are rather thin possibilities potentially floated to provide an otherwise elusive rationale for this re-examination. Although Chua et al. (2021) clarified the use of visual inspection of funnel plots and quantitative assessment of their symmetry or asymmetry as essentially an evaluation of the presence or absence of small-study effects (as described in the method section and specifically reported as such in the results section), and publication bias assessment was supplemented with the use of additional analyses in the form of removing “high-risk” studies (i.e., with extreme effect sizes), trim-and-fill analyses, p-curve analyses, and the selection model-based approach of Mathur and VanderWeele (2020), the current authors seemed to have misrepresented the previous authors by selectively focusing only on the assessment results obtained from the analyses of funnel plots to infer their conclusion(s) in regard to publication bias (p. 6 lines 94-106). Most importantly, given the rather substantial difference in findings of the Chua et al. (2021) and the current authors’ analyses, and existence of substantial meta-analytic evidence (i.e., from six other meta-analysis papers) in support of Chua et al.’s findings, this discrepancy deserves much stronger explication. And, a whole field is condemned with “Our findings impugn the evidential basis for the superiority of an external focus of attention. However, rather than establishing nil or trivial benefits from focusing externally, uncertainty remains. The posteriors include interesting effects, there may be important moderators, and our estimates may have overcorrected for bias. We simply do not know if an external focus provides meaningful benefits to motor learning and performance or not.” (Current authors, p. 22-23, lines 398-403). These conclusions implying a downgrade of evidential quality from the lack of determinism (i.e., “uncertainty remains”) to total non-existence (i.e., “simply do not know”) seem groundless and contradictory in consideration of the current authors’ citing seven meta-analyses showing findings favoring the use of an external focus of attention focus as synthesized across a myriad of studies published over at least two decades, and none providing synthesized results pointing in the opposite direction. Additionally, large meta-meta-analysis studies, each synthesizing more than 60 meta-analyses of peer-reviewed studies across various research fields, have been published to provide supporting evidence of the absence of any publication bias that should easily alarm the scientific community into doubting meta-analytic results (Mathur & VanderWeele, 2020; van Aert et al., 2019). Hence, there is no clear objective that can address an obvious research gap to justify the use of another meta-analytic method to re-examine a large meta-analytic dataset, other than perhaps the apparent desire to apply a different meta-analytic technique for the sake of its relative novelty to meta-analysts who are mostly more familiar with frequentist formulations.

This is our response

Assignment: **Brad**

[RC 3.7.] It can be exciting, and sometimes valuable for addressing gaps in knowledge, that meta-analytic researchers can use subjectivity to construct the parameter distribution prior to arriving at conclusive beliefs about an emerging or less studied field of research. This approach can provide important insights especially if the field is relatively new and extant literature has not accumulated a sufficient amount of empirical evidence conclusive. The Bayesian method fundamentally requires the analyst to think of all possible scenarios to create a distribution of parameters of scientific interest in a well-defined hypothesis space because the accuracy of the posterior distribution (and thus, useful validity of models) depends on the complete coverage of approximations including all alternatives that could possibly exist. Realistically, however many different specifications any analyst can think of, there will always be others that do not come to the mind, but cannot be dismissed “before the event”, due to the inherent limitation of the Bayesian approach. In light of this non-trivial technical problem, the current authors should elaborate on the need to introduce a priori guesswork through the use this beliefs-based method in the current study despite the existence of substantial corpus of research-based evidence in regard to the direction of attentional focus effects on motor performance and learning from more than two decades of empirical work contributions from the scientific community. Additionally, for the current authors to go ahead with using a Bayesian approach despite the existence of substantial empirical evidence (synthesized or otherwise) and admitting that “the true model is unknowable” (p. 7 line 139), they should provide details on the prior distribution that they started with whose support indeed covers all hypothetical approximations and their alternatives that could be covered to justify the results obtained from the Bayesian models thus produced that are supposedly not mis-specified. For example, important information such as the μ value in relation to the point, or otherwise, prior distribution on the effect size, the type of prior distribution on Hedges’ g effect size used to represent the hypothesis about the presence, or not, of the effect of interest so that the range of effect sizes can be assessed by readers to be typical or not for the field of attentional focus, τ value for point, or otherwise, prior distribution representing the hypothesis that heterogeneity is absent, or not, specifications of weight functions in terms of cut-offs, p -values, and direction of the effect, Cauchy distributions used in the PET and PEESE methods, are missing in this manuscript version. Also provide the sensitivity analysis results for checking the impact of using different priors.

This is our response

Assignment: **Brad**

[RC 3.8.] There seems to be an obvious and important approach to reducing the uncertainty—instead of employing (perhaps only nominally) dueling mathematical/statistical meta-analytic techniques to the published literature regarding the attributes of the unpublished (gray) literature, amounting to an innuendo of sorts, formally conduct a careful, comprehensive, and systematic review of the actual unpublished literature (e.g., Willett et al., 2019). Databases of the gray literature, such as “ZETOC,” “Conference Proceedings Citation Index”, and “ProQuest Dissertations and Theses Global” can be used. Combine it and/or compare it to the relevant published literature. The null hypothesis of a publication-only focused review on publication bias might be that there is no evidence of bias, thus, regardless of pre-registration, it is important to do more than cast every net on the basis of the published literature only.

This is our response

Assignment: **Brad**

[RC 3.9.] For the evaluation of potential post hoc selection bias, the current authors used the approach of synthesizing non-primary outcomes reported in the original articles, but not selected by Chua et al. (2021) for the calculation of effect size estimates. This re-examination of effect size estimates, however, was conducted only for outcomes of retention tests (p. 8 line 160, p. 9 lines 169 and 178, and p. 16 lines 313-317) with no elaboration on the basis for choosing this particular subset of outcomes, and omission of outcomes of the other type of motor learning test (i.e., transfer) as well as other test domains (i.e., performance, electromyographic activity, and distance effect).

This is our response

Assignment: **Brad**

[RC 3.10.] For the re-analysis, why were data selectively extracted for up to only three additional outcome variables for the retention test in each experiment/study? If the objective of this re-analysis is to detect potential selection bias introduced in Chua et al. (2021), why did the current authors choose to exercise selection themselves, rather than avoid this process altogether by re-analyzing all outcome variables reported in the original studies included in the meta-analyses?

This is our response

Assignment: **Brad**

[RC 3.11.] A “preregistered priority list” (p. 10 line 205) was used to form the decision matrix for guiding the selection of three supposedly “appropriate” outcomes if more than four outcomes were reported in a particular study. Assuming the appropriateness of such a decision matrix in the current authors’ re-analysis approach, does it have adequate complexity to allow suitable applicability for all types of task goals (e.g., distance in jumping tasks, movement time or speed in running tasks, muscle activity in coordination tasks, oxygen intake in treadmill walking tasks)? The implied conceptual difference between “outcomes most connected to the goal of the task” and “outcomes only correlated with success” is unclear without further elaboration (p. 10 lines 208-209). Is it not a given that task success should be assessed according to task goal?

This is our response

Assignment: **Brad**

[RC 3.12.] Were the outlier studies identified via “externally standardized residuals and Cook’s distances” (p. 11 lines 224-225) for removal from the sensitivity analysis in addition to, or a subset of, those outlier studies removed in Chua et al. (2021)? For clarity, the outlier studies for each meta-analysis (p. 13 line 269, p. 15 line 279, p. 15 line 289, and p. 16 line 304) should be listed so that readers could refer to their sample sizes and effect sizes for better interpretation of the meta-analytic estimates of the mean values of the true effect distributions. What proportions of the five datasets used by Chua et al. (2021) do they constitute? Discuss if their removal could have influenced the current meta-analytic estimates, degrees of heterogeneity, and publication bias assessment results to be different from those of Chua et al. (2021).

This is our response

Assignment: **Brad**

- [RC 3.13.] In relation to Point 6, missing information, or at least insufficient details, on the hypotheses used in the Bayesian methodology of the current study precludes a more comprehensive review of the findings reported in the current study. This includes the statement “our results suggest that an internal focus may be superior to an external focus in nearly as many situations as the reverse” (p. 21 lines 357-358). Were hypotheses other than the null hypothesis of no effect showing an external focus being more beneficial than an internal focus used? An apparent understanding of that statement is that this current study found meta-analytic estimates and Bayes factor values in support of an alternative hypothesis of an internal focus being more beneficial than an external focus in each of the five re-examined meta-analyses, but no such results were reported anywhere in this manuscript version. Secondly, only one of five g values is negative (i.e., in the opposite direction to the expected effect), but even so, the 95 % credible interval of -0.38 to 0.30 denotes no statistically significant difference. Please clarify.**

This is our response

Assignment: **Brad**

Minor issues

- [RC 3.14.] The statement regarding “attentional focus literature has not made use of preregistration or Registered Reports” (p. 8, line 155) should be clarified in terms of scope (e.g., studies assessing effects of all types of attentional focus, attentional focus studies that assessed effects on motor performance and/or learning) and substantiated with the provision of reference(s) from which such a conclusion can be drawn.**

This is our response

Assignment: **Brad**

- [RC 3.15.] Which two authors extracted all the additional data (p. 10 lines 202-203)? For transparency of author contributions, provide their initials in the manuscript.**

This is our response

Assignment: **Brad**

- [RC 3.16.] For required data that were not reported in the text of the included articles, or shared in public repositories, statistical information on the use of indirect sources and/or plot digitization should be provided. For example, the number of studies for which plots were digitized, and for which authors were contacted (including success rates).**

This is our response

Assignment: **Brad**

- [RC 3.17.] Without clarifying the ranges of values for interpretation of Bayes factors, it may be difficult for readers to assess the impact of their magnitudes. For example, how should readers interpret the lower and**

upper bounds of the Bayes factors (i.e., 1.33 and 5.74) for measuring the evidence favoring no benefit of external focus of attention over internal focus of attention?

This is our response

Assignment: **Brad**

[RC 3.18.] Consider defining “mu” before introducing it in symbol form (p. 13 line 264).

This is our response

Assignment: **Brad**

[RC 3.19.] What does the missing posterior plot for EMG mean in Figure 1?

This is our response

Assignment: **Brad**

[RC 3.20.] With the removal of one outlier study for the distance effect meta-analysis, was a total of only nine effect size estimates used? If so, discuss the limitation regarding uncertainty (high or low) in the estimated weights used for further analysis. Likewise, elaborate on the use of other low numbers of effect size estimates in terms of the affected analyses and their respective interpretation regarding the certainty of the results they contribute to.

This is our response

Assignment: **Brad**

[RC 3.21.] Do “previous studies” refer to meta-analytic studies as well as those primary studies that they synthesized (p. 20 line 339)? Did all these studies not report bias-corrected estimates as their primary estimates for the current authors to conclude that “reporting bias was ignored”? For example, in the set of meta-analyses that the current study re-analyzed, Chua et al. (2021) used meta-analytic estimates after bias-correction (via the removal of outliers with extreme effect sizes and checking the results using sensitivity analyses of publication bias) in their reporting of primary results.

This is our response

Assignment: **Brad**

[RC 3.22.] The phrase “somewhat larger” can neither be quantified nor qualified because “somewhat” is subjective and vague such that readers cannot interpret objectively, and results reported by the current study showed that there is no statistical difference.

This is our response

Assignment: **Brad**

[RC 3.23.] Elaborate on “20 studies missing due to insufficient information to calculate effect sizes in the original data set and another four missing effects from our extraction of secondary outcomes” (p. 22 lines 385-386). They should be identified and categorized according to which analyses in the current study were affected. What were the reasons? Consider using an appendix or supplementary document for these purposes.

This is our response

Assignment: **Brad**

[RC 3.24.] The statement “our analysis only investigated whether instructions impacted performance” brings forth ambiguity. Did the current authors use only a subset of the included studies in Chua et al. (2021) because they were not interested in the data from studies that induced attentional focus through feedback?

This is our response

Assignment: **Brad**

References