

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

To Dr. Paul Verhaeghen,

Thank you overseeing the review of our manuscript and the opportunity to resubmit our work for publication in *Psychological Bulletin*. Based on comments from the Reviewers 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 are included in this response letter (in a textbox) and appear in blue text in the revised manuscript.

We would like to emphasize that the primary purpose of our paper was not meant as a commentary on Chua et al. (2021). In our response letter, we have described different points where we refrained from taking a commentary approach (e.g., identifying and handling influence cases). As described in the manuscript, Chua et al. (2021) was the most comprehensive of the seven previous meta-analyses. This made Chua et al. a logical choice for re-analysis. Each of these previous seven meta-analyses had the same general conclusion of an external focus advantage. Yet, this conclusion in each meta-analysis was based on the results of naive random effects models that assume zero reporting bias in the primary estimates. We do not think this is a reasonable assumption for motor learning based on recent research, which we have made more clear in the revised manuscript, as well as highlighting key limitations of these models in response to Reviewer 1's comments. When reporting bias was accounted for, we are the first of these meta-analyses to arrive at a novel conclusion based on estimating trivially small effects for external focus of attention. In line with Reviewers 1 and 2, we believe this is not only an important finding, but it is a finding that is consistent with recent motor learning metascience (e.g., McKay et al., 2023).

As our meta-analysis encompasses all the other meta-analyses and so we do not think a meta-review would add new information. Additionally, the Reviewer comments themselves highlight the plausibility of reporting bias in this literature. Reviewer 3 is clearly protective of the mainstream view of external focus and put great effort into making a case for rejecting our paper. We have carefully refuted each of Reviewer 3's points in our point-by-point reply. Regardless of the merit of their points, the zeal with which they critiqued our paper is an illustration of how past studies that did not support the dominant narrative may have been reviewed, and in many cases, censored/rejected.

Thank you again for considering our work and overseeing the review process.

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.

Thank you for the overall positive assessment of our paper and its importance, novelty, and timeliness for the field of motor control and learning.

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

Thank you for the positive comment.

[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.**

Thank you for the positive comment. We are in complete agreement that this should become the standard.

[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.

Thank you for this suggestion. We have added this information to the manuscript:

On page 6, lines 97-100:

Importantly, a random effects model assumes no reporting bias and has been shown to be quite biased in the presence of selective reporting for statistical significance (Bartoš, Maier, Shanks, et al., 2023; Bom & Rachinger, 2019; Carter et al., 2015, 2019; Kvarven et al., 2020; Stanley et al., 2017, 2022).

[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).

We have added the following sentences to the Discussion section:

On page 21, lines 340-341:

However, it has become clear that such an assumption may not be appropriate for motor learning research (McKay, Yantha, et al., 2022; McKay et al., 2023).

On page 21, lines 353-358:

This is a serious limitation of the previous external focus meta-analyses as simulation studies have clearly demonstrated that random-effects result in large biases and high rates of false positives in the presence of publication-selection bias (Bartoš, Maier, Shanks, et al., 2023; Bom & Rachinger, 2019; Stanley et al., 2017, 2022), which have been further supported when random-effects are compared with preregistered multilab replications (Kvarven et al., 2020).

[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?

Thank you and we agree with you about the importance of this becoming the default. We have added the following to the recommendations section:

On page 24, lines 426-430:

Moving forward, it will therefore be critical for researchers conducting meta-analyses in motor learning and related areas (e.g., psychology, neuroscience, sport and exercise science) to adopt the use of leverage statistics as the default approach for identifying outliers (see Deeks et al., 2023; Viechtbauer & Cheung, 2010 for discussions).

[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.

We have added the following on to the Discussion:

On page 22, lines 365-369:

Thus, similar to previous simulation studies our findings illustrate that reporting bias can cause random effects models to produce even large effect estimates when the true model is null. The random effects estimates reported by (Chua et al., 2021) ranged from small to large, while our corrected estimates range from essentially nil to trivial at best.

[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.

Although we agree with your view of science and in particular the importance of "calling out" theories in the

field that do not appear to be well-supported by the evidence when that evidence has been synthesized and evaluated appropriately. However, a more general commentary on OPTIMAL theory is beyond the scope of the current paper as this work only focused on one of the three pillars within the theory. However, we believe such a commentary is necessary and have such a paper in the works currently.

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

Thank you for your helpful comments. We feel it has helped to strengthen our paper in terms of its importance, novelty, and theoretical contribution to the field.

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.

Thank you for the positive comments about our paper. We have made some minor additions to the revised manuscript based on your helpful comments.

[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.

We share your view regarding OPTIMAL theory as more of a framework at best rather than a proper theory. We focused our discussion on OPTIMAL theory rather than a more broad discussion that included other perspectives that we highlighted in the introduction for two main reasons. First, OPTIMAL theory was highlighted as a leading explanation for the supposed benefits of an external focus in the Chua et al. paper. Additionally, the senior author on the Chua et al. paper is also one of the authors of OPTIMAL theory so this is perhaps not too surprising. Second, an impetus for our reanalysis of Chua et al. was the fact that

findings from recent meta-analyses really questioned the evidential value of the findings that have been used as support for the two other pillars of OPTIMAL theory. Given this, we felt it was critical to ensure the impacts of reporting bias were also explored for the external focus literature. As such, and in line with one of our responses to Reviewer 1, a more broad discussion of the impact of our findings for different theories is beyond the scope of the current paper. Additionally, in a more general sense our conclusions on page 23 in the original submission with respect to OPTIMAL theory are applicable to any other theories and explanations that have relied on the external focus literature to date as support. That is, the mechanisms in any of these theories could still be valid; however, it is the literature itself that lacks evidential value for any of the possible theories. We have added a footnote in the revised manuscript to better capture this.

On page 23:

This conclusion also applies to other theories and perspectives (e.g., ecological dynamics) that have been forwarded based on the extant attentional focus literature to account for a supposed external focus advantage (e.g., Davids et al., 2003; Gottwald et al., 2023; Hommel et al., 2001; Prinz, 1990; Wulf & Prinz, 2001).

[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.

We agree that there has been relevant empirical work providing mixed results and we have now made reference to some of this work in a footnote (see below). We have kept the use of consensus as we believe this is an accurate reflection of the field (e.g., how external focus is presented in the main undergraduate and graduate textbooks), but this is why we also tried to temper this statement with “at least most of the time”. We now hope that the novelty and importance of the current paper will be the catalyst in changing this consensus.

On page 5:

While we acknowledge this as the general consensus in the field, it is important to note that there are mixed findings and alternative discussions in this area of research (e.g., Bernier et al., 2016; Brick et al., 2014; Canning, 2005; Collins et al., 2016; Emanuel et al., 2008; Lawrence et al., 2011; Maurer & Munzert, 2013; Peh et al., 2011; Perkins-Ceccato et al., 2003; Schorer et al., 2012; Zentgraf & Munzert, 2009).

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.

Thank you for the extensive review of our manuscript. Due to the length of each major issue highlighted in your review, we have included responses inline where warranted.

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?

The current study was an analysis of the two decades of experimental results that Reviewer 3 references. It is not appropriate to use the results of the previous meta-analyses, all of which are (mostly) embedded in the dataset reported by Chua et al., to form the prior for an analysis of those very same results. Instead, we must consider what our prior would be had we not yet seen these two decades' worth of evidence and the seven previous meta-analyses that synthesized them.

It is critical to consider the standard design of the experiments in question. Participants are asked to practice or perform a skill with the only difference between groups or conditions being a couple of words. For example, the difference between asking participants to focus on keeping their feet level versus keeping the markers on a stabilometer level. For an insight as to what prior an expert in motor learning would have held before the data in question had accumulated, consider the response of the reviewers to the first study in Wulf, HoB, & Prinz (1998). Here is an anecdote from Dr. Wulf:

"Obviously, we were excited about these results. Yet when we first tried to publish this study, the reviewers were skeptical. Admittedly, it was hard to believe that such a small difference in the focus of attention could have such a strong influence on performance and learning. Among those reviewers of the manuscript was a friend and colleague from Texas A&M University, Professor Charles Shea. To our disappointment, he asked us to try to replicate the findings with another task before he would be able to recommend the study for publication." (Wulf, 2007, pp.36-40).

We echo the skepticism of Dr. Shea in the above quoted. Indeed, our own prior was not that there was an even chance of benefits or no effect, but rather that it was more likely that the subtle differences in instructions or feedback studied in this literature were largely unimpactful. Nevertheless, we chose a neutral prior and allowed the data to dominate.

Regarding the presence of heterogeneity, we believed that there was most likely heterogeneity among study effects in these analyses. However, we also note that Chua et al. reported that there was no significant heterogeneity among performance effects following outlier removal. Again, we chose a neutral prior that did not tilt the odds in favor of either hypothesis. Nevertheless, we wish to stress that an important purpose of a meta-analysis is to estimate the extent of the heterogeneity. While the RoBMA approach does provide a statistical test of this hypothesis (as does a Q-test in frequentist treatments), the more important result here is the tau estimate.

Did we think it was equally likely there would be evidence of reporting bias compared to not? No. As we noted in the introduction, previous meta-analyses have found substantial evidence of reporting bias among studies that form the other two pillars of OPTIMAL theory, and external focus forms the third. Again, we adopted a conservative approach and specified a neutral prior that did not tip the odds in favor of finding evidence of reporting bias.

[RC 3.3.] 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.

Please note, the model-averaged posteriors generated by the RoBMA analysis employed in this study will accurately recover the true effect size in an unbiased sample of studies, as can be demonstrated via Monte Carlo simulation (see Figure 1 below). In this simulation—reported previously in McKay et al. (2023)—49 studies with a mean effect of $d = .54$ and heterogeneity of $\tau = .35$ were simulated in the absence of reporting bias. These studies were then analyzed with the same robust Bayesian meta-analysis approach used in the current study. Note that the posterior distribution is centered on the true effect and shows almost no density on the spike null. The analysis does not favor the null nor does it systematically deflate point estimates. It is the consistency of the observed results with our models of reporting bias that causes the model-averaged posterior to favor the null, not the use of neutral priors.

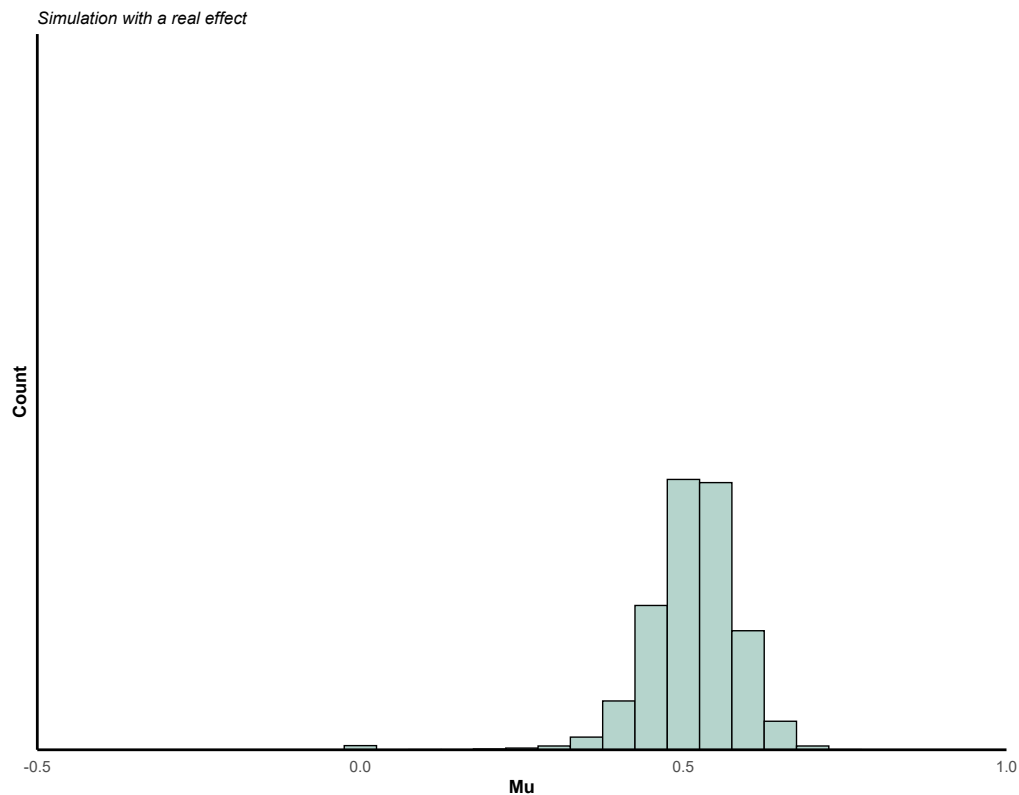


Figure 1: Posterior simulation with a real effect

We conducted sensitivity analyses with what we would describe as “optimistic” priors (see Table 1 below). We set the prior expectation that there is an effect of external focus as twice as likely as no effect, and we centered the prior effect distributions on the primary estimates reported by Chua et al. As we have argued above, these are not the appropriate priors for our primary analyses. Nevertheless, these optimistic priors are more aligned with Reviewer 3’s position and allow us to evaluate whether adopting such priors would substantively change the results. While the results of each analysis changed slightly, the qualitative inferences remained unchanged. This is not surprising. There is sufficient data in each analysis to dominate the priors.

Table 1: Results of RoBMA analyses with neutral and optimistic priors

Analysis	<i>Mu</i> Neutral	<i>Mu</i> Optimistic	<i>BF</i> ₁₀ Neutral	<i>BF</i> ₁₀ Optimistic
Retention	<i>g</i> = .15	<i>g</i> = .26	.75	.795
Performance	<i>g</i> = .01	<i>g</i> = .02	.174	.172
Transfer	<i>g</i> = .09	<i>g</i> = .16	.568	.565
EMG	<i>g</i> = .06	<i>g</i> = .13	.474	.413
Distance	<i>g</i> = -.01	<i>g</i> = -.01	.255	.249

Note. Neutral priors for the effect of attentional focus were 50% probability of an effect \sim Normal(1,0), 50% probability of a point null $d = 0$. Optimistic priors adopted the primary point estimate reported by Chua et al. (2021) for each analysis as the mean of the Normal distribution \sim Normal(Mean of Chua, 1). The optimistic priors also gave two to one odds that an effect was present.

The priors we adopted for heterogeneity are based on empirical evidence reported by van Erp (2017). Specifically, the prior for tau was an inverse gamma distribution with shape = 1, scale = .15. We tested the impact of changing this prior to a half-normal distribution with mean = .1, sd = 1, and it had almost no impact on the tau estimate (originally .400, changed to .408 in analysis of performance data). Again, this is not surprising. The priors included for both effect and heterogeneity encompassed all plausible values.

With respect to the prior for reporting bias, as we state in the manuscript, the random effects estimates reported in previous meta-analyses assume no reporting bias. Our prior gave a 50-50 chance of reporting bias, despite ample reason for thinking it was more likely. While the results would naturally be different with different model specifications, our results would only become more pessimistic about the external focus effects if we increased the prior odds of reporting bias. Decreasing the odds of reporting bias would increase the optimism for an effect, up to the limits reported by Chua et al.

[RC 3.4.] 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.

Model convergence checks were conducted for all analyses and have now been summarized in a footnote on page 14. Model performance was weighted based on how well they predicted the data (note, all models were included in Bayesian model averaging). This was explained in the methods section (now page 13, lines

261-265 in the revised manuscript). We included these model comparisons for all five analyses in aggregate in Figure 3. To facilitate the examination of model comparisons for each specific analysis, we have included five new supplemental tables (referred to in the revised manuscript as Supplementary B). The current class of models was chosen because they model specific reporting bias mechanisms. The PEESE model often fit the data quite well, but even when it did not, we are not going to fit any sort of expanded models that are not based on a theoretical or causal model of reporting bias.

[RC 3.5.] 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?

We thank the Reviewer for their attention to detail. This apparent discrepancy is due to rounding. We report the Bayes factors to two decimal places in the manuscript but used three digits when converting BF10 to BF01 in the abstract. We did make an error, as we failed to round the BF01 value up to 5.75. This has been fixed. We reported the results with all included studies in the abstract because our conclusions were not sensitive to their exclusion. We have no reason to expect the effect sizes calculated for the outlier studies to be in error, and simply being discordant with the remaining studies is not a justifiable reason for exclusion.

Given the length constraints of the abstract, we chose to focus on the primary results readers will be interested in. None of the results reported in the abstract were chosen based on “significance”. While we feel our investigation of dependent variable selection was important and adds to our understanding of the root causes of reporting bias in the external focus literature, the results of this analysis do not modify the key takeaways most readers will have from this manuscript.

[RC 3.6.] 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.

We respectfully disagree and reject the premise that “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.” We are not sure where this impression came from, but it is simply not true. For a few counterexamples published in *Psychological Bulletin*, see Liu & Almeida (2023; <https://doi.org/10.1037/bul0000393>), Cafferata, Hicks, & von Bastian (2021; <https://doi.org/10.1037/bul0000325>), and Ragni & Johnson-Laird (2018; <https://doi.org/10.1037/bul0000146>).

[RC 3.7.] 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).

We use the term “reporting bias” consistent with its usage in the Cochrane handbook linked here: (https://handbook-5-1.cochrane.org/chapter_10/table_10_1_a_definitions_of_some_types_of_reporting_biases.htm). Importantly, all reporting bias is due to missingness – whether because of publication status, failure to report certain outcomes, failure to report sufficient information to meta-analyze certain outcomes, reporting in a language not accessible to the meta-analyzers, or even failing to report results with preplanned analysis methods and sample size, opting instead for an analysis or sample size that produces a specific result. Our analysis attempted to model specific reasons for this missingness.

We have added the following to the revised manuscript:

On page 5, lines 81-84:

Note that reporting bias encompasses various forms of selection bias that limit the availability of data. Potential reporting bias mechanisms can be modeled, though models cannot determine the specific reason for censorship within a literature.

Our analysis of the results selected by Chua et al. examined the evidence that Chua et al.’s selection process

biased the results, an effect we failed to find support for. This analysis does not address or rule out the possibility of outcome reporting bias at the level of the original studies.

[RC 3.8.] 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 Mathur & VanderWeele method does not rely on empirical data from previous meta-analyses and several fields to determine the amount of publication bias required to attenuate the observed effect estimate to the null. The method (as applied by Chua et al.) assumes all publication bias is based on whether a result is statistically significant in the positive direction. If we accept this assumption as true, then when we observe a positive effect among all non-significant results in the positive direction, we can conclude that no amount (of the assumed type of publication bias) could reduce the effect to the null. The logic here is straightforward: there is already a positive effect among the ‘remaining’ results, therefore it does not matter how much less likely they were to be published than significant results, there would still be an effect. We simply take issue with the assumption that the only form of reporting bias likely to be present in this literature is a preference for significant results in the positive direction. We believe it is highly unlikely for that assumption to hold, and if it does not, then the conclusion that “no amount of publication bias could attenuate the effect to the null” is no longer justified. Indeed, our analysis highlights how untenable this conclusion is.

[RC 3.9.] 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.”

We are not sure who Reviewer 3 is quoting here. This quote is not from our manuscript or from Chua et al. Are they quoting Mathur & VanderWeele? If so, how does quoting them give cause for us to acknowledge their quote?

[RC 3.10.] 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.

The Mathur and VanderWeele method as applied by Chua et al. does not provide evidence for or against the possibility of reporting bias based on the direction of results, non-significant trends, sample size, or otherwise. It is simply a version of a 3-point selection model with a $p = .05$ cut-point. As can be seen in Figure 3, we observed ample evidence that models accounting for these additional sources of reporting bias provided a better fit to the data than a 3-point selection model.

It is worth noting that the primary estimates from the Chua et al. meta-analysis as well as the six others were all based on random effects estimates that assume no reporting bias. As we discussed in the introduction, there was substantial reason to suspect reporting bias in this literature based on previous meta-analyses of this field, as well as of the two adjacent pillars of the OPTIMAL theory. We feel this was a strong rationale to conduct this re-examination.

[RC 3.11.] 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.

It is important to note that Chua et al. found evidence consistent with reporting bias in their assessment of funnel plot asymmetry, their Mathur & VanderWeele selection model, and their trim-and-fill analyses. We did not “misrepresent” the previous authors by noting they found evidence of reporting bias.

We chose not to spend time criticizing Chua et al.’s use of trim-and-fill, p-curve, or their approach to outlier assessment and removal because the purpose of this study was not to provide a commentary on Chua et al.’s work, but rather to assess the evidential basis for external focus superiority. However, please note that trim-and-fill is not a reliable method of correcting for reporting bias (see Carter et al., 2019). The p-curve results reported in the supplement appear problematic – the effect sizes often increased to implausible sizes (i.e., $d = 1.146$, $d = 1.163$, $d = 0.916$, $d = 0.943$, $d = 0.816$, $d = 1.586$, $d = 1.412$). Without access to the analysis script, it is difficult to know why correcting for reporting bias resulted in larger effect size estimates in many cases. However, it is not plausible that reporting bias has deflated the meta-analytic estimates in this literature. Further, it is worth noting that p-curve is also a selection model with a $p = .05$ cut-point. Therefore, while Chua et al. should be commended for applying multiple types of reporting bias correction methods, their use of p-curve in addition to the Mathur & VanderWeele model does not address the additional reporting bias mechanisms that the present study found substantial evidence of.

Please be aware that all the previous meta-analyses analyzed subsets of the same literature and used the same method to form their primary estimate (random effects models). The existence of multiple meta-analyses of the same data with the same method reaching the same conclusion does not build the formidable counterargument that Reviewer 3 is implying.

[RC 3.12.] 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.

While uncertainty remains, it is important to note that our analyses found greater support for the point null hypothesis than for the presence of an effect of external focus.

As noted above, the previous meta-analyses of these data do not provide some cumulative counterweight of evidence against the results of our analyses. They share the same data (in many cases), the same methods, and the same weaknesses.

[RC 3.13.] 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.

We believe there may be some confusion about the threat of reporting bias and perhaps what the selection model employed by Chua et al. can provide. The papers cited above cannot be used to evidence a lack of reporting bias in the present literature because they do not examine it. The results of McKay and colleagues (2022) demonstrate the severity of reporting bias effects among literatures pertaining to the other two pillars of OPTIMAL theory, a tripartite theory that includes external focus as its third pillar. Lohse and colleagues (2016) provide stark evidence of reporting bias in the motor learning literature at large. Psychology as a field has been gripped by the effects of reporting bias and the so-called replication crisis for over a decade, at least (Bishop, 2019).

[RC 3.14.] 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.

We respectfully disagree about the purpose of meta-analyses. In particular, the notion that meta-analyses are best suited when the “extant literature has not accumulated a sufficient amount” is, frankly, preposterous.

While Reviewer 3 may think it impossible to imagine all possible values of the mean effect of external focus in the population, we believe $N(0,1)$ more than adequately captures all realistic possibilities. Even Reviewer 3 has previously suggested our prior for the effect was too “diffuse.” Which is it? Did we fail to think of every value imaginable or was our prior too diffuse?

[RC 3.15.] 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.

We have included detailed descriptions of our priors in a footnote on page 13.

[RC 3.16.] 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.

Unfortunately, exploring the gray literature is not a feasible solution to reducing uncertainty. First, the quality of unpublished studies is either difficult or impossible to assess, unlike the studies included in the present study that were scrutinized for risk of bias. The accuracy of meta-estimates depends on the quality of the included studies; as unpublished studies have not been vetted by peer review, they may be of a lower overall quality. This can cause the problem of so-called “garbage in, garbage out”. Second, the unpublished nature of the gray literature creates ‘location bias,’ a subtype of reporting bias that has been found to exacerbate bias in meta-analysis (Ferguson & Brannick, 2012; <https://doi.org/10.1037/a0024445>). Relatedly, the format that many unpublished studies take (i.e., conference abstracts) can limit access to all outcomes, including perhaps the focal outcomes, exacerbating outcome reporting bias.

We note that most of the previous meta-analyses, including the one by Chua et al., included a statistical examination of publication bias. Indeed Reviewer 3 has previously argued that the treatment conducted by Chua et al. is sufficient to rule out reason for alarm from publication bias. Our study challenges this conclusion and finds substantial evidence that assumptions made by Chua et al. do not hold, and when we account for more plausible forms of reporting bias there appears to be substantial cause for alarm.

[RC 3.17.] 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).

We have added the following rationale with a footnote in the revised manuscript:

On page 10:

We chose to focus on retention effects because the performance estimates were already small. The retention estimates were substantial, and retention tests are often the focal learning measure in an experiment. Almost all transfer tests were from studies that also included a retention test, so focusing on retention outcomes was the simplest way to test our research question.

[RC 3.18.] 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?

There are only so many ways one can measure motor performance relative to the goal of the task. By selecting up to three additional measures, we considered up to four total measures of performance for a given study. We set a limit on the total number of measures we would extract *a priori* and employed a priority list to ensure the measures we selected were most relevant because we wanted to avoid the possibility that a few studies would flood our analysis with dozens of kinematic outcomes. As it turned out, this would not have happened, but we did not know this ahead of time. There were only 3 studies that included more than four measures and none of the measures beyond the fourth could reasonably be considered a measure of performance relative to the task goal. These measures were kinematic (blade angle on a rowing task, elbow and shoulder angles on a juggling task) or procedural (did the server return to the baseline after completing a serve).

[RC 3.19.] 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?

We agree that task success should be assessed according to the task goal. This is why we applied this priority list, which was created *a priori* and used in previous meta-analyses (see McKay et al., 2022; Bacelar et al.,

2022). Reviewer 3 points to outcomes that are not indicative of performance relative to a task goal: Muscle activity is not the goal of a coordination task – though it may be measured and is potentially correlated to success. Likewise, oxygen intake is not usually the goal of a walking task (unless it is). Nevertheless, the priority list allowed for these measures that may only be tangentially related to task performance to be included as “other”. As mentioned in the previous response, we collected (nearly) all possible measures of performance and tested whether those measures selected by Chua et al. differed from the measures that were not selected. Our analysis failed to reject the null. This result is robust to whatever differences of opinion Reviewer 3 may have regarding the specifics of the priority list we applied.

[RC 3.20.] 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).

We have added a citation of each of the excluded outliers in the results section and a footnote that acknowledges the difference in approaches and the associated outcome.

On page 12:

Our approach to influential case screening differed from the approach employed by Chua et al. (2021) and we therefore arrived at a different number of outliers for each analysis (see Supplementary A for more details).

In both the Chua et al. study and our own, all studies were included in the first analyses. Subsequently, we differed in our approach to outlier detection and removal. We found four influential cases in the performance data; Chua et al. deleted 18. We found two outliers in the retention data; Chua et al. deleted 10. We found one outlier in the transfer data; Chua et al. deleted three. We found no outliers in the EMG data; Chua et al. deleted three. We found one outlier in the distance effect data; Chua et al. also deleted one.

We chose not to elaborate on the differences in our approach to the one employed by Chua et al. in the manuscript because we did not want to be overly critical of their work. The primary differences between our studies are not the identification and treatment of outliers, but the assumptions made about reporting bias and the subsequent methods used to model it. For the purposes of this response letter, we will detail the differences between our approach to outliers and the one employed by Chua et al. We believe Chua et al. employed an inappropriate approach and miscited Viechtbauer & Cheung (2010) in the process.

In their paper, Chua et al. describe their approach to outlier detection as the following (see pages 623-624): “Furthermore, to eliminate the influence of extreme ES values on the meta-analytic results, a statistical outlier analysis was conducted to determine values of overall Hedges’ g , Q , I^2 , and s^2 after the removal of individual ES both containing a Hedges’ g value with a 95% CI that falls completely outside of the 95% CI of the overall Hedges’ g value for the meta-analysis of interest (Viechtbauer & Cheung, 2010).”

This approach described above has the effect of substantially reducing heterogeneity in the dataset but not for principled reasons. For example, Chua et al. removed 18 out of 100 effects as supposed outliers from their analysis of performance data, changing their tau-squared estimate from .219 to .006.

Conversely, the paper by Viechtbauer & Cheung recommends identifying outliers using influence analysis (i.e., standardized residuals and Cook's d). This is the approach we adopted. Note, Viechtbauer & Cheung join many other authors in calling for outlier deletion as a sensitivity analysis, and one that should be employed judiciously (see, for example: <https://training.cochrane.org/handbook/current/chapter-10#section-10-10>)

[RC 3.21.] 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.

The statement quoted above is missing critical context from the preceding sentence: “This heterogeneity could imply that focus of attention has a range of effects that depend on situational factors. If so, . . .”

Heterogeneity means variability in true effects. Having large heterogeneity, *if due to situational factors*, implies that effects could vary widely in practice. Since the point estimates are all centered near zero, *if this heterogeneity is due to situation factors*, then we would expect effects in both directions with roughly equal likelihood. Therefore, “an internal focus may be superior to an external focus in nearly as many situations as the reverse.”

Minor issues

[RC 3.22.] 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.

We have clarified the scope of this statement by substituting “external” for “attentional”. This meta-analysis will be the reference that substantiates the claim because it is our review of the extant literature that revealed this.

[RC 3.23.] Which two authors extracted all the additional data (p. 10 lines 202-203)? For transparency of author contributions, provide their initials in the manuscript.

We have added the initials of the six researchers who extracted data on page 10. We have also changed the sentence to: “Data were extracted in duplicate by a team of six researchers working independently.” to avoid confusion.

[RC 3.24.] 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).

We have added the number of studies requiring plot digitization and the number of authors emailed along with the success rate. It is not typical to name (and potentially shame) the authors who were contacted and did not respond and we will refrain from doing so here.

[RC 3.25.] 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?

Bayes factors are intended to be interpreted as continuous measures of evidence. We provide the meaning of Bayes factors when we report the results (e.g., the data were over 5 times more likely under the null model than the alternative, etc). We do not wish to impose any arbitrary thresholds for interpreting the Bayes factors.

[RC 3.26.] Consider defining “mu” before introducing it in symbol form (p. 13 line 264).

We have defined Mu on page 14.

[RC 3.27.] What does the missing posterior plot for EMG mean in Figure 1?

There were no outliers identified in the EMG data.

[RC 3.28.] 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.

The fewer the studies included, the less powerful our model is. Smaller meta-analyses are less precise and are therefore less able to provide convincing evidence for the presence of an effect, reporting bias, or heterogeneity.

Uncertainty in each parameter estimate is captured by the 95% credible intervals we report. Note that the width of the 95% credible interval differs between analyses with several studies (performance, for example) and analyses with fewer studies (distance effect).

[RC 3.29.] 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.

We have substituted “meta-analyses” for studies for clarity. We have also changed “ignored” to “unaccounted

for.” The primary estimates reported in the abstract of each study were from random effects analyses. These analyses do not correct for reporting bias. Removing outliers does not correct for reporting bias. Conducting sensitivity analyses, but reporting uncorrected estimates in the abstract does not correct for reporting bias.

[RC 3.30.] 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.

No statistical difference does not mean no actual difference. We do not treat null results as equivalent to evidence of no difference. Readers can interpret the difference we observed by looking at the data, including the corresponding confidence intervals.

[RC 3.31.] 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.

We have added supplementary data (referred in revised manuscript as Supplementary B) indicating which studies had insufficient statistical information for inclusion in the original dataset, as well as the studies that did not have sufficient information for inclusion in the present study.

[RC 3.32.] 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?

We have added “or feedback”.