Dear Dr. Zoltan Dienes,

Thank you for allowing us to submit the 2nd revised version of our Stage-1 manuscript titled "Do task-irrelevant cross-modal statistical regularities induce distractor suppression in visual search?" to PCI RR.

We would like to thank you and the reviewers for their constructive comments and helpful suggestions. Below you can find our responses to all comments in bold.

We have submitted the revised Stage-1 Registered report (file name: "Registered_Report_Stage-1_Proposal_v3.pdf"). We have also uploaded a PDF document indicating modifications in Tracked changes.

We look forward to your and reviewer's comments.

With kind regards,
Kishore Kumar Jagini (on behalf of authors)

Recommender

The reviewers are largely happy with your changes. Vadillo raises a couple of points, one of which I want to highlight here - namely how the figure for an effect size of 0.45 in particular can be justified. I realize in almost all other papers which are not RRs no one really justifies their effect sizes used in power analyses. But we as a rule do for Registered Reports. Thus, while I realize you are already running more subjects than typical, there remains the point that a non-significant result only counts against there being any effect of interest for the H1 in question if the study was well powered for detecting such effects. Thus the power analysis is only as good as the reasons relating the minimally interesting effect size to the scientific problem in question. It is only by addressing this problem that you can justify rejecting your H1. One heursitic in the paper I previously refered you to is to use the lower limit of a confidence interval on the effect from relevant previous studies - if the lower limit is still interesting, then there is a case for that being the smallest effect of interest that is plausible (roughly treating the CI as a credibility interval). Or you may think about it some other way.

(The Meyen method for equating direct and indirect task performance that Vadillo refers to assumes equal signal to noise ratio for each trial for the tasks, which is implausible - it makes the same assumption for trials that Vadillo points out shouldn't be made for tasks, so repeats the same issue at another level.)

Response: Thank you for suggestions. Based on your suggestion, we revised the manuscript to include appropriate justification for effect sizes for each proposed hypothesis test. Relying on the effect size from the previous study at the face value for an a priori power analysis is not recommended, as this might lead to underpowered studies (Dienes, 2021; Perugini et al., 2014). To guard against the underpowered study, we determined the smallest effect size of interest as the lower-bound limit of the effect size by following the advice of Perugini et al. (2014). Using the determined smallest effect size of interest, we conducted an a priori power analysis. Please see the uploaded PDF document indicating these revisions in Tracked changes.

Reviewer #2

The authors have done an excellent job at addressing my comments to the previous version. I appreciate in particular that they are now willing to test a substantially larger number of participants and that the ms now addresses the question of the low sensitivity in awareness tests. I only have relatively minor comments to the present version:

On page 7 the authors write "We hypothesise that if the participants are aware of the the relationship between auditory and visual distractor location regularities, we expect that the score received by each location linearly decreases from its distance from the actual HpValD location". The sentence sounds a bit mysterious because the reader still has no clue as to how locations will be scored. This is not explained until page 15.

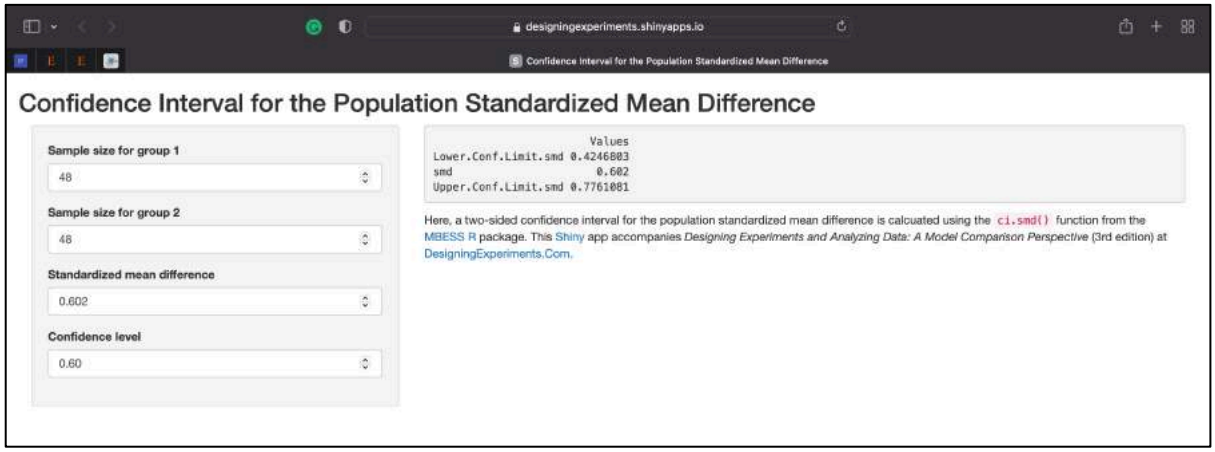
Response: Sorry for the confusion. We have revised the manuscript accordingly. Please see the uploaded PDF document indicating revision modifications in Tracked changes.

The new power calculations basically assume that the expected effect size is roughly the same in reaction times and in the awareness test (i.e., d = 0.45). But the former are measured over hundreds of trials and the latter are measured in just six questions. Implicitly, this means that the authors expect each question of the awareness test to be much more sensitive and informative than each trial of the visual search task, which is an arguable assumption, in my opinion. I am not asking the authors to make any change in the ms regarding this. I am just trying to highlight a recurrent problem in this area of research. There is a great paper about this problem by Sascha Meyen in JEP:General. https://www.tml.cs.uni-tuebingen.de/team/luxburg/publications/MeyenEtal2021.pdf

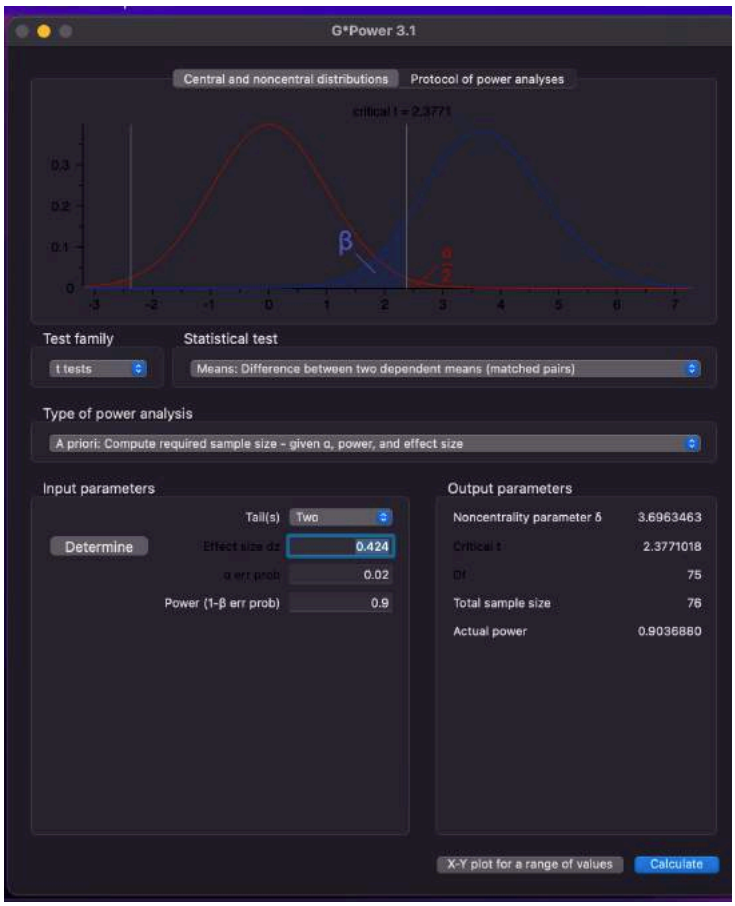
Response: Thank you for pointing out the problem of assuming the same effect size for both the proposed hypothesis tests. In the revised manuscript, we have determined the smallest effect size of interest for each hypothesis test separately. Then, we have conducted an a priori power analysis on the determined the smallest effect size of interest for each hypothesis test. Please see the uploaded PDF document indicating these revisions in Tracked changes.

Appendix:

Screenshot of Shiny R web app for estimating confidence interval for the effect size (standardised mean difference). Web app link: https://designingexperiments.shinyapps.io/ci_smd/



Screenshots of power calculation using G*Power 3.1



References:

Dienes, Z. (2021). Obtaining Evidence for No Effect. ***Collabra: Psychology*****, *7*(1). https://doi.org/10.1525/collabra.28202**

Perugini, M., Gallucci, M., & Costantini, G. (2014). Safeguard Power as a Protection Against Imprecise Power Estimates. ***Perspectives on Psychological Science : A Journal of the Association for Psychological Science*****, *9*(3), 319–332. https://doi.org/10.1177/1745691614528519**