

Dear Dr. Zoltan Dienes,

Thank you for allowing us to submit the revised version of our Stage-2 manuscript titled “No reliable effect of task-irrelevant cross-modal statistical regularities on distractor suppression?” to PCI RR.

We would like to thank you and the reviewers for providing suggestions to improve the quality of the manuscript. Below you can find our response to the review suggestions/comments in bold.

We have submitted the revised Stage-2 Registered report (file name: “Registered_Report_Stage-2_Version_3.pdf”). We have also uploaded a PDF document indicating modifications in the manuscript with Tracked changes.

We look forward to your and the reviewers’ comments on the revised manuscript.

With kind regards,
Kishore Kumar Jagini,
(On behalf of the authors).

Recommender

The reviewers have thoroughly reviewed your Stage 2 manuscript and make some very useful suggestions for your consideration. There is a question of whether you have evidence that the auditory stimuli were heard at all. One reviewer recommends Bayes factors; I prefer them myself to significance tests, but they do not substitute the pre-registered analyses of course, so you may include them in an exploratory section (you may find this useful: <https://psyarxiv.com/yqaj4>).

Make sure you report the actual t values in the text for the main t-tests you conduct.

In the Discussion, the issue of subjective vs objective tests could be considered in a more structured way. At the moment you first bring it up then side with objective tests; then in a subsequent paragraph starting "The asymmetry in participants’ objective measures of awareness of statistical regularities between Experiment 1 and Experiment 2 is unclear" deal with one issue, then revert somewhat unrelatedly to the general issue of subjective vs objective measures, siding with subjective measures. It would be better to discuss weakness and strengths of subjective vs objective measures in one place, then in a way consistent with that discussion, interpret your results.

Our Response: Thank you for the suggestions. We have updated the text based on review suggestions/comments in the revised manuscript. Also, we have uploaded the sound discrimination tasks’ raw data and scripts in the OSF archive associated with this project.

Reviewer #1

In the present version of the article, the authors report the empirical results of the two preregistered studies approved in Stage 1. The method and results sections follow the preregistered protocol and all the new analyses, not originally included in the protocol, are clearly identified as such. The general discussion is consistent with the analyses reported in the results section. Therefore, I think that the manuscript should be accepted for publication. I only have minimal comments that the authors might want to take into account for the final version.

In the general discussion the authors mention that one potential explanation of the results is that participants may have paid little attention to the task-irrelevant stimuli. Indeed, this seems like the most plausible explanation and an interesting idea for future research is to test this hypothesis again in an experimental task that ensures that sounds receive some attention. But I even wonder if participants heard the tones at all! The preregistered protocol mentions that participants would perform a test to ensure that they could discriminate the task-irrelevant stimuli, but I could not find any information about that manipulation check in the manuscript. Did I miss it? Or is it simply not reported? I think the final version of the ms needs to include some information about the results of this discrimination test.

Our Response: Thank you for the suggestions. We have now updated the relevant information in the revised manuscript. We apologize for missing out on the sound discrimination test results in the earlier manuscript. Overall, the mean percent accuracy of sound discrimination judgments was reasonably high in both experiments. Sound discrimination accuracy for Experiment 1 was: Mean (in %) \pm SEM: 97.8030 ± 0.3697 , while for Experiment 2 was: Mean (in %) \pm SEM: 90.5906 ± 0.6996 . This evidence indicates that the participants heard the sounds clearly and could discriminate them accurately.

On page 27, the authors conclude that “Moreover, these results suggest that, at least under the conditions of Experiments 1 and 2, the participants are unable to learn associations between the location of the visual distractor and the auditory stimulus.” In truth, it is arguable that this is the case, because the awareness test of Experiment 1 suggests that participants did learn something, even if this was not translated into faster responses in valid trials. Of course, another possibility (and I find this one tempting) is that participants didn’t learn anything at all about the sounds and that the results of the awareness test in Experiment 1 are entirely driven by the test question itself. That is, if the awareness test question asks participants to rank positions when the left tone was presented they can speculate (ad hoc) that the most likely locations might have been on the left. And the opposite for the right tone. In other words, performance in the awareness test might be entirely driven by inferences made during the test itself and not by anything learned during the visual search task. Note that this kind of inferences are easier to make in Experiment 1, where participants might assume that sounds on the left side might be associated with distractors on the left side. It is less clear how participants can make a similar assumption about high/low pitch sounds in Experiment 2.

I wonder if anything could be done to test for this possibility in Experiment 1. If participants performance is driven by inferences made at test, then it follows that participants will be more likely to mention locations on the right-hand side for the right-tone and on the left-hand side for the left tone, but within each hemifield, there is very little reason to expect participants to perform above chance. In other words, if the analysis is restricted to responses on the same side as the tone, participants might not show above chance performance. If they do, this would show quite convincingly that their responses were driven by something they learned during the first stage of the experiment and not something based on inferences made at test.

Our Response: Thank you for the suggestion. According to your suggestion, we have performed the analysis by restricting the responses on each hemifield separately for Experiment 1. Figures R1 below show the summary of responses when response analysis was restricted to each hemifield (left and right) of search displays for Experiment 1. A statistical significance threshold of 0.02 was used to interpret the results (like the pre-registered analysis methods).

A linear mixed model with random intercepts for participants indicated that mean scores for each location were not significantly decreased as a function of its distance from the HpValD location for both left ($b = 0.125$, $t(359.691) = 1.730$, $p = 0.085$) and right hemifields ($b = 0.004$, $t(370) = 0.056$, $p = 0.955$) of Experiment 1. In other words, the participants' responses were indeed influenced by the inferences made at the awareness test in Experiment 1.

We have updated the relevant text in the General Discussion section of the revised manuscript. In the supplementary material section, we added the results of this exploratory statistical analysis.

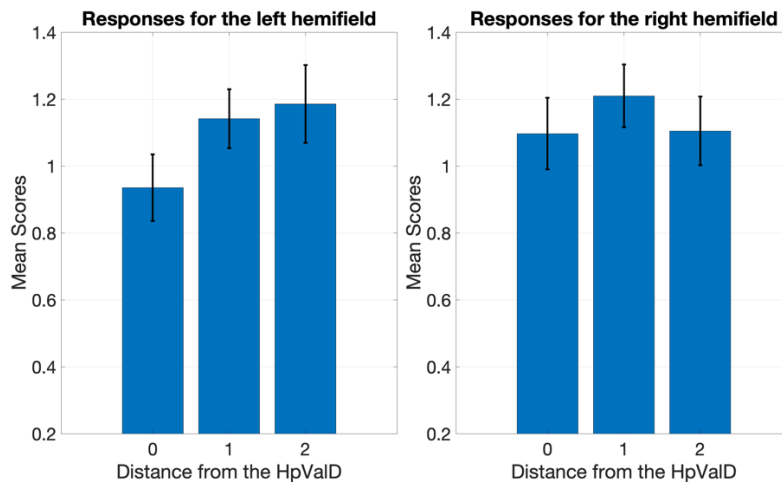


Figure R1: Summary of responses received by participants for the awareness test in Experiment 1. Left panel for the left hemifield; Right panel for the right hemifield. Error bars indicative of SEM. Note: We chose only 3 categories (0, 1, and 2) on the x-axis when restricting the response analyses for left and right hemifields, separately. We have not included the category 3 for the analysis. This is because when restricting the response analysis one each hemifield separately, the category 3 values can be obtained only for two out of four valid distractor locations on each hemifield (indexes: 1, 4, 5, and 8, please see the example displays with index numbers shown in the questionnaires)

Minor comments

p. 2, line 7, remove comma after “known”

Our Response: Thanks. It is now corrected in the revised manuscript.

p. 3, lines 24-25, rewrite “(for review see, (Frost et al., 2019))” as “(for a review, see Frost et al., 2019)”. That is, remove the middle parenthesis and move the comma.

Our Response: Thanks. It is now corrected in the revised manuscript.

p. 4 “However, in recent studies utilizing similar probabilistic tasks, testing the awareness of statistical regularities with more sensitive measures indicated the evidence of explicit knowledge of awareness (Giménez-Fernández et al., 2020; Vadillo et al., 2020). These studies cast doubts on the implicit nature of learning distractor statistical regularities in additional singleton tasks.” -à I understand that no changes should be introduced at this stage, but the authors might want to know that we have just published similar results with the additional singleton task: <https://link.springer.com/article/10.3758/s13414-022-02608-x> This might also be relevant in the general discussion, on page 28, when the ms states that “The relative contributions of whether the participants’ are “aware” or “unaware” of regularities on distractor suppression is not clear from the previous literature (Theeuwes et al., 2022).”

Our Response: Thank you for sharing the recent publication details. We have updated the relevant information in the revised manuscript. Kindly see the manuscript document with tracked changes for a quick review.

p. 9. Sample size planning for awareness is conducted in the d scale, but the reference value from Vadillo et al is in the h scale. I am not sure ds and hs can be considered equivalent. Wouldn't it make more sense to run the power analysis on the h scale?

Our Response: We have used the ‘d scale’ because it is most familiar and frequently used by researchers to calculate the power, using GUI-based open-source software (e.g., GPower), and helps future researchers to use the same metric across studies. In any case, we compared the power analysis results obtained using the d scale and h scale, and the results are in close approximation (please see the Table R1 below). We believe that with the large samples (a total of 124 samples in our study), the difference between the results of power obtained using either the d scale or the h scale would be small enough to be ignored.

Table R1: Results of Power analysis using ‘pwr’ package in R software

<u>Power analysis on the d scale:</u>	<u>Power analysis on the h scale:</u>
<p>Code:</p> <pre>pwr.t.test(d=0.426,n=75,sig.level=0.02, type="paired", alternative="two.sided")</pre>	<p>Code:</p> <pre>pwr.p.test(h=0.426,n=75,sig.level=0.02, alternative="two.sided")</pre>
<p>Output (results):</p> <pre>n = 75 d = 0.426 sig.level = 0.02 power = 0.9023394 alternative = two.sided</pre>	<p>Output (results):</p> <pre>h = 0.426 n = 75 sig.level = 0.02 power = 0.9135462 alternative = two.sided</pre>

Also related to sample size, the authors didn’t commit to a specific N. They simply preregistered that N would be larger than 121. The final N was 124 in both experiments. I assume that no analysis took place before those 124 datasets were collected. Could this be stated explicitly in the ms?

Our Response: Although we pre-registered to have a minimum sample of N = 121, our total sample that was included for the statistical analysis was N=124 after counter-balancing the two High-probability Distractor Locations in the search displays across participants. We did not perform any statistical tests until we had the data of 124 participants who achieved a minimum of 75% overall accuracy in the search task. We have updated this information in the revised manuscript.

In the methods sections, some verbs are now presented in the past tense, but others remain in the present or future. Sometimes, the text sounds a bit weird, like for instance: “The target (shape singleton) was [PAST] present in all the trials, and the target can be [PRESENT] either circle or diamond with equal probability. A blank display with intertrial interval (ITI) will be [FUTURE] randomly...”. I’d suggest changing all the verbs in the method sections to the past tense.

Our Response: We have updated the text with corrections in the revised manuscript.

p. 17, line 23 “HpValD: 1022.227ms ± 137.409...” and elsewhere, please explain what ± refers to. In the next paragraph it seems to stand for SD, but in the figures the error bars stand for SEM. Perhaps it would be clearer to use the same dispersion statistic throughout the text.

Our Response: Thank you for this suggestion. In the revised manuscript, we have used SEM statistics and clearly mentioned them, throughout the text.

p. 17 and elsewhere, would it be possible to report exact p-values instead of just “ $p > 0.02$ ” or “ $p < 0.02$ ”?

Our Response: In the revised manuscript, we reported exact p-values, except for p values < 0.001 (JASP output did not provide us with exact values for less than 0.001).

Figure 3, left panel: it is not easy to appreciate differences across conditions. Would it make sense to rescale the figure in the y-axis? Perhaps including just values from 700 to 1200 or so? Consider the same change for figure 6.

Our Response: Thank you for the suggestion. We have rescaled the figures and updated them in the revised manuscript.

Figure 4. Perhaps the figure caption could remind the reader of the specific text in the awareness question? That would make the figure much easier to understand without going back to the method section. Also, change “definitely” to “definitly” in the x-axis. Incidentally, I found a bit weird that there was a bimodal distribution in participants responses, surprisingly similar for both experiments. Is there any potential explanation for this? Just curiosity...

Our Response: Thanks. The typo has been in the revised manuscript. We regret that we could not figure out a potential explanation for the observed bimodal distributions in the responses.

p. 27 “Indeed, prior research suggested that allocating attention 1 to sensory events is required for statistical learning” Incidentally, a recent study from our lab provides converging evidence for this in a related visual statistical learning paradigm: <https://link.springer.com/article/10.3758/s13423-020-01722-x>

Our Response: Thank you for sharing the paper. In the revised manuscript, we have incorporated the citation of mentioned paper.

Reviewer #2

The present study investigated whether the spatial and non-spatial statistical regularities of task-irrelevant auditory stimulus could suppress the salient visual distractor and their results indicate no reliable effect of task-irrelevant cross-modal stimulus regularities on distractor suppression, irrespective of participants' awareness of the relationship between distractor location and predictive auditory stimulus. I have several concerns.

Will the ability to discriminate the spatial location or two sound frequencies influence the main experimental task? If the participants can't determine the spatial location of sound or the sound frequencies well, they can't surely learn the cross-modal statistical regularities. It seems only those participants who showed a minimum of 75% accuracy were selected for participation in the experiment. But there are chances they can't judge the sound location or frequencies correctly. This will bias all the results reported in this study.

Our Response: Thank you for the comment. The mean percent accuracy of sound discrimination judgments was reasonably high in both experiments. Sound discrimination accuracy for Experiment 1 is: Mean (in %) \pm SEM: 97.8030 ± 0.3697 , while for Experiment 2 is: Mean (in %) \pm SEM: 90.5906 ± 0.6996 . This evidence indicates that the participants heard the sounds clearly and could discriminate them accurately. However, it may be possible that the participants' sound discrimination ability is much lower during the visual search task due to reduced attentional allocation to sound stimuli. Unfortunately, we did not test this possibility in the study. Future experiments would be needed test for the same. We have now updated the relevant information in the revised manuscript.

The auditory stimuli were always simultaneously presented with visual search displays. Will the SOA between the two modalities influence how the stimuli were encoded? Could the SOA be a factor to influence cross-modal statistical learning?

Our Response: The encoding of the auditory stimuli is known to be approximately 50ms faster than a visual flash (Vroomen & Keetels, 2010, <https://doi.org/10.3758/app.72.4.871>). Moreover, in our study experiments, there were multiple visual stimuli in the search display, which can further delay the processing time of visual items compared to a single flash. In such a scenario, we expected that the auditory signal might be sufficiently processed earlier in time to provide anticipation regarding distractor locations in the context of experimental design in our study. However, we do not know whether adding an SOA delay would influence cross-modal statistical learning between auditory stimuli and visual distractor location. We believe that exploring the SOA as a factor to test the learning of cross-modal statistical regularities on distractor processing would be interesting for future studies.

Given the authors mainly reported non-significant differences between valid and invalid distractor locations, the Bayes factor should also be reported to confirm the null effect.

Our Response: Thank you for the suggestion. The Bayesian analysis results supported the null hypothesis more likely than the alternative hypothesis for Experiment 1 ($BF_{01} = 8.2885$) and Experiment 2 ($BF_{01} = 5.338$). We have reported the Bayesian analysis results in the revised manuscript.

I suggest that the authors also show the results of the RTs for different distractor conditions across experimental time (or epoch) in the main results (especially the figure) which are now mentioned in the general discussion only.

Our Response: Thank you. Based on the suggestion, we have made changes to the revised manuscript.

Please correct the citation format and typo in the text.

Our Response: Thank you. We have corrected typos in the revised manuscript. Regarding citation format, we could not find specific guidelines on the website of PCI RR. However, we chose to use a citation format that is consistent with one of the PCI RR-friendly journals.