

Dear Dr. Reshanna Reeder,

Thank you for the opportunity to resubmit a revised copy of our manuscript, “How Perceptual Ability Shapes Memory: An Investigation in Healthy Special Populations” (Article ID #782). We were pleased to read the comments from the reviewers and yourself, and we thank you all for the insightful feedback on our work. We have carefully considered and addressed these comments in our revised submission, as summarized below. We have also updated the figures and tables as we realized the previous versions were too distorted, the sizes of the circles were inconsistent, and the text was too small to read. Please find these changes in Table 1, Table 2, Figure 1, Figure 2, and Figure 3. Additionally, Table 3 has been changed to Figure 4, as it depicts a figure of a table as opposed to a table.

We look forward to hearing from you.

Sincerely,

On behalf of all authors.

**Reviewer 1: Dr. Ariel Kershner**

**1A: Scientific Validity**

- 1. This research question is scientifically justifiable according to the enhanced processing theory. The prior research has only investigated the ventral pathway, while investigating both pathways simultaneously would better address the question.*

**Answer:** Thank you very much for your evaluation on the scientific validity of our research question.

**1B: Logic, rationale, and plausibility of hypotheses**

- 2. The logic, rationale, and plausibility of the three hypotheses are well-conceived and plausible. The hypotheses fully address the research question and the underlying theory.*

**Answer:** Thank you very much for your evaluation on the logic, rationale, and plausibility of our hypotheses.

## **1C: Soundness and feasibility of methodology and analysis pipeline**

### **Methodology**

- 3. Given the preponderance of drop out rates in a longer-term study such as this, as well as the difficulty in finding suitable participants, I would err on a larger initial sample size (currently only 12 participants can be excluded from each group).*

**Answer:** Thank you for raising this point. We have now adjusted our recruitment strategy to continue recruiting participants until we reach our target number in each group, taking our exclusion criteria into account. This adjustment can be found on page 9 where we state, "If, however, after exclusions, we lose more than 12 participants in a group, we will continue to target recruitment specifically for the group that has not yet achieved the target sample size."

- 4. Are participants who skip a day in the sequence excluded, or do they pick up again when they return to the experiment?*

**Answer:** Thank you for your question. We will not exclude participants who skip a day in the sequence or pick up later. We will only exclude participants if at least 50% of their data is missing in the visual perceptual task, visual short-term memory task or the visual long-term memory task. However, to address this point, we have now decided to also include a second conservative analysis in addition to the first, where we will only include participants with complete datasets to see if the results are still consistent. We have made the above amendment on page 29 stating the following, "We will also repeat the planned analyses outlined below using a more conservative approach, including only participants with complete datasets. This will help us determine if the results remain consistent compared to when participants with less complete datasets (but more than 50% of data) are included."

- 5. Inter-rater reliability, synesthesia tests, and no overlapping scores are all good controls.*

**Answer:** Thank you for your evaluation.

- 6. Can the sequence of the method (perceptual task, WM, and LTM tasks, as well as loads within WM) be counterbalanced? The spatial and featural reports are already counterbalanced.*

**Answer:** Thank you for your comment. We considered counterbalancing the sequence of the tasks, but we have specific reasons for not doing so. First, there are practical implications for maintaining a fixed sequence. We need a few days between the study and testing phases in the visual long-term memory task. During this interval, presenting other tasks might interfere with the encoding and retrieval processes of the visual long-term memory task. Additionally, there could be potential interference between two memory tasks. This fixed order of task presentation was also used by Ovalle-Fresa et al. (2021). Furthermore, regarding counterbalancing the loads for the visual short-term memory task, we decided to start with the easiest load first to establish a baseline performance level. This sequential approach of gradually increasing the load aligns with previous research, where loads are presented in increasing order. We have clarified this with the following sentences on page 13, "All participants will perform these three tasks in the sequence outlined above. This sequence is necessary because a few days are required between the two phases of the visual long-term memory task (see Table 1). During this interval, other tasks cannot be presented as they might interfere with the encoding and retrieval processes of the visual long-term memory task."

7. *Will the report of features for the stimuli in the STM 3 and 5 load tasks be randomized? Or will it always go clockwise in stimulus order, for example?*

**Answer:** Thank you for your question. Yes, the report of features for the stimuli in visual short-term memory task loads 3 and 5 will be randomized. We have now clarified this on page 21 for load three by amending the following statement, "Participants will then be probed for the color and spatial location association of each image in randomized order, and will be asked to retrieve these to the best of their ability." And on page 23 for load five, "As before, images will be presented in randomized order in the retrieval segment."

8. *Also in the higher VWM load task, will the beginning point on the wheel be randomly generated with equidistant stimulus positions from there, or will the beginning position remain constant?*

**Answer:** Thank you for bringing this up. Yes, the beginning point on the wheel will be randomly generated with jittered stimulus positions from there. We have now clarified this

on page 24, “For all loads of the visual short-term memory task, the starting point on the wheel will be randomly generated.”

9. *Does the color wheel remain constant or does it rotate? I’m a bit concerned about having the same type of wheel for reporting space and color, as these could confound each other. Imagine a green bucket that is in the spatial position of where the color green is on the color wheel. Regardless of group assignment, one feature (e.g., color for a color expert) may help the participant remember the other feature (e.g., space for a color expert), potentially be recoding the space as where it would be on the color wheel. If the color wheel rotated, I believe this confound would be alleviated.*

**Answer:** Thank you for this comment. Yes, the color wheel will rotate on every trial. For example, red, which is at 0 degrees, will not always appear at the top of the circle. Instead, it will rotate around the circle on every trial. This was initially explained on page 17 in the visual perceptual task section. However, we have now moved and expanded on it on page 13 with the following statement, “In addition, in all three tasks, the color wheel will rotate on every trial to ensure that the same color does not appear in the same position on the circle each time. Instead, it will appear in different positions on the circle on every trial.”

10. *The analysis is appropriate for the hypotheses.*

**Answer:** Thank you for your comment.

#### **1D: Clarity and degree of methodological detail and analysis to prevent flexibility**

11. *The specificity of the methodology and analysis provide littler room for researcher bias, are fully linked to the hypotheses, and can be replicated.*

**Answer:** Thank you for your evaluation on the clarity and degree of the methodological detail and analysis.

#### **1E: Outcome-neutral conditions**

12. *Other than the checks and exclusion criteria to ensure that no group overlaps, there are no applicable manipulation checks. The design is very well-controlled.*

**Answer:** Thank you very much for your evaluation on the outcome-neutral conditions of our experiment.

## Reviewer 2: Dr. Katherine Moore

1. *This is an interesting investigation, as there are open questions about what drives superior perception and memory performance in synesthetes compared to control subjects. The paper was well-written and easy to follow, and provided detailed information in terms of design as well planned statistical analyses. I was especially impressed with the links to various aspects of the study, including the ChatGPT output and especially the task itself. I wish all published articles would include a link to the task for readers to fully grasp what the participants are experiencing. While the methods were described clearly, there was no substitute to experiencing the task itself. I look forward to learning the results of this investigation. I believe this is a strong experiment, and I only have a few minor questions to be addressed in the next version of the paper.*

**Answer:** Thank you very much for this feedback on our paper and we are glad you enjoyed the links to tasks.

2. *My primary question is with regard to the pilot data and power analyses to determine the appropriate sample size. The authors are clearly quite skilled in their statistical analyses; I trust that they are capable of performing adequate power analyses to determine an appropriate sample size for the current investigation. However, I was confused as to how the pilot study was conducted. The authors describe “simulated” data as well as “pilot” data. It is unclear how “simulated” participants were generated and why they could hold value in determining outcomes the same way true participants could. I had expected power analyses to be performed based on prior literature and/or new pilot participants, so the “simulated” participants threw me. More detail is needed here.*

**Answer:** Thank you for asking clarification on this point. The pilot participants, non-synesthetic controls recruited from the student participant pool at UniDistance Suisse, were used in simulating data for the power calculation. The power calculation for this study was conducted in two steps. First, we simulated an artificial dataset from the pilot data collected from the three tasks through bootstrapping. Second, for each task, we built a model from

the artificial dataset and used the effect sizes from this model to determine the required sample size. Regarding previous studies, there hasn't been one using a similar paradigm with the same research question, except for the Ovalle-Fresa et al. (2021) study published in Cortex. However, that study only examined one aspect of the enhanced processing account and did not use the same paradigm as ours. Our pilot data and all the power analysis scripts will be shared on OSF so that researchers can look at it. However, we understand that the power analysis might be confusing to read and interpret. To simplify this, we have created a new section called "A priori power calculation" and have divided this into two further sections (pages 29 to 31). The first section, "Artificial dataset simulation," describes how we created an artificial dataset based on the pilot data. The second section, "Power analysis," explains how we conducted the power analysis for our study.

- 3. The rest of my concerns are minor. Also in terms of participant recruitment, I would expect there to be considerable overlap among groups. It is common for people with synesthesia to have multiple forms of it, with both grapheme-color and sequence-spatial to both be quite common. As for the non-synesthetes, the color and spatial experts are also likely to have considerable overlap; indeed, the ChatGPT output listed many of the same professions and hobbies for both color and spatial expertise. Finally, it would be unsurprising to find that color synesthetes are likely to share in the color expertise hobbies/professions. That said, the authors include a detailed plan to recruit participants with a screening questionnaire and will continue to seek participants in batches of 1000 until they have the appropriate pool. I did, however, have one question about the target sample sizes. The authors indicated that they targeted 40 participants per group, and would recruit 52 per group to aim for that number, in order to account for eliminations. Would these eliminations be due to performance on the experimental tasks, or would they also include issues such as "overlap" in groups? If the former, I think it is an appropriate plan. If the latter, I would recommend handling overlap cases (since there may be many) in the screening procedure prior to the main experiment.*

**Answer:** Thank you for this very good point. We have now taken both possibilities into account by clarifying the exclusion criteria based on performance in experimental tasks on page 7. We have extended the original statement to include: "To ensure 90% power at an

alpha level of 0.02 (cf. power analysis) and account for participant exclusions based on performance in the experimental tasks (cf. exclusion criteria), we will increase the total by 30%, resulting in 208 participants, with 52 in each group.” Furthermore, on page 9, we have added, “This exclusion of participants with overlapping or missing consistency values will be carried out after the screening procedure (i.e., these participants will not perform the experimental tasks).”, to clarify that overlapping cases will be handled in the screening procedure prior to the main experiment.

- 4. In terms of the results, the authors make a prediction of an interaction between domain and task in that color experts will perform better on the color task than the spatial task, and spatial experts will perform better on the spatial task than the color task. I wonder if this prediction should be softened to merely predict an interaction including “relative” superiority rather than a more specific outcome. In other words, if one task happens to be easier than another, perhaps both groups will perform better on the color task than the spatial task, even if the color experts perform more better. Either way, the result will be impactful.*

**Answer:** Thank you very much for this suggestion, we also believe that this is a good idea and so have decided to change the wording to soften the prediction of the interaction. We have now implemented this change on pages 7, 25, and 27.

- 5. Also in the predictions, the authors state that they predict on the short term memory task for domain, task, and “load” to interact because at load of five there will be floor effects in memory performance, so the domain/task interaction may change compared to loads one and three. The authors then state that for the long term memory task, the critical interaction will be among domain, task, and “day,” but there is no prediction about how “day” will affect the primary interaction. I think it would be useful to include such a prediction, even if it is that “day” will not interact with the variables of interest.*

**Answer:** Thank you very much for pointing this out. This was an oversight on our part. The key analysis of interest in this task is the interaction between domain and feature in both the testing phase and the study phase of the visual long-term memory task. We have now updated Table 2 on page 26 and clarified the wording on page 27 by revising the original

paragraph to, “In all tasks, we predict a significant interaction between domain and feature. This two-way interaction between domain and feature is the primary effect of interest for the visual perceptual task and the study phase, and the testing phase of the long-term memory task (which is the key phase of interest in this task). In these analyses, we predict that participants will perform relatively better in the feature that is congruent with their synesthesia or expertise. Specifically, grapheme-color synesthetes and color experts are expected to perform relatively better in the color feature, while sequence-space synesthetes and spatial experts are expected to perform relatively better in the spatial feature. During the study phase of the visual long-term memory task, we anticipate seeing performance improvements in the feature congruent with participants' synesthesia or expertise on each day.” We have also updated Table 2 on page 26.

6. *I finally have a few questions about the synesthesia screen as well as how synesthetes might behave in the primary task. The authors mention that in order to investigate sequence-spatial synesthesia, participants will indicate locations of sequences on the computer. However, many sequence-spatial synesthetes experience their spatial percepts in three dimensions. Will there be a way to account / screen for this experience? If not, how will this dimension limitation affect the screening process for this kind of synesthete?*

**Answer:** Previous research has found that sequence-space synesthetes sometimes experience their spatial percepts in three dimensions and sometimes in two dimensions. In our task, we will instruct participants to use the screen as a reference frame for their synesthetic experiences, allowing them to refer to their three-dimensional experience and intuitively place stimuli in different spatial locations (mentioned on page 16). The instructions and procedure will follow those of Rothen et al. (2016), who found good sensitivity and specificity for their sequence-space synesthesia consistency test.

7. *During the task itself, do you anticipate an issue with synesthetes (particularly color-based ones) experiencing a synesthetic reaction to a stimulus? For example, the example piano stimulus jumped out at me as one that might elicit a color reaction to it. For those with chromesthesia, the synesthete might associate the timbre of a piano with a particular color that may or may not map onto the color the object is*



*given in the task. It could undermine the results if pre-existing synesthetic mappings for particular objects might be incongruent with the assigned colors in the experiment. Perhaps a questionnaire asking for feedback or reactions could take care of this concern, and any unusual observations could be reported in the discussion.*

**Answer:** Thank you for this insightful point. We will now include a question on this in the strategy questionnaire to identify any such occurrences. If we find that many synesthetes indicate interference due to incongruent synesthetic mappings for particular objects, we will consider a follow-up session. In this session, the synesthetes will be re-invited and asked to indicate their color experiences of the objects.

8. *Is there a reason that there are very slightly different numbers of stimuli / trials in the different loads of the memory task? (This doesn't need to be changed because they are close enough, but I am just curious.)*

**Answer:** Thank you very much for your question. The number of stimuli in the different loads of the visual short-term memory task remains the same (i.e., 45 images in each load condition). However, the number of trials participants are presented with in each load varies because we wanted to keep the total number of stimuli presented constant. Therefore, there are 45 trials in load one, as participants need to encode and recall the color and spatial features of 45 individual stimuli. In load three, there are 15 trials, as participants need to remember the color and spatial features of three images presented simultaneously on each trial (45 images / 3 images per trial = 15 trials). Finally, in load five, there are nine trials, as participants need to remember the color and spatial features of five images on each trial (45 images / 5 images per trial = 9 trials). We realize that this is probably not described well in the manuscript so on page 21 we have changed the original sentence reading, "Each block will comprise of 45 stimuli" to, "The number of stimuli presented in each load will be kept constant at 45, so participants will be presented with a total of 45 stimuli in each of the three loads. Therefore, there are 45 trials in load one, as participants need to encode and recall the color and spatial features of 45 individual stimuli. In load three, there are 15 trials, as participants need to remember the color and spatial features of three images presented simultaneously on each trial (45 images / 3 images per trial = 15 trials). Finally, in load five, there are nine trials, as participants need to remember the color and spatial features of five images on each trial (45 images / 5 images per trial = 9 trials)."

9. *This might have been a coincidence, but I noticed in the demo of the task that in loads 3 and 5, the objects appeared in spectrum order in terms of color (e.g. clockwise: red, golden, green, blue, purple). I wonder if this affects strategy, because in that case it might be necessary only to remember spatial location and then fill in the colors based on where they might appear in order. If the colors are in fact truly randomized, ignore this question! I did appreciate the methodological explanation that described how the colors were selected on each trial to ensure proper color distance; I'm just curious about how they were arranged spatially within a single trial.*

**Answer:**

Thank you for trying out our tasks; it's very helpful to know how they were experienced. The colors and locations are randomized on each trial with some jitter added to them.

Essentially, the circle is rotated on every trial, and a color and spatial location value are chosen. Once one value is chosen, the other colors and spatial locations are selected to be the same distance away from each other in each trial to avoid color and spatial overlap.

Initially, in load three, the colors and locations were spaced at least 100 degrees apart with jitter added, and in load five, they were spaced at least 60 degrees apart with jitter added.

However, to address this point, we have now increased the jitter to make the task more challenging for participants. With this change, in load three, the colors and locations are spaced at least 50 degrees apart when jitter is added, and in load five, they are now spaced at least 40 degrees apart when jitter is added. We have updated this section on page 23 to read, "For load three, the minimum separation between the three presented images will be 50 degrees, and for load five, it will be 40 degrees between the five presented images".