Point-by-point Response to Reviewers' Comments

Thank you to all the Reviewers, and to the Recommender, for taking the time to carefully review the manuscript and give detailed and constructive comments. Below is our point-by-point response to each comment.

1. Review by Janina Neufeld

Comment 1

The research questions seem valid to me. One aspect could be made a bit clearer in the introduction, namely to what extent whether the link between perception and memory has been explored in other populations previously.

Response

Thank you for raising this. We note that another reviewer has made a similar comment, and therefore we have included additional information regarding representational accounts of perception and memory, with an emphasis on other special populations where possible (e.g., amnesics and aphantasics), in paragraphs two and three of Section 1. Introduction.

Comment 2

The hypotheses are plausible. I have one comment regarding hypothesis 2A): Correlations between the precision of perception and the precision in memory. If a person cannot differentiate between, for instance, 2 colours, it should be impossible to for them to remember which of those 2 colours was presented. A high accuracy in perception should therefore be a pre-requisite for high accuracy in memory, but someone being highly accurate in perception does not necessarily need to have a highly accurate memory. Can the data somehow be used to differentiate whether a relationship between accuracy and memory could de due to high accuracy being a pre-requisite for high memory precision?

Response

Perceptual ability does provide a ceiling for performance on tests of perceptual memory. However, previous studies have shown that this ceiling is only reached after multiple learning episodes (Ovalle Fresa & Rothen, 2019) so enhanced perceptual memory in earlier learning attempts cannot be trivially explained in this way.

Comment 3

Are only grapheme colour synesthetes recruited who score below the cut-off for all graphemes or also those only scoring below the cut-off for either letters or digits but not both? If a person only reports, for instance, colours for digits, will that person only be tested for 10 digits? The validated consistency test used all graphemes, I believe, so many more items. Of course there are known cases of digit-colour synesthetes without letter-colour synaesthesia, but it might be relatively easy to memorize colours for only 10 items. So perhaps, there should be a minimum of items in the consistency test, even if only as "distractors"?

Response

The standard consistency test in our lab presents 36 stimuli (10 digits and 26 letters) although a 'no colour' option is available. We will prioritise recruitment of participants who report colours for most letters and digits as this ensures a more robust assessment of synaesthesia. If any synaesthete has only colours for letters or numbers then this will be reported. We have added information regarding the number of stimuli presented to Section 2.2.1. Consistency Tests, and information about prioritising the recruitment of grapheme-colour synaesthetes with a greater number of colour experiences to Section 2.1. Participants.

Comment 4

What types of sequence-space synaesthesia will be tested for? Month, weekdays, alphabet, numbers, any reported?

Response

The test involves months, weekdays, and numbers and we will select participants who report these. We have updated Section 2.2.1. Consistency Tests in the manuscript to stipulate that these are the types of sequence-space synaesthesia assessed.

Comment 5

The power calculation seems to have been based on power for comparison of 2 independent means. Relatives of synesthetes are not really an independent group. It would be of interest to discuss expected effect sizes for comparisons between synaesthetes and their relatives, especially since you do not expect to be able to recruit more than the minimum of 61. I understand that it is difficult since the study is the first on memory in relatives of synaesthetes, but maybe one can get an idea based on a different measure or a condition other than synaesthesia…

Response

If the difference between synaesthetes and their relatives is of a similar magnitude to the difference between synaesthetes and non-synaesthete controls then we do have sufficient power to detect this (at power = 0.9, alpha = 0.02), but at smaller effect sizes we do lose power (detecting a difference of Cohen's d = 0.4 would be a power of 0.65 at $p < .02$). It is to be noted that we also conduct multivariate analyses such that multiple smaller univariate effects between groups may yield more robust differences. This information has been added to Section 2.1.1. Sample Size Determination in the manuscript.

Comment 6

As I understand your methods plan, you will only ask relatives once whether they have synaesthesia, and if they claim not to have it, it is assumed that is true. How high is the chance that some of them are actually synaesthetes and what consequences does that have for your results?

Response

In light of this comment, we have changed the screening procedures slightly. As well as asking relatives and controls to indicate via response-box whether they believe that they have synaesthesia, we will provide consistency tests (both grapheme-colour and sequence-space) to all relative and control participants to verify that this is so. As we are mindful that taking these tests may be irrelevant for the relative and control group, we implement a stopping rule for these tests such that if after presenting all stimuli >90% of stimuli have not been associated with a colour or spatial location, participants are advised that they "do not appear to associate stimuli with colours/spatial locations" and are given the option to end the consistency test there or to continue for two further repetitions of the consistency test. As stated in the original manuscript, synaesthetes will not be asked to repeat these consistency tests as we already have this data. As these additional screening measures will take approximately 5-20 minutes to complete, we have separated the screening tests for colour blindness and synaesthesia consistency into a short, separate session which takes place prior to the main experiment. Eligible participants will have the option to complete Session One immediate after completing screening or can complete it later. Revisions have been made to Figure 1, Section 2.1.2. Exclusion and Inclusion Criteria and to Section 2.2.1. Consistency Tests to reflect these changes.

Comment 7

What do you mean by corrected to normal colour vision? Do you mean corrected to normal vision?

Response

Yes, it should be corrected to normal vision. We have deleted the word "colour" in the first sentence of Section 2.1.2. Exclusion and Inclusion Criteria.

Comment 8

Could you please comment why the feedback given during the learning phase of the long-term memory task will be based on a non-linear scale?

Response

The same non-linear scale is used throughout. Response accuracy on a given trial can fall between 179 degrees and 0 degrees so the feedback needs to be anchored to these endpoints. However, random guessing would produce a response of 90 degrees (on average) and, in practice, most responses fall within 50 degrees. On a linear scale most feedback would be within 25% of the perfectly correct anchor. On our non-linear scale this distance increases to about 50% and is therefore more sensitive (i.e. one half of the scale is 180 to ~50 degrees and the other half is from ~50 to 0). Random guessing gets pushed more towards the incorrect anchor rather than being at the midpoint of the scale.

Comment 9

I am assuming that there will be some overlap between grapheme-colour synaesthesia and sequence-space so that some individuals contribute to both groups. How much overlap are you expecting?

Response

We would aim for roughly equal numbers (so a third with grapheme-colour, a third with sequence space, and a third with both).

Comment 10

Are you planning to also analyse the influence of having several synaesthesia types?

Response

Other reviewers have requested this and we plan this as an exploratory analysis, noting that our main hypotheses concerns the differences between colour and spatial experiences. A new section (Section 3.6. Exploratory Analysis: Impact of Number of Types of Synaesthesia) has been added to the manuscript for this purpose.

Comment 11

One thing that was a bit difficult to follow was how exactly the guess rate is estimated. I understand that the guessing likelihood should be higher if answer error degree differences are distributed randomly. Does this mean that the similarity between perfect randomness and the error distribution will be somehow expressed as a model fit score or alike?

Response

Our main analyses (model-free) do not rely on an estimation of guessing rate (the dependent variable is just the average distance from target in degrees). However, the model-based analysis breaks down the distribution of responses into two components: a guess rate (uniform distribution) and a memory estimate (normal distribution). Effectively, it tries to separate the mechanisms into absence of memory (guessing) versus quality of memory and gives more information as to underlying mechanism. We refer to Figure 5. Summary statistics (absolute error, precision and [for the short-term memory task] swap errors) are produced by the R package 'Mixtur' (Grange & Moore, 2022), which we shall report. It is also possible to obtain a model-fit score; however, we note that previous papers utilising this model (e.g., (Gresch et al., 2024; Hu et al., 2023) have not reported the model-fit score.

Comment 12

Why did you choose Bayes factor of at least 6 as a threshold? Is this based on previous articles or recommendations you could cite here?

Response

The Peer Community Registered Reports does not specify a BF threshold value but some journals using this service do (e.g. Cortex has BF > 6). Doing an RR doesn't prevent arbitrary decisions being made, it just puts these decisions upfront (to guard against p-hacking etc.).

2. Review by David Brang

Comment 1

The study provides a strong framing and testable hypotheses. Though it should also be noted that the absence of an effect within family members is not evidence for the alternative view. E.g., if synesthetes but not family members demonstrate a benefit this could be due to either the dual coding account or the enhanced processing account, if the phenotypical differences in synesthetes are weaker in their relatives and thus creating non-detectable effects on memory and perception. Some of this is of course the common difficulty of arguing for the null hypothesis and will depend on the strength of the observed data in each group (and potentially explainable using proposed baysean statistics). Though, given the results of Ward & Filiz, 2020, the absence of an effect for family members is perhaps the less likely outcome.

Response

We acknowledge that a null result for the relatives (i.e. that they are indistinguishable from controls) would be hard to interpret, particularly if Bayes Factors are also insensitive. Our final multivariate analysis may be more sensitive than ANOVAs because it considers all variables simultaneously (i.e. so multiple weaker effects can constitute a larger aggregate effect).

Comment 2

It is reported that controls will be recruited through internet-based participants databases, such as Prolific. Are the researchers concerned about confounding group differences separate from synesthesia? It is a strength to the design that the authors are using a motivational scale questionnaire to try and quantify some of the differences that may exist between the groups but a more closely yoked control group may prevent ambiguous results if large motivational differences are observed. Additionally, is there any dimension of the data that can be used to show the selectivity of the results (e.g., a measure that synesthetes and non-synesthetes would not be expected to differ on)? This level of control condition/analyses would help rule out effects purely due to how subjects were selected. Another option would be a synesthetic control group (e.g., Mirror Touch Synesthetes) who would not be expected to show enhancements but would be motivated nonetheless.

Response

These are important points and we acknowledge these limitations. Our hypothesis that grapheme-colour and sequence space may differ from each other (according to colour versus spatial memory) is similar to what you are proposing, although there is a risk that these two types of synaesthesia may yield similar results. Following suggestions from other reviewers we will also run an exploratory analysis in which we determine whether the number of types of synaesthesia has an effect. This exploratory analysis is described in a new section of the manuscript (Section 3.6. Exploratory Analysis: Impact of Number of Types of Synaesthesia).

Comment 3

Please confirm that for the Visual Perception task, the dependent measure is the recalled location/color as opposed to the left/right color/position matching accuracy.

Response

The reviewer may have misunderstood the task. It does involve matching, i.e. the target colour and location remain visible on the screen whilst the participant makes their selection (on a different part of the screen).

Comment 4

Additional information is additionally needed for the Visual Perception task. How many blocks are presented and what is the total number of trials (it is noted 15 objects per block). Do subjects recall location and color for each object sequentially or only one feature for each trial? If both features for each trial is the order of recalled feature randomized to prevent order effects?

Response

Whether colour or location is selected first is randomised between participants (i.e. consistent within each participant; see Section 2.3. Procedure). So the target visual stimulus (coloured object in a particular location) stays the same for both colour and location selections, but the task (on the other side of the screen) is either colour or location selection (not both). Recall for both features is required, and each feature is sequentially probed per trial (the recall order is counterbalanced across participants). In Section 2.3. (Procedure) we state that the visual perception task will be completed in three blocks of 15 trials (45 objects).

Comment 5

For the Short-term Memory Task the text description states "Following a delay of one second, participants will be cued to recall the colour and location of each object from the array in turn" whereas the figure states "Following a delay of one second, participants are asked to report the colour and location of one of the previously presented items." Are subjects presented which each item or only one single item after each trial? Similar to the Visual Perception task, do subjects report both the spatial position and color or only one (and if both, please confirm the order of feature recalled is randomized).

Response

In the test phase, participants are shown a greyscale object as a cue. They then either drag the object to its location on the wheel or click on the relevant part of the colour wheel to change the colour of the object. Colour and location decisions are done sequentially (location and colour order randomised across participants). All objects within the array are sequentially probed, in random order (e.g., all three objects are individually probed for location and colour in the load three condition). We have updated the caption for Figure 4 (Short-term Memory) to be clearer about the number of objects probed for each array.

3. Review by Tessa van Leeuwen

Comment 1

In the first two paragraphs of the introduction, however, I miss a rigorous motivation for studying the link between perception and memory specifically. It is mentioned that this link is important, and the authors refer to it later in the introduction as well, but I feel the importance of the set-up study can be emphasized more at this point.

Response

Thank you for raising this point. As other reviewers have requested that further information to this effect be included in the introduction, we include additional reference to research on representational accounts of perception and memory, with an emphasis on other special populations (e.g., amnesiacs and aphants), in paragraphs two and three of Section 1. Introduction.

Comment 2

The power analysis is well motivated and the sampling plan is realistic. I would, however, like the authors to comment on the feasibility of recruiting 61 first-degree non-synaesthetic relatives from 100 synaesthete participants. Perhaps the authors can refer to other papers that have successfully done this, or comment on their own experience with this? Also, I was wondering whether there would be any preference to recruit siblings of synaesthetes (rather than parents or children), to avoid large age differences between the groups.

Response

We anticipate that this will indeed be challenging and are prepared for this part of the study to take time. Our strategies here are: (1) to prioritise synaesthete recruitment from people who have already identified a candidate family member and (2) to carry on recruiting family members after the N=100 target of synaesthetes is met (e.g. we may have to test 130 families to find 100 synaesthetes and 61 willing non-synaesthetes). We can see how recruiting siblings would have certain benefits but it would also make it more challenging to recruit overall. As such, our aim is to match the mean of the groups by age rather than yoke relatives. Information about the prospect of recruiting relatives after the target of N=100 synaesthetes has been met has been added to Section 2.1.1. Sample Size Determination.

Comment 3

Consistency tests: With regard to synaesthetes already in the Sussex database, it is mentioned that recruitment ID will be retrieved to verify which types of synaesthesia the participants have. This is a rather crucial aspect as the number of synaesthesia types matters for the cognitive profile to be expected, e.g., the strength of attention to detail-like profiles (including perhaps, colour sensitivity, as it belongs to the parvocellular visual system), for instance. In this respect, it has been shown that the more types of synaesthesia someone has, the more autism-like the profile is and the more deviant the Glasgow Sensory Questionnaire scores are (Ward et al., Cortex, 2017). I therefore feel the number of synaesthesia types is a confound to control for when assessing especially perceptual performance across the groups of synaesthetes, as synaesthetes with a stronger ‘synaesthetic profile’ might score very

different from synaesthetes with only 1 type (Hypothesis 3). Additionally, synaesthetes with a stronger profile might differ more clearly from control participants in their accuracy (Hypothesis 1).

Response

Following suggestions from several reviewers we have added an exploratory analysis that considers the number of types of synaesthesia. This can be reviewed in a new section of the manuscript: Section 3.6. Exploratory Analysis: Impact of Number of Types of Synaesthesia.

Comment 3

Consistency test for SSS: perhaps the authors can, in the paper itself, briefly comment on the ability of their questionnaire (in combination with the consistency test) to discriminate between synaesthetes and non-synaesthetes. It is our experience that with the consistency test only, this can be hard to do because non-synaesthetes can easily adopt a structured order for e.g., the days of the week, artificially lowering their consistency score (Van Petersen et al., Cortex, 2020). So it would be relevant to stress the additional qualities of the questionnaire in the paper itself. As an alternative verification task, I can suggest a drawing tasks that previously worked well to discriminate between synaesthetes and non-synaesthetes (Van Petersen et al., Cortex, 2020), although I acknowledge this would be impractical to implement online, perhaps.

Response

Thank you very much for this suggestion. We have added further information in Section 2.2.1. Consistency Tests, which discusses how the use of a questionnaire allows us to better distinguish between sequence-space synaesthetes and non-synaesthetes. The drawing task looks to be a very useful alternative verification task; however, we are unable to adopt it at this time due to the online nature of this research (as acknowledged by the reviewer). We shall bear this in mind for future in-person research.