Grace Edwards

Dear Dr. Karakashevska,

Thank you for your Stage 1 submission to PCI-RR. We have received comments from two expert reviewers who both enjoyed reading your manuscript and were impressed by the design of your study.

We are pleased that the reviewers are positive about our proposed study.

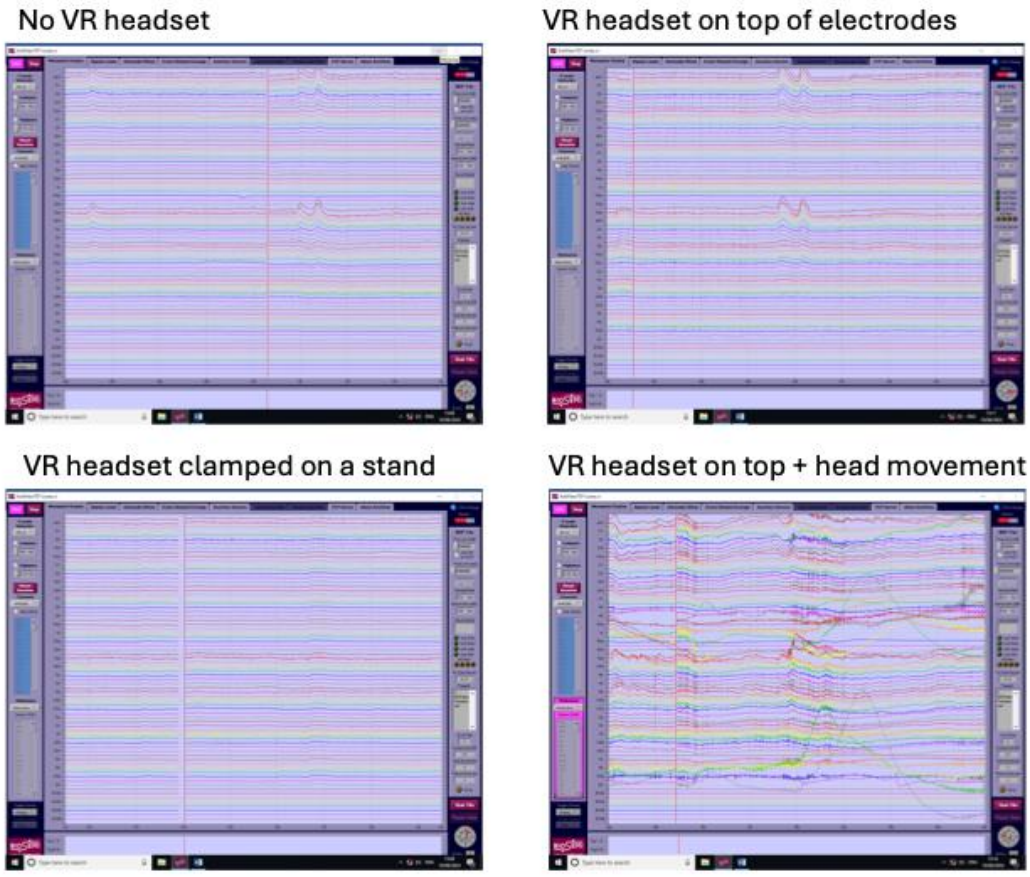
Dr. Baker requests some further discussion about the interaction between the virtual reality (VR) headset and the EEG recording regarding the potential noise in the signal. They also suggest including Bayes factors in the planned analyses. Although it has become common practice for people to report both Bayesian and frequentist statistics in their manuscripts, PCI-RR prefer the authors avoid mixing their hypothesis testing frameworks. As the authors have powered their study for the equivalence test, I believe they can conclude the probable absence of an effect, if that case arises.

As recommended, we have not added Bayesian analysis.

Our Anonymous Reviewer 1 (AR1) provides some useful feedback regarding methodological considerations and clarifications when employing a VR environment, which should be addressed in detail. AR1 also echoes the concern of Dr. Baker regarding the quality of the EEG signal with the addition of the VR headset. They further highlight that with the potential drop in signal to noise, the effect the authors wish to detect may become smaller. I support AR1's request in considering a smaller effect size in the power analyses.

To test this problem, we have tried multiple set ups with the VR headset and the BioSemi system. The VR headset will be placed on top of the electrodes with the straps, carefully clasping them so the weight is equally distributed. Participants will then be instructed to place their head on a chin rest and retain from performing any harsh movements. As seen in the Figure below, it is only head movements which introduce substantial noise. The mere presence of the VR headset does not. This is supported in a study by Tauscher et al. (2019). They systematically show that if

participants are static, the EEG signal quality is not altered significantly and can be improved by modifying the headset strap.



Screenshots of VR headset set up with BioSemi versus no VR headset.

As an additional check, we will record additional EEG data from each participant for with and without VR. If the VR helmet significantly degrades the EEG signal, we will adjust pipeline parameters proportionally. This check is now explained in the manuscript:

"Signal quality check

The power analyses presented were informed by previous SPN research. While there is a possibility that integrating Virtual Reality (VR) could reduce signal quality below typical standards, necessitating larger sample sizes to detect the expected effects, prior research (Tauscher et al., 2019) and our testing with the

VR headset suggest this is unlikely. To assess the signal quality in this project, we will record EEG data in one 32-trial block with the VR headset and one 32-trial block without it (same stimuli shown on a screen) for each participant after the main experiment. Signal quality will be assessed by the number of trials rejected based on the criteria outlined above.

Ideally, we aim to find evidence of no significant difference in signal quality, rather than absence of evidence. However, as we lack a precise definition of what constitutes a meaningful difference in EEG signal quality, a non-significant difference in trial rejection ($p > 0.05$) will be considered sufficient to act as if no substantial differences exist. An initial signal quality analysis will be conducted at N=24. If the VR headset significantly degrades the EEG signal, it is crucial to identify this before collecting a larger sample. If signal quality is compromised, we will adjust the analysis pipeline parameters to accommodate this reduction. First, we will adjust the +/- 100 microvolt trial exclusion thresholds until trial exclusion aligns with the non-VR condition. Second, if there is a 20% reduction in EEG signal quality with the VR headset, we will increase the maximum sample size by 20% (to 144) and recalculate the look points for the sequential analysis. These adjustments will only be made if there is a significant reduction in signal quality when using the VR headset."

On a different note, AR1 requests a clarification on what analyses would be done (if any) if no significant SPN is detected in for hypothesis 1.

Note that we have changed hypothesis 1 to 'there will be an SPN in the frontoparallel conditions. If hypothesis 1 is wrong and there is no SPN in the frontoparallel conditions, no other analyses would be meaningful. However, this is very unlikely. We have now explained this in the paper.

In general, AR1 finds the references to Karakashevska et al. (forthcoming 1 and 2) difficult to evaluate as they weren't able to access the articles. I suggest adding links to your preprints in the current submission. I believe Karakashevska forthcoming 2 is the Stage 2 article you have under review with PCI-RR currently.

We have clarified this. Karakashevska et al. (forthcoming 1) is under review and will now be cited as Karakashevska and Makin (2024). Karakashevska et al. (forthcoming 2) which is the Stage 2 article under review will be Karakashevska et al. (RR1) with a hyperlink of the preprint on psyArxiv.

From my perspective as a PCI-RR recommender, I have a couple of further comments:

1. Could the authors clarify the exclusion criteria regarding the behavior on page 7? Will >80% performance need to be upheld for all conditions?

We have now clarified this:

> *"We will replace participants whose performance is below 80% correct on either task. There is no requirement that they should exceed 80% correct on every stimulus condition within each task. Performance is usually > 90% correct on similar tasks with similar stimuli (Karakashevska et al., RR1)*

2. Your sample of 120 participants is determined for zero perspective cost (i.e. less that -0.35 microvolts) at 95% power. Does this sample give you enough power to detect the effect sizes for your other analyses, especially given that you may stop data collection at 48 participants? Please be explicit regarding expected effect sizes. For hypothesis 2 in your Study Plan Table you state "The final sample size of 120 was chosen to detect smaller effects and is thus adequate to detect the main effect of Task, which is likely to be large." How large do you expect? And what if the final sample is actually 48?

We have now worked through the power analysis in far more detail including these considerations. First, note that we have change the order of hypothesis. Second, note that we have changed hypothesis 1 so that it only predicts SPNs in the frontoparallel conditions. It is likely that there will be SPNs in the perspective conditions, but this is not an essential criterion for the project. The hypothesis section now looks like this:

> *"Hypothesis 1*
>
> *There will be an SPN in the frontoparallel conditions of both tasks (black bars < 0 in Figure 3). Specifically, amplitude will be lower at symmetrical than asymmetrical conditions between 300 and 600 ms post stimulus onset at*

posterior electrodes. Establishing the presence of an SPN in the frontoparallel conditions is essential step when measuring perspective cost.

Hypothesis 2

SPN perspective cost will approximate zero in both tasks (red bars in Figure 3). This is the critical test of our research question.

Hypothesis 3

SPN amplitude will be larger (i.e. more negative) in the Regularity task than the Luminance task. This is of secondary interest, however, plausible given previous work.

We have now redone the power analysis section and the study plan table. This explains how power is 0.9 with alpha 0.02 for the foundational hypothesis 1 (even if we finish at N=48) and the theoretically interesting hypothesis 2. This also applies to the less interesting hypothesis 3, although estimated effect size is more speculative. This should allow submission to more selective journals.

In the original submission, we used a naïve approach to sequential analysis, and did not adjust the alpha level for the number of looks. We have now improved this by using the Pocock-like correction factor to adjust alpha. We now plan three looks, rather than four. Furthermore, we have decided to yoke the sequential analysis to analysis of perspective cost in the Regularity task only. These decisions were not demanded by reviewers, but they are justified and explained in the new manuscript:

*"**Power analysis***

Power analysis for Hypothesis 1

As explained in the next section, we may finish data collection at N=48. To confirm Hypothesis 1, we need to find a significant SPN the frontoparallel conditions of both tasks. The frontoparallel SPN is likely to be smaller in the Luminance task. Karakashevska et al. (RR1) found that SPN amplitude in the Luminance tasks with a static frame was -0.93 microvolts, with a Cohen's d_z of

-0.95. If this is the true effect size, our minimum sample already exceeds 99% power (N = 48, alpha = 0.02, one-sided one-sample t test). Conservatively assuming that true effect size is merely medium sized (Cohen's d_z = -0.5), power exceeds 0.9 with our minimum sample of 48.

Minimum effect, sequential analysis and power analysis for Hypothesis 2

Our theoretically interesting Hypothesis 2 predicts zero perspective cost. This requires finding evidence of absence, not just absence of evidence. A non-significant one sample t test is only absence of evidence and is thus inconclusive. We will therefore take an alternative approach and demonstrate that SPN perspective cost is significantly less than our a priori definition of a small but meaningful effect. As explained in Karakashevska et al. (RR1), a good definition of a small but meaningful SPN perspective cost is 0.35 microvolts.

We will employ two separate one-sided, one-sample t tests to analyse perspective cost. The first t test examines whether perspective cost significantly less than 0.35 microvolts. If this is significant, we will conclude that there is NO perspective cost. The second t test examines whether perspective cost is significantly more than 0 microvolts. If this is significant, we will conclude that there IS a perspective cost.

It is logically possible for perspective cost to be both significantly more than 0 microvolts and significantly less than 0.35 microvolts. This outcome would be hard to interpret, but it is very unlikely to happen: if perspective cost is significantly less than 0.35, it will almost certainly be statistically indistinguishable from 0. It is also possible, but unlikely, that we will see a significant SPN perspective advantage. We can set aside these two unlikely outcomes when planning the research.

We will use a sequential analysis and apply the one-sided one-sample t tests when sample size reaches certain pre-defined cut points (N = 48, N=96 and N=120). The first analysis will be conducted at N=48. If neither one-sided one-sample t test is significant, we will collect more participants and re-analyse. We are constrained to use sample sizes which are multiples of 24, because this covers all possible block orders.

It could be that the perspective cost is eliminated in one task but not the other (contrary to our predictions, Figure 3). The sequential analysis will thus be guided by emerging patterns in the Regularity task only. The decision to yoke the sequential analysis to emerging results in the Regularity task, while ignore emerging results in the Luminance task, has logistical advantages. One disadvantage is that we risk terminating the experiment while results of the Luminance task remain statistically ambiguous. Another alternative would be to yoke the sequential analysis to mean perspective cost across both tasks. However, this is less sensible if trends unexpectedly diverge (again contrary to predictions, Figure 3). Yoking sequential analysis to the Regularity task only is thus a reasonable compromise.

Power analysis builds on known SPN effect sizes from database called the 'Complete Liverpool SPN catalogue' (https://osf.io/2sncj/). We can estimate that a within subject 0.35 microvolt SPN modulation would likely have a Cohen's d_z of 0.344 (Makin et al., 2022). If the true mean perspective cost is 0 microvolts, we can estimate it will be around 0.344 SDs away from the 0.35 microvolt threshold. Conversely, if true mean perspective cost is 0.35 microvolts, then we can estimate that it will be around 0.344 SDs away from 0 microvolt threshold. The 0 and 0.35 microvolt thresholds were used and found to be appropriate in Karakashevska et al. (RR1).

Our sampling plan aims for 90% power an alpha level of 0.02 when testing Hypothesis 2. Our maximum sample 120 provides approximately 95% power for finding an effect of 0.344 standard deviations with a one-sided one-sample t test. Therefore, if Hypothesis 2 is correct, and true perspective cost is 0 microvolts, N=120 exceeds 95% power for finding that perspective cost is significantly less than 0.35.

However, the sequential analysis allows three looks at the data (at N=48, N= 96 and N=120). To avoid increasing type 1 error through multiple comparisons, we will use the Pocock-like correction factor to adjust alpha at each look. The Pocock correction, similar to the Bonferroni correction (which divides alpha by the number of looks), offers increased efficiency by placing greater emphasis on earlier looks in the analysis. The adjusted alpha levels and statistical power are shown in Figure 5. There is approximately 50% chance of finding a significant effect and terminating the experiment at N=48, an 80%

chance of doing so at N=96, and a 90% chance at N=120. The sequential sampling and analysis thus achieves the desired 90% power with cumulative alpha of 0.02.

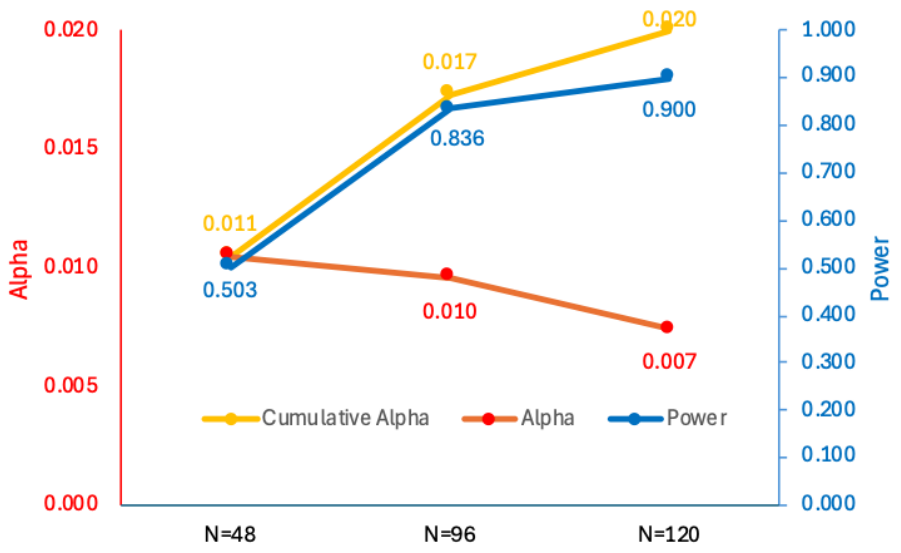


Figure 5. Change in power across the looks of the sequential analysis. *We will test Hypothesis 2 with a minimum effect testing approach. If perspective cost is significantly more than 0 microvolts, or significantly less than 0.35 microvolts at the pre-defined sampling points shown on the X axis we will terminate the experiment. If these results are non-significant at N=48, we will increase to N= 96, and then again to N=120. The final alpha level (0.02) accumulates across the three looks (orange). Alpha at each look is adjusted using the Pocock-like correction factor (red). Power increases at each look to reach the desired 0.9 threshold at N=120 (blue).*

Power analysis for Hypothesis 3

To confirm Hypothesis 3, we need to find a significant main effect of Task in a Task X Angle repeated measures ANOVA. An a priori estimate of effect size comes from a recent unpublished SPN study. Much like the planned experiment, this unpublished study also had Regularity and Luminance tasks in separate blocks of a within subject's design. Furthermore, it used very similar frontoparallel symmetrical and asymmetrical dot stimuli. The effect size ηp^2 *associated with the main effect of Task was 0.3. If this is the true effect size, our minimum sample provides 98% power (N = 48, alpha = 0.02). We will not add more participants if this effect is not significant. This ANOVA may also*

reveal an unexpected main effect of Angle and/or an unexpected Task by Angle interaction. We have not powered the experiment to find these unexpected effects, so we might miss them if they are small.

3. Please examine the requirements of the PCI-RR friendly journals. If you wish to publish your registered report to a journal with high power thresholds following peer review at Stage 2, you may be required to collect data beyond 60% power (which you achieve with 48 participants).

As explained above, our sequential analysis does achieve the higher-level 90% power with alpha 0.02 for the all-important hypotheses 2. Look one has 50% power (rather than 60%) now that we use the appropriate alpha correction for multiple looks, but data collection will continue if the results are not significant at N=48, so overall power is 90%.

Following these positive reviewer comments, and subsequent edits following these comments, I believe your manuscript has potential for a Stage 1 in-principle acceptance. I therefore request a revision and resubmission addressing the reviewers and recommenders feedback. Please note that PCI-RR is closed for resubmissions until the 1st September to accommodate reviewer and recommender holiday schedules.

Yours sincerely,

Grace Edwards

Daniel Baker

Review of Karakashevska, Batterley & Makin, 'Do they look virtually the same: extraretinal representation of symmetry in virtual reality', stage 1 registered report submitted to PCIRR.

Summary

This study proposes to extend some recent work by the authors by using virtual reality. It is an excellent candidate for a registered report, as the previous work permits

credible and precise estimates of effect sizes. The main purpose is to see if VR environments cause EEG signals relating to symmetry to become fully perspective-invariant. The stage 1 report is well-written and exceptionally clear, and so I have only some minor suggestions and requests for clarification.

We are pleased reviewer 1 is positive about the manuscript. As described in the response to the editor, we have also improved our approach to the sequential analysis, although this was not demanded by either reviewer. We now adjust alpha for multiple comparisons. We have expanded the power analysis section substantially.

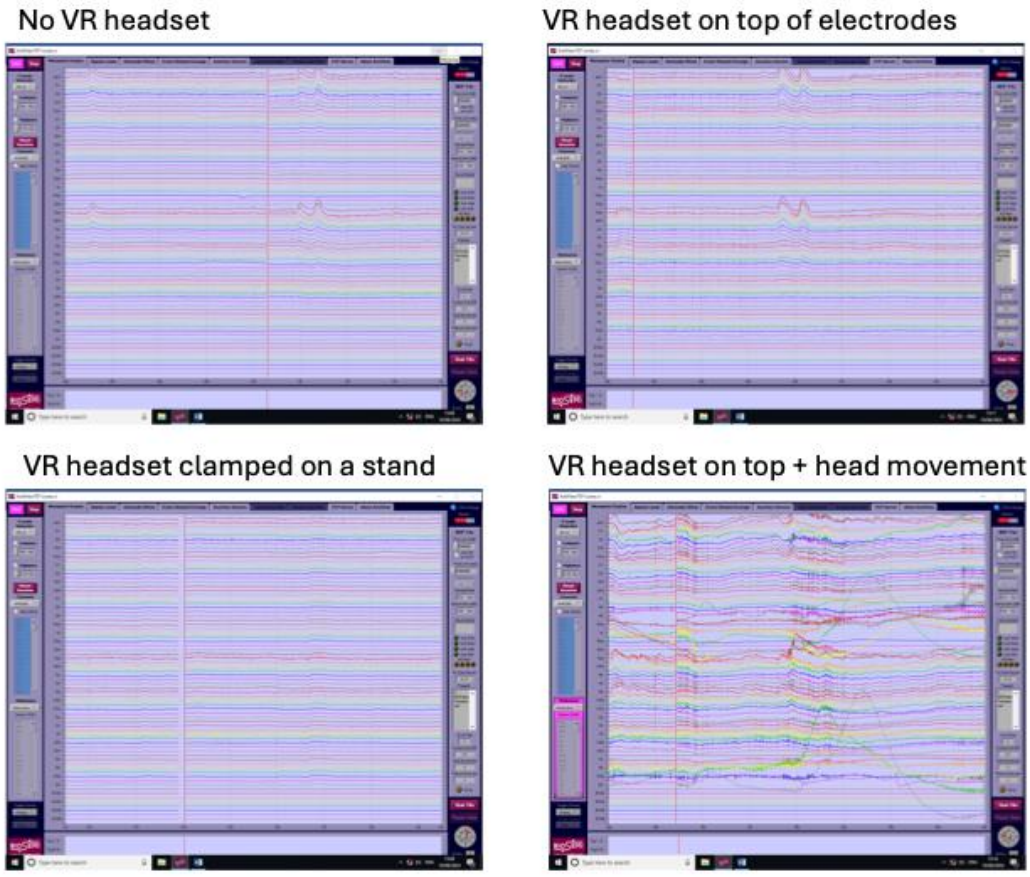
Specific points

1. I think the use of equivalence testing is appropriate and well thought-through here. However it is now quite common to report Bayes factors alongside the results of more traditional frequentist tests. These help to distinguish between null effects that are underpowered, versus being caused by the absence of an effect. I'd recommend including these statistics here in addition to the planned analyses.

This is a reasonable suggestion, and we are not opposed to Bayes factors. However, the PCIRR guidelines prevent us from mixing Bayesian and frequentist hypothesis testing methods (as explained by the editor above). Therefore, we will keep the frequentist analysis in the main manuscript and include the Bayesian analysis in the supplementary material.

2. Does the VR headset interact with the EEG system, either physically (i.e. straps moving electrodes), or electrically (greater line noise)?

To test this problem, we have tried multiple set ups with the VR headset and the BioSemi system. The VR headset will be placed on top of the electrodes with the straps, carefully clasping them so the weight is equally distributed. Participants will then be instructed to place their head on a chin rest and retain from performing any harsh movements. As seen in the Figure below, it is only head movements which introduce substantial noise. The mere presence of the VR headset does not.



Screenshots of VR headset set up with BioSemi versus no VR headset.

As an additional check, we will record additional EEG data from each participant for with and without VR. If the VR helmet significantly degrades the EEG signal, we will adjust pipeline parameters proportionally. This check is now explained in the manuscript:

"Signal quality check

The power analyses presented were informed by previous SPN research. While there is a possibility that integrating Virtual Reality (VR) could reduce signal quality below typical standards, necessitating larger sample sizes to detect the expected effects, prior research (Tauscher et al., 2019) and our testing with the VR headset suggest this is unlikely. To assess the signal quality in this project, we will record EEG data in one 32-trial block with the VR headset and one 32-

trial block without it (same stimuli shown on a screen) for each participant after the main experiment. Signal quality will be assessed by the number of trials rejected based on the criteria outlined above.

Ideally, we aim to find evidence of no significant difference in signal quality, rather than absence of evidence. However, as we lack a precise definition of what constitutes a meaningful difference in EEG signal quality, a non-significant difference in trial rejection ($p > 0.05$) will be considered sufficient to act as if no substantial differences exist. An initial signal quality analysis will be conducted at N=24. If the VR headset significantly degrades the EEG signal, it is crucial to identify this before collecting a larger sample. If signal quality is compromised, we will adjust the analysis pipeline parameters to accommodate this reduction. First, we will adjust the +/- 100 microvolt trial exclusion thresholds until trial exclusion aligns with the non-VR condition. Second, if there is a 20% reduction in EEG signal quality with the VR headset, we will increase the maximum sample size by 20% (to 144) and recalculate the look points for the sequential analysis. These adjustments will only be made if there is a significant reduction in signal quality when using the VR headset."

3. In Hypothesis 1, please clarify that lower amplitude means more negative. You do this for Hypothesis 2, but would be good to have it in earlier too.

It is easy to get things back to front when talking about small and large SPNs. We have thus added a useful clarification the legend of Figure 1:

"Note that a large SPN is one that falls a long way below zero. If the SPN is said to be 'reduced' or 'weaker', it does not fall so far below zero. A relatively large SPN would be -3 microvolts, while a relatively small SPN would be -0.5 microvolts."

We have also changed wording of Hypothesis 1:

"We expect to observe an SPN difference wave at posterior electrodes between 300 and 600ms post-stimulus onset in all conditions. In other words, we expect

mean amplitude to be lower for symmetrical than asymmetrical trials in all blocks (grey and black bars in Figure 3).”

4. The very last point in the table at the end says “Power for the one-sided t tests used in these analyses = 0.95”. But this is only true if the full sample of N=120 is tested – an earlier bullet point in the same column explains this more clearly. So I’d simply omit this last point to avoid any confusion.

We have removed this point now to avoid any confusion. We have also redone the power analysis section and the study design table (see response to editor).

5. Should the first part of the title have a question mark? It feels like it should, but looks weird if it’s before the colon, and then also seems wrong at the end!

We agree, so we have slightly tweaked to title to avoid the question mark awkwardness!

Anonymous reviewer

The authors suggest a study which will investigate whether previous findings regarding the brain’s processing of symmetrical vs asymmetrical stimuli in different viewing conditions also hold in immersive virtual reality (VR). More precisely, the experiment shall test the hypothesis that the additional information (e.g., stereoscopic depth cues) available in such immersive settings cancel out an effect previously found for the SPN (Sustained Posterior Negativity; an ERP component). Namely, showing the stimuli with a perspective distortion (i.e., like looking at them from an angle) leads to a reduction in the (absolute) amplitude of the SPN (“perspective cost”), particularly if participants focused on other properties of the stimuli (e.g., their luminance) rather than their symmetry. The motivational argument for this new approach is that VR provides strong and intuitive depth-cues which might support the brain in forming a viewpoint independent representation of the stimulus. A truly viewpoint independent representation should (by definition) not vary as a function of the viewpoint dependent “retinal” representation of the stimulus. Therefore, if the “perspective cost” was zero in immersive (i.e., more naturalistic) conditions, this would be evidence that the SPN can reflect symmetry processing based on a viewpoint independent representation.

The authors therefore suggest a VR-based experiment which implements a design that (in similar forms) was previously used in conventional 2D-screen settings to investigate the SPN. The data gathered via this experiment shall (centrally) test the hypothesis that in such immersive conditions the amplitude of the SPN does not differ between presentations of the stimuli with or without perspective distortions. To this end, they plan to test (at least) 48 healthy participants in a combined EEG+VR setup and use equivalence testing on the resulting EEG data to test whether there is evidence to reject the null hypothesis that the SPN amplitude is different in VR conditions with and without perspective distortion.

I enjoyed reading the study proposal and learning about the field of symmetry processing and the SPN. The study appears to be based on an impressive body of research addressing similar questions. The authors demonstrate extensive experience and experimental insights into how the SPN behaves under certain conditions and how to study it effectively. Using VR to expand this knowledge base and to investigate the phenomenon of “perspective cost” under conditions that might substantially facilitate the formation of viewpoint-independent representations is a promising and informative endeavor. I look forward to reading about the results.

We are pleased reviewer 2 is positive about the manuscript. As described in the response to the editor, we have also improved our approach to the sequential analysis. We now adjust alpha for multiple comparisons. We have expanded the power analysis section substantially.

I noticed a few aspects in the registration that could benefit from clarification, which I would like address below:

Validity of the research question(s)

Above, I attempted to articulate the underlying research question in my own words. I hope it accurately reflects the authors’ actual aims. (The only explicitly stated “research question” I found in the report was in the table at the end: “Can we achieve extraretinal representation of planar symmetrical dot patterns in virtual reality?”. However, this

appears to be more of a subsidiary question related to Hypothesis 1, while Hypothesis 3 seems to be the central focus of the study.) The (assumingly) central question seems well-derived from previous findings regarding the SPN as well as assumptions and insights gathered in other studies and fields about VR as an experimental tool. I would recommend keeping the scope/formulation of the (explicitly phrased) research question narrow enough—for example, focusing on the modulation of the SPN rather than about how the brain generally processes (a)symmetry—so that the suggested experiment can provide the data to answer it. Based on the introduction and the framing of the study's motivation, I conclude that the authors have a concrete and valid research question in mind. I recommend that a specific formulation of this (central) research question be added to the study plan (e.g., in the table at the end of the document).

Hypotheses

The authors suggest three hypotheses, all of which seem logically and plausibly derived from previous research. However, it would be helpful if the authors clarified the function of each hypothesis. H3 formulates the core claim of the study. H1 seems to describe a necessary (?) pre-condition for studying H3. H2 appears corollary and independent of H3 (i.e., H3 can be tested irrespective of the outcome for H2). This makes the role of H2 somewhat unclear. It could serve as a form of positive quality control, but this function is not explicitly mentioned.

This understanding of the function of each hypothesis is correct. It was perhaps confusing how the less interesting and independent H2 was sandwiched between the background H1 and theoretically interesting H3. We have now reordered the hypothesis, so H1 is the foundation, H2 is the interesting one, and H3 is the secondary expectations one:

> *Hypothesis 1*
> *We predict that there will be an SPN difference wave at posterior electrodes between 300 and 600ms post-stimulus onset in all conditions. In other words, we expect amplitude to be lower for symmetrical than asymmetrical trials in all*

blocks (grey and black bars < 0 in Figure 3). Unless VR dramatically changes symmetry processing, this essential foundation is likely to be in place.

Hypothesis 2
We predict SPN perspective cost will approximate zero in both tasks (red bars in Figure 3). This is the critical test of our research question.

Hypothesis 3
We predict SPN amplitude to be larger (i.e. more negative) in the Regularity task than the Luminance task. This is of secondary interest, but plausible given previous work.

The main prediction is that SPN amplitude will be the same in frontoparallel and perspective conditions. That is, perspective cost should be zero. To interpret this as zero perspective cost, it is essential that we observe SPNs at least in the frontoparallel conditions (Hypothesis 1 confirmed). We do not want perspective cost to be absent simply because the SPN is absent!

I have a few concerns with the statements made in the columns “Interpretation given different outcomes” and “Theory that could be shown wrong by the outcomes” (in the final table):

H1:

- The authors will conclude that “something in the experiment went wrong” if H1 is not supported by data from both frontoparallel conditions. I think, this is a good but strict criterion. Does this mean, that if there is no significant SPN in one of these two conditions, the rest of the data can and will not be analyzed and interpreted in any case? What happens for the case that there is evidence for H3 in the regularity condition but no evidence for H1 in the luminance condition (or vice versa)?
- The authors write that “We are also confident we will observe SPNs, **albeit smaller** in the perspective conditions given the results of Karakashevska et al. (forthcoming 1,2)” which seems to contrast with H3. If the authors expect smaller SPNs in the perspective conditions (i.e.,

perspective cost), wouldn't they want to test this hypothesis (and reject the H0 that there is no perspective cost) instead of the other way around?

- Furthermore, if H1 is not supported by data from the two "perspective" conditions, the authors will conclude that "in a virtual reality environment, the brain is blind to extraretinal symmetry". This claim is way too strong, in my opinion. (A) If participants are behaviorally capable of performing the symmetry task in the perspective condition, this is strong evidence that "the brain" is not blind to this kind of symmetry. Any conclusions should be restricted to the SPN and the processes it reflects). (B) Additionally, the generalization to VR environments as such is not justified. The results might be specific to the design, environment, setup, hardware, or stimuli used in this study. Whether such a finding generalizes to other VR experiments needs to be tested explicitly. (C) Finally, the credo "absence of evidence does not imply evidence of absence" also applies here.
- As the authors write themselves, "the brain is not sensitive to symmetry presented in virtual reality environments" is not a particularly interesting or probable theory to disprove. Isn't the aim (of H1) rather to demonstrate that the SPN can also be measured and studied in immersive settings? This would refute the claim that the SPN is merely an artifact of unnaturalistic, simplified, abstract 2D lab experiments.

H2:

- As with H1, I am not a fan of the (potential) conclusion that "the task modulation of SPN amplitude does not apply in virtual reality environments [if the data does not support H2]". I would advise against generalizing such findings to all virtual reality environments/studies.

H3:

- As with H1, the claim that "symmetry presentations in VR are not sufficient for achieving extraretinal symmetry representation [if there is perspective cost for both tasks]" is too strong, in my opinion. This should

be more focused on the SPN and the experimental design/setup of the study.

- Furthermore, it would be valuable to know what the interpretation will be if there is support for H3 in only one of the two tasks.
- “The brain codes extraretinal symmetry in a different way that it codes frontoparallel symmetry” appears overly general. Even if there is no difference in SPN observed in this experiment, it does not justify conclusions about how “the brain” universally processes symmetry. Regarding the sentence "We will acknowledge that it is not possible to achieve equivalence in the symmetry signal for retinal and extra-retinal representations of symmetry," it seems unclear

Considering these comments and other things, we have substantially reworked the study design table. There are many possible outcomes and counts against a different theory. We have now taken a more systematic approach. Please see the all-new study design table at the end of the manuscript.

Experimental setup/design:

The experimental setup and design seem feasible, sound, and mostly well-thought through. Potential challenges might arise from the fact that (in comparison to the previous experiments which the authors refer to) this study will be conducted in VR. Besides the positive aspects of VR (which the authors outline), it also brings additional obstacles. Foremost, putting a VR headset on top of an EEG cap is likely to introduce additional noise into the EEG measurements, potentially leading to a lower signal to noise ratio (SNR) compared to previous data sets.

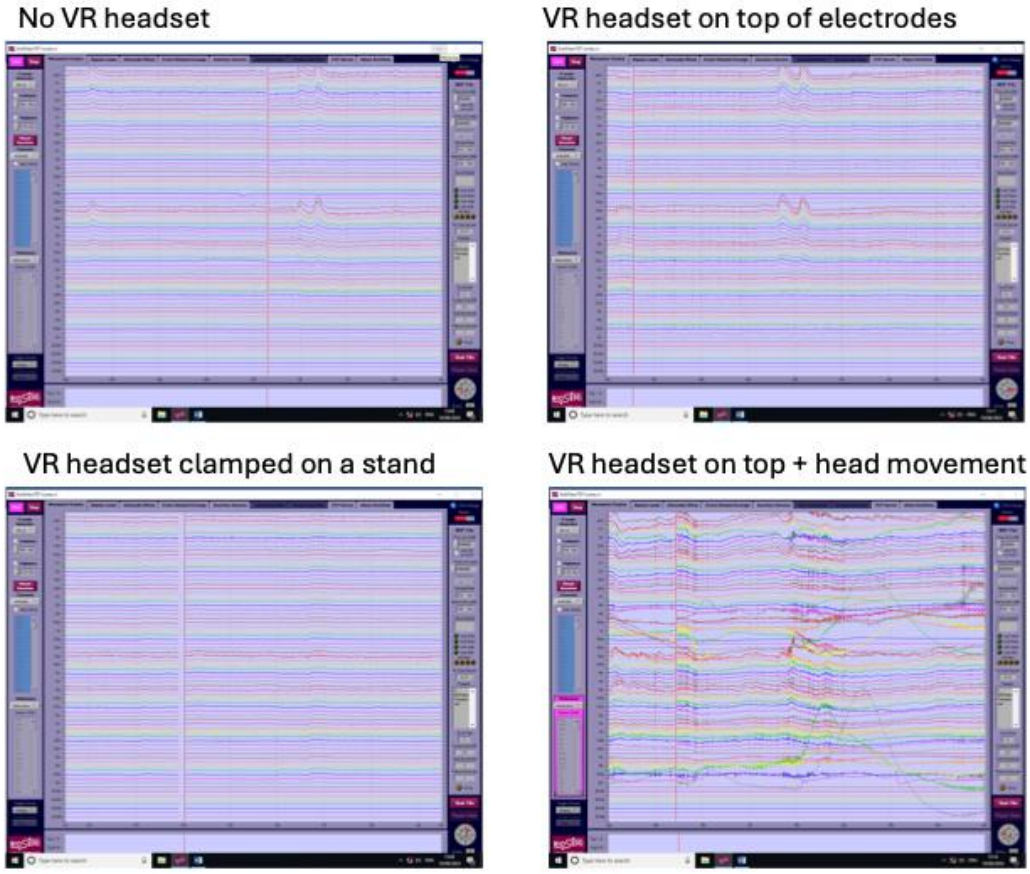
Consequently, effect sizes in the data may be smaller than those observed in previous studies. It is difficult, if not impossible, to estimate the magnitude of this impact beforehand. Therefore, I believe it is reasonable to base power calculations on recent non-VR studies (as done by the authors). However, to err on the side of caution, the authors might consider adjusting power calculations to account for the potentially lower SNR and reduced power due to the VR setup. This could involve increasing the

number of trials or participants to compensate for any anticipated decrease in data quality/SNR.

At least, it should be discussed (at latest when interpreting the results) that the power analyses conducted may be overly optimistic as they do not reflect potentially interfering effects of a VR setup.

This is a very reasonable concern that was also raised by reviewer 1.

To test this problem, we have tried multiple set ups with the VR headset and the BioSemi system. The VR headset will be placed on top of the electrodes with the straps, carefully clasping them so the weight is equally distributed. Participants will then be instructed to place their head on a chin rest and retain from performing any harsh movements. As seen in the Figure below, it is only head movements which introduce substantial noise. The mere presence of the VR headset does not. This is supported by the findings of Tauscher et al. (2019).



Screenshots of VR headset set up with BioSemi versus no VR headset.

As an additional check, we will record additional EEG data from each participant for with and without VR. If the VR helmet significantly degrades the EEG signal, we will adjust pipeline parameters proportionally. This check is now explained in the manuscript:

> ***"Signal quality check***
>
> *The power analyses presented were informed by previous SPN research. While there is a possibility that integrating Virtual Reality (VR) could reduce signal quality below typical standards, necessitating larger sample sizes to detect the expected effects, prior research (Tauscher et al., 2019) and our testing with the VR headset suggest this is unlikely. To assess the signal quality in this project, we will record EEG data in one 32-trial block with the VR headset and one 32-trial block without it (same stimuli shown on a screen) for each participant after the main experiment. Signal quality will be assessed by the number of trials rejected based on the criteria outlined above.*
>
> *Ideally, we aim to find evidence of no significant difference in signal quality, rather than absence of evidence. However, as we lack a precise definition of what constitutes a meaningful difference in EEG signal quality, a non-significant difference in trial rejection ($p > 0.05$) will be considered sufficient to act as if no substantial differences exist. An initial signal quality analysis will be conducted at N=24. If the VR headset significantly degrades the EEG signal, it is crucial to identify this before collecting a larger sample. If signal quality is compromised, we will adjust the analysis pipeline parameters to accommodate this reduction. First, we will adjust the +/- 100 microvolt trial exclusion thresholds until trial exclusion aligns with the non-VR condition. Second, if there is a 20% reduction in EEG signal quality with the VR headset, we will increase the maximum sample size by 20% (to 144) and recalculate the look points for the sequential analysis. These adjustments will only be made if there is a significant reduction in signal quality when using the VR headset."*

Another difference to the previous studies is the (more naturalistic and therefore) less

controlled background against which the stimuli pattern will be presented. I understand that this is a core feature of the study and do not want to criticize it. However, this might introduce confounds in the data, as, for example, in the perspective condition not only the stimulus will be non-symmetric (in the visual field) but also the background which may lead to changes in EEG potentials which are not related to stimulus processing (e.g., the background is asymmetric also for “symmetric pattern” trials in the perspective condition).

We had considered this, but we think it is not a problem because of the way the SPN is computed (symmetry – asymmetry). Overall, the screen will be less symmetrical in perspective blocks. But this will subtract equally from both symmetrical and asymmetrical trials within the perspective block, and not alter the difference between them.

I have some clarification questions regarding the sizing of the stimuli (i.e., the actual dot patterns). The authors write that the patters have a size of “approximately 7.5° of visual angle”, show dots with a diameter of 0.25°, and are presented at a distance of 4.13m (Fig. 6). To my understanding, this translates to an absolute width of the whole pattern of ~0.54m and 0.018m diameter for a single dot (diameter = tan(0.25°/2) * 4.13 * 2). This seems small for stimuli in VR (at 4m distance). Is the resolution of the Vive Pro Eye high enough to clearly see stimuli (dots) of this size (at the given distance)? The example environments which were provided by the authors seemed to hold placeholders for the stimuli patterns, but these looked substantially (by far) larger than the numbers mentioned above. Maybe these placeholders are/were not representative of the final design? To ensure a concrete understanding of the actual experimental setting it would be beneficial to see screenshots (or even blender models or Unity scenes) of the final layout of the scene containing an actual stimulus.

We have completed programming of the experiment, including the parameter choices like stimuli size and camera position. Unity uses arbitrary units that can be treated as equivalent to 1m when working with real-world scale objects and environments.
We have now revised this section of the manuscript as follows:

The Unity scenes, including the folders containing each experiment are available on OSF for peer review (https://osf.io/mzvy9/, Stage 1, Stimuli and experiment folder). Additionally, we are confident that participants are able to do the task with the size of the stimuli chosen. We have conducted a behavioural regularity discrimination task in VR where one of the conditions is the same as this EEG study, and participants have no trouble completing the task (median RT = 0.6068 s, mean error rate = 6%).

Furthermore, I find it challenging to reproduce the placement of the cameras with the provided written information. It's unclear around which point and axes the cameras will be rotated. Are these coordinates based on Blender rather than Unity? This issue aside, Figure 6 is very helpful and mostly self-explanatory (however, also here the numbers do not really add up: if the triangle C1-C2-Stimulus is equilateral [4.13], all inner angles should be 60 degree). What concerns me is the “tilt”: if the camera is rotated 15° downwards, the center of the stimulus pattern will be approx. 15° above the participant’s straight line of sight (i.e, the center of the field of view). That is quite a big eccentricity for VR. I know from own experience that stimuli with an eccentricity >15° (i.e., the upper half of the pattern in this setup) can become quite blurry in the Vive Pro Eye (due to the Fresnel lenses). Might this become a problem? Or am I misunderstanding the setup?

We apologise for the confusion with the added tilt. Please refer to the updated Figure 6 as well as Figure 4, illustrating this. By tilt we do not mean eccentricity, as the stimulus will remail foveal. The tilt of now 15° is on the x axis in the Unity coordinate system or in other words, a vertical rotation of the camera upwards. What this achieves is positioning the participant in a view slightly looking upwards towards the stimulus to destroy any regularity remaining in the centre of the stimulus.

Related to the size of the stimuli is another challenge that I envision for the experiment: participants will most likely perform eye movements to explore the patterns. The larger the patterns, the larger these eye movements will be. It is possible that there could be systematic differences between conditions (e.g., perspective vs frontoparallel) in terms of eye movements, which may/will influence the EEG signals. Do the authors have a plan to address this issue? How will participants be instructed regarding fixation

behavior? Will fixation and gazing behavior somehow be monitored? The Vive Pro Eye allows for measuring eye tracking. This could be an option (e.g., in order to show post hoc that there were no systematic differences between the conditions). Relying purely on ICA to clean the data from eye movement artefacts and gaze related EEG components (which do not need to be artefacts) might not be sufficient. I am not requesting the authors to add eye tracking, but I want to send a sign of warning. We see in comparable VR experiments a lot of eye movements which often are confounded with experimental manipulations and correlate with EEG findings (also in parieto-occipital sensors). This is not a bad thing per se but should be factored in when setting up a new study.

It is true that eye movement artefacts could be a problem, even though participants will be asked to fixate, and the relevant information will always be in a small and foveal region. We will take advantage of the eye tracker build into the VR system as recommended. We will also examine the number of eye movement related ICA components identified by the Adjust procedure. This is now explained in the manuscript:

> ***"Eye tracking***
> *In this project, it is essential to ensure that participants maintain consistent fixation across all four experimental blocks. We will use the eye tracker built into the VR headset to monitor fixation, defining it as less than a 2.5-degree change in eye position during the stimulus period. We expect the number of fixation breaks to be similar across all blocks. Additionally, we will analyse the output from the Adjust procedure to evaluate whether the number of eye movement-related components remains consistent across blocks. A non-significant difference in fixation breaks ($p > 0.05$) will be considered sufficient to proceed as if there are no significant differences in fixation compliance between the blocks."*

<u>EEG preprocessing</u>

The pipeline seems reasonable and well thought through. I only have minor comments/questions here:

- ICA rejection: I do not know the `Adjust()` function in MATLAB. I assume it has some settings or parameters which can be chosen to adjust the rejection criteria. For reproducibility, it would be good to mention/register the choice for these settings. Will the function make use of the EOG channels? Will ICA be run on continuous or epoched data?

Adjust is an algorithm that allows for automatic rejection of ICA components, optimised on large EEG data sets, combining stereotyped temporal and spatial artefact features. ADJUST features are optimised to capture blinks, eye movements and discontinuities of a dataset (Mognon et al., 2011). This removes the human bias of visually inspecting components and performing manual rejection. In our case, algorithm uses epoched data and does not use EOG channels (which we no longer plan to record). The function does not take arguments from the user. It just takes EEG data and name for text output (ComponentsToRemove = ADJUST(EEG,'ADJUST_Report.txt').

- Channel rejection: also here, it'd be great to register which criteria (even if applied visually/manually) will guide the selection of channels which are to be rejected. Will this rejection be performed on continuous or epoched data?

Channel rejection is based on visual inspection of EEG variance in epoched data. When viewed in this way, bad channels appear as outliers with extremely high variance. This works better without excessively rigid a priori criteria (e.g. variance > 1000 = remove channel). We have now explained this in the manuscript:

- Trial rejection: the authors plan to reject every trial with an amplitude >100uV. Does this apply to all channels (i.e., will a single channel which reaches >100uV at one point of the trial lead to rejection of the entire trial)? Or only to channels in the ROI?This might be a very strict criterion (especially when applied to all channels) in VR-EEG settings that leads to high rejection rates.

This will apply to all channels, as now clarified in the manuscript. We apply this criterion to the majority of EEG experiments we run, and we believe it should be applied here.

Statistics

H1: Will the significance criterion for the four t-tests be corrected for the number of tests (if so, by which procedure)?

[The main effect of Task is now H3, and no perspective cost is now H2].

We have now changed H1 so that we only predict two significant SPNs, in the frontoparallel conditions. These will not be corrected for multiple comparisons because these are two a priori predictions (although we think it is unlikely they would become non-significant if they were).

H2 predicts that the SPNs will be just as large in the perspective block as the frontoparallel block. This means we silently predict that SPNs will be significant in perspective conditions. However, the research does not critically depend on SPNs in the perspective blocks on the same way, so we have focused H1 on the frontoparallel blocks.

H2: What will be the interpretation if (instead of only the factor "*Task*"—as hypothesized) also or only the interaction between the two predictors ("*Task*" and "*Angle*") turns out significant? What if the main effect "Angle" is found significant—will this influence the interpretation of H3? Will the testing of additional participants continue even if after 48, 72, … participants a solid effect of "Angle" manifests?

We will not use this ANOVA to increase sample size – that is, we will not collect more participants if this the main effect of Task is non-significant.

It is true that we may find an effect of Angle or Task X Angle interaction, contrary to our predictions. A main effect of Angle alone does not influence interpretation of H3.

We have now addressed all these considerations in a new power analysis section:

"Power analysis for Hypothesis 3

To confirm Hypothesis 3, we need to find a significant main effect of Task in a Task X Angle repeated measures ANOVA. An a priori estimate of effect size comes from a recent unpublished SPN study. Much like the planned experiment, this unpublished study also had Regularity and Luminance tasks in separate blocks of a within subject's design. Furthermore, it used very similar frontoparallel symmetrical and asymmetrical dot stimuli. The effect size □p2 associated with the main effect of Task was 0.3. If this is the true effect size, our minimum sample provides 98% power (N = 48, alpha = 0.02). We will not add more participants if this effect is not significant. This ANOVA may also reveal an unexpected main effect of Angle and/or an unexpected Task by Angle interaction. We have not powered the experiment to find these unexpected effects, so we might miss them if they are small."

H3: To provide evidence that there is no "perspective cost" the authors plan to apply an equivalence testing strategy by refusing the hypothesis that there is a meaningful difference in the SPN for the perspective as compared to the frontoparallel condition. Hereto they only specify an "upper" boundary (-0.35uV) for the equivalence test. To my knowledge, it is common to also provide and test against a lower boundary if one wants to show equivalence. It would be great if the authors provided concrete arguments why they think that testing only one side of the equivalence boundaries is sufficient.

This planned analysis is consistent with Karakashevka et al. (RR1). We should technically use the term 'minimum effects test' rather than equivalence test. We have now explained this fully in our expanded section:

"Minimum effect, sequential analysis and power analysis for Hypothesis 2
Our theoretically interesting Hypothesis 2 predicts zero perspective cost. This requires finding evidence of absence, not just absence of evidence. A non-significant one sample t test is only absence of evidence and is thus inconclusive. We will therefore take an alternative approach and demonstrate that SPN perspective cost is significantly less than our a priori definition of a small but meaningful effect. As explained in Karakashevska et al. (RR1), a good definition of a small but meaningful SPN perspective cost is 0.35 microvolts.

We will employ two separate one-sided, one-sample t tests to analyse perspective cost. The first t test examines whether perspective cost significantly less than 0.35 microvolts. If this is significant, we will conclude that there is NO perspective cost. The second t test examines whether perspective cost is significantly more than 0 microvolts. If this is significant, we will conclude that there IS a perspective cost.

It is logically possible for perspective cost to be both significantly more than 0 microvolts and significantly less than 0.35 microvolts. This outcome would be hard to interpret, but it is very unlikely to happen: if perspective cost is significantly less than 0.35, it will almost certainly be statistically indistinguishable from 0. It is also possible, but unlikely, that we will see a significant SPN perspective advantage. We can set aside these two unlikely outcomes when planning the research.

We will use a sequential analysis and apply the one-sided one-sample t tests when sample size reaches certain pre-defined cut points (N = 48, N=96 and N=120). The first analysis will be conducted at N=48. If neither one-sided one-sample t test is significant, we will collect more participants and re-analyse. We are constrained to use sample sizes which are multiples of 24, because this covers all possible block orders.

It could be that the perspective cost is eliminated in one task but not the other (contrary to our predictions, Figure 3). The sequential analysis will thus be guided by emerging patterns in the Regularity task only. The decision to yoke the sequential analysis to emerging results in the Regularity task, while ignore emerging results in the Luminance task, has logistical advantages. One disadvantage is that we risk terminating the experiment while results of the Luminance task remain statistically ambiguous. Another alternative would be to yoke the sequential analysis to mean perspective cost across both tasks. However, this is less sensible if trends unexpectedly diverge (again contrary to predictions, Figure 3). Yoking sequential analysis to the Regularity task only is thus a reasonable compromise."

A meta comment

Throughout the report, the authors refer to (some of their) previous works by citing "Karakashevska et al. (forthcoming …)" that have relevant explanations and method descriptions for the present study. As this previous work seems to be unpublished and

not (yet) accessible, this makes it difficult/impossible to fully understand, evaluate, or reproduce the according sections.

We have now clarified this throughout the manuscript. Karakashevska et al. (forthcoming 1) is now Karakashevska and Makin (2024) which is under review. Karakashevska et al. (forthcoming 2) is a now Karakashevska et al. (RR1). Stage 2 registered report currently under review, the preprint for which can be accessed here https://osf.io/preprints/psyarxiv/z9c28.