Dear Dr. Karakashevska,
We have been fortunate to receive insightful and thorough comments from four expert reviewers. We agree with the reviewers that your research question is well-motivated and you have a strong proposed design. There are several areas where the reviewers suggest clarifications, to highlight a few:

Drs Apthorp and Cottereau both indicate the need for more information related to task difficulty and its potential impact on the results. They suggest if the task is too easy, the participants will be able to focus on the regularity of the stimuli, which would undercut the automaticity you would like to study and could reduce the perspective cost, according to a study you previously ran (Makin et al., 2015). Please note, there seems to be some inconsistency with the mention of task-based effects in Makin et al., (2015), and the earlier assertion in the manuscript (Page 2) which suggests "The SPN is comparable when participants are classifying stimuli in terms of symmetry or in terms of a different dimension such as colour…" and "Task relevance of symmetry has a relatively small effect on SPN amplitude…". It is plausible these statements are in reference to frontoparallel stimuli only, but please clarify.

We are pleased the editors and expert reviewers see value in our proposed research.

Regarding the apparent inconsistency: Yes, these statements are about frontoparallel stimuli only. Our current theory is that retinal symmetry, such as symmetry in the frontoparallel plane, is processed automatically, whatever the participant's task. In contrast, extraretinal symmetry, such as that viewed from an angle, may only be processed, when it is task relevant, using a set of elective processes. We use the term retinal image to refer to planar patterns of symmetry present in the input, and therefore defined in retinotopic coordinates. The current work tests the second claim. It could be that when there are sufficient visual cues supporting 3D interpretation, extraretinal symmetry is also processed automatically.

We have now changed this sentence to clarify:

> *"Indeed, analysis of a database called the complete Liverpool SPN catalogue (https://osf.io/2sncj/) suggests that symmetry is processed automatically whenever it is present in the retinal image (Makin et al., 2022). The current work investigates SPN responses to stimuli seen from perspective viewpoints, that distort symmetry in the retinal image. Such 'extraretinal' symmetry might not be processed automatically."*

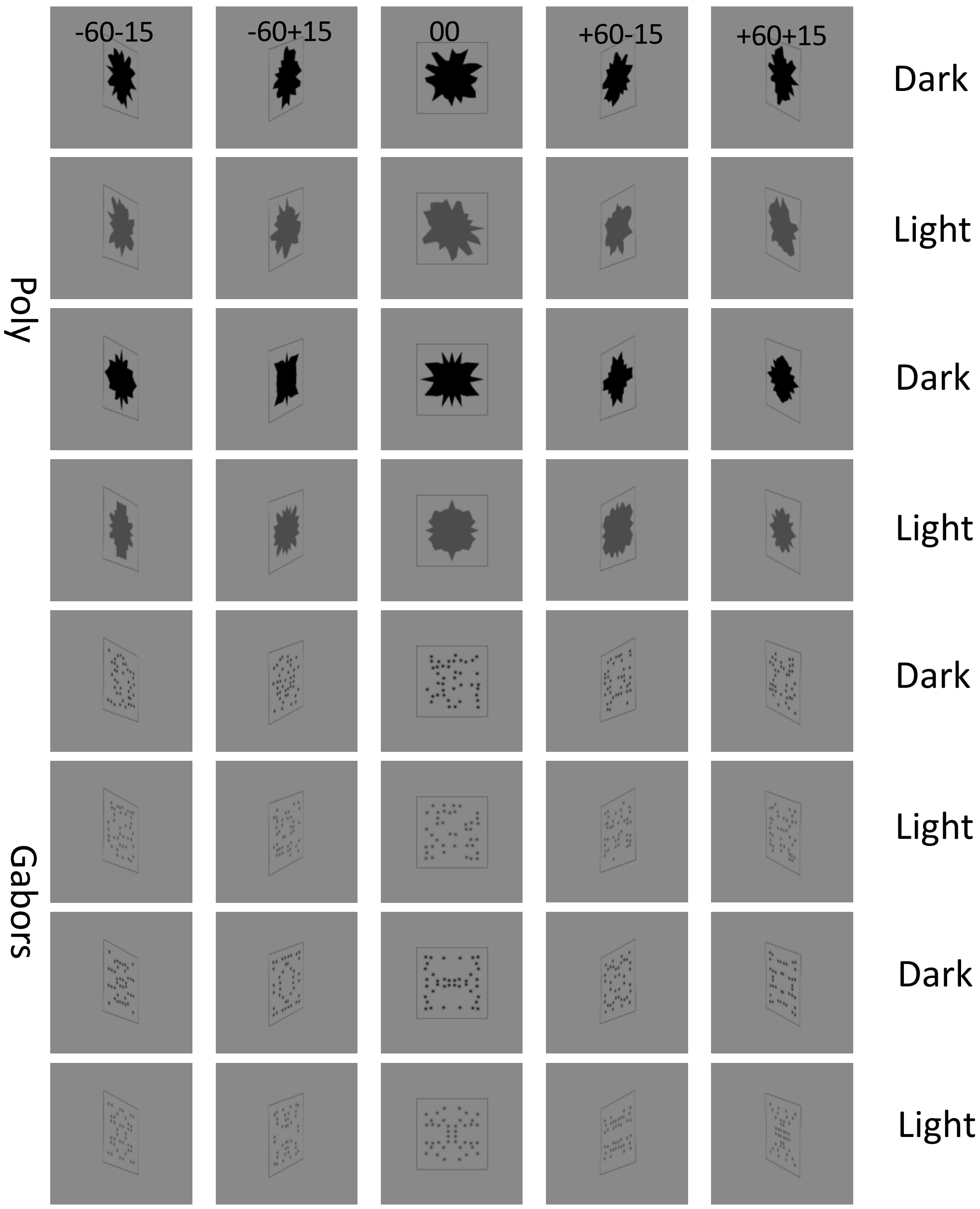
We address concerns about task difficulty from Dr Apthorp and Dr Cottereau extensively below.

Our anonymous reviewer also indicated some previous studies where participants struggled to detect mirror-symmetry in dot-stimuli, which was overcome with the use of contours. Please motivate the use for the dot-stimuli in your Stage 1.

Some aspects of our design are justified by unpublished results from a recent experiment with frontoparallel and perspective dot-stimuli and polygons. We refer to this as previously currently unpublished paper as 'Karakashevska et al. (forthcoming). To be clear, Karakashevska et al. (forthcoming) will be a different publication and is not part of the registered report.

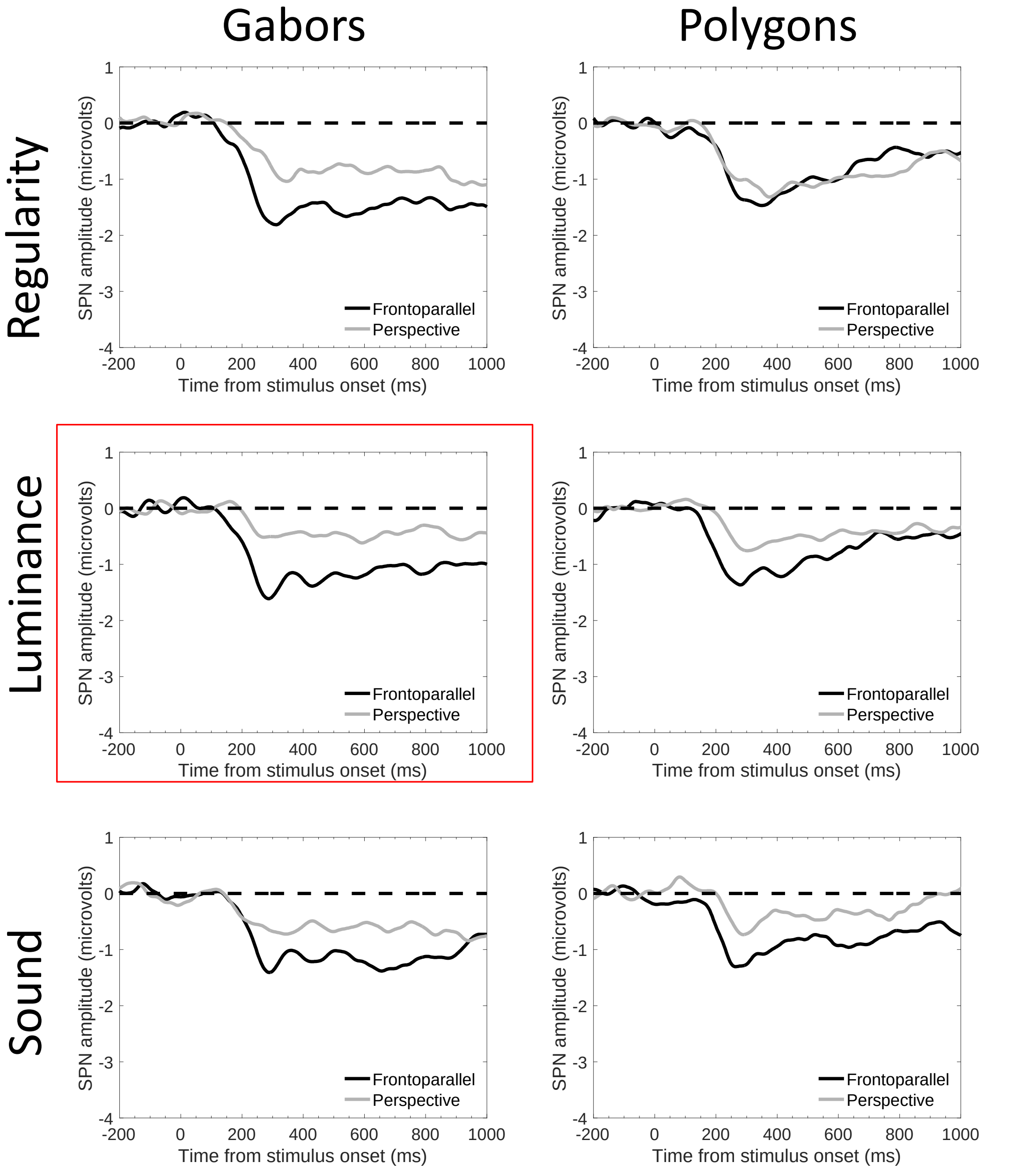
We did not discuss Karakashevska et al. (forthcoming) in the Stage 1 RR because it has not yet been peer reviewed, and the results analysis was not completed at the time of submission. The analysis is now finished, and it is relevant to addressing reviewers' comments, especially regarding polygons and task difficultly.

The stimuli of Karakashevska et al. (forthcoming) are shown in Figure 1 below:



***Response to Editors Figure 1**: Stimuli used in Karakashevska et al. (forthcoming). The Gabor stimuli are similar those we intend to use in the proposed research.*

Karakashevska et al. (forthcoming), included three tasks. In the first task, participants discriminated regularity. In the second, participants discriminated luminance (the same as the task in our proposed research). In the third, they judged congruence between element luminance and the pitch of a concurrent sound. There were 40 participants in each task. As shown in Figure 2, an overall perspective cost was found (Main effect of Angle, $F(1,117) = 18.459$, $p < 0.001$, $\eta p^2 = 0.136$).



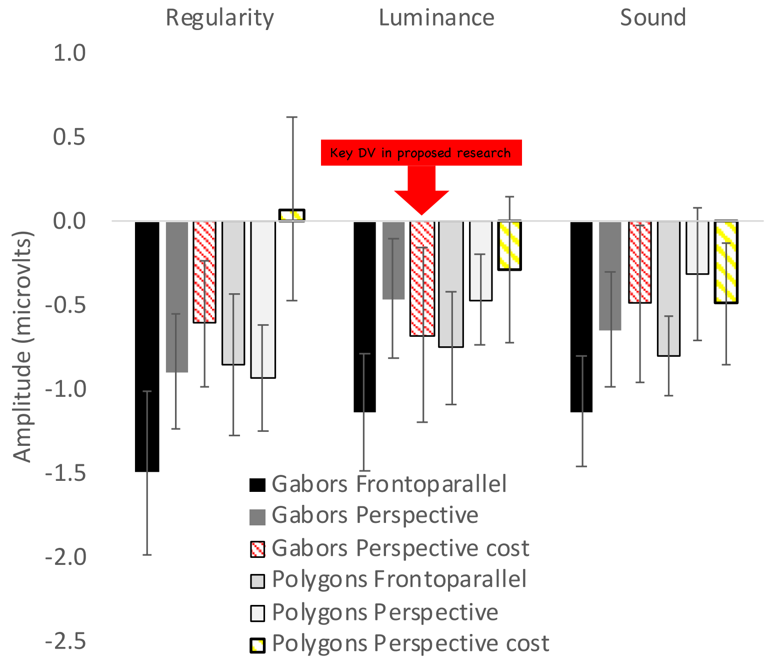
***Response to editors Figure 2.** SPN waves from Karakashevska et al. (forthcoming). The SPNs generated by Gabors during a luminance task (red frame) informs our planned research.*

Crucially there was a perspective cost during the Luminance task with Gabors (red frame panel in Figure 2, t (39) = 2.629, p = 0.012, dz = -0.416). Although the luminance task was easy, with performance > 95% correct, perspective cost was not eliminated by participants spontaneously attending to regularity.

Furthermore, stimulus presentation duration will be reduced to 500 ms in our new experiment (it compared to 1000 ms in Karakashevska et al. forthcoming). This reduces time available for spontaneous task switches and gives a clearer measure of perspective cost.

As predicted by Reviewer 1, polygons significantly reduced perspective cost ($F(1,117) = 4.013$, $p = 0.047$, $\eta p^2 = 0.033$). This can be seen by comparing left and right columns in Figure 2, and by comparing red and yellow bars in Figure 3. We use Gabors rather than Polygons in the current work, because we require a large perspective cost in the baseline block. This is now explained in the manuscript.

In summary, the Karakashevska et al. (forthcoming) project greatly reduces uncertainty about our proposed stimuli and task.



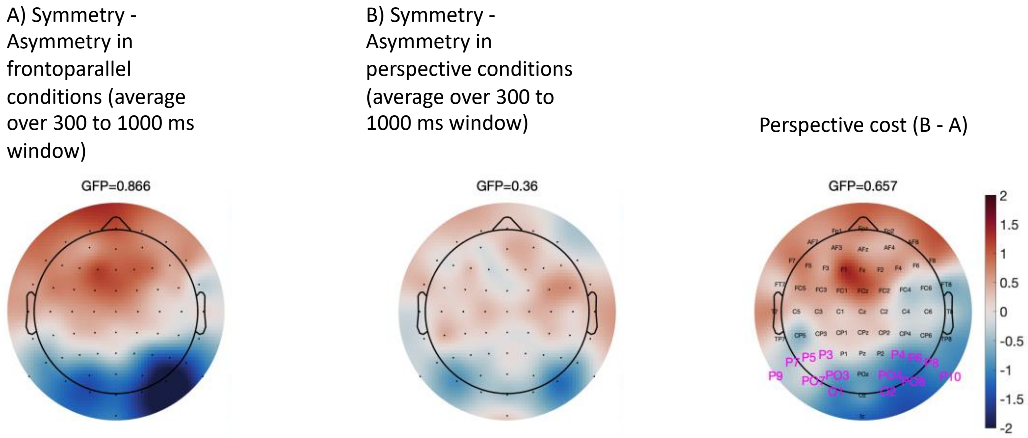
Response to Editors Figure 3*. Mean SPN and perspective cost amplitude from Karakashevska et al. (forthcoming). Perspective cost during a Luminance task with Gabors is the key DV in the proposed research (red arrow). This effect was 0.68 microvolts (t (39) = 2.629, p = 0.012, d = 0.414). Error bars = 95% CI. We expect it to be larger than this in the baseline Block of the proposed experiment because there will be no frame.*

Dr. Rousselet suggested using the Greenhouse-Geisser correction by default, so your power analyses do not need to include the potential testing of Sphericity. They also gave a potential design proposal for boosting your signal-to-noise with the addition of a localizer.

As recommended, we will use the Greenhouse-Geisser correction by default. This is now mentioned in the Study design table.

However, we think it would be safter not to use a localizer. Although we are not experts on localizers, we think it could overweight electrodes and time windows where effects were unusually strong by chance. There are also complications about deciding which effects to optimize for.

The results of Karakashevska et al. (forthcoming) help again here because they support informed decisions about the a priori spatiotemporal cluster. Figure 4 shows the topography of the perspective cost. The planned electrodes are shown in the right panel.



Response to editors Figure 4. *SPN difference between frontoparallel and perspective conditions in the Key condition of Karakashevska et al. (forthcoming) (GFP = global field power, the SD of amplitudes across the 64 electrodes).* Pink electrodes are those we plan to use in the new project.

From our perspective as recommenders, we have a further few items to address in relation to formatting for registered reports:

1. Hypothesis testing:

Your hypotheses are direct and well-motivated (with the addition of a couple of suggestions from the reviewers). In this case, the analyses you plan to run should be equally succinct:
Hypothesis 1: Performing a 2x4 repeated measures ANOVA (regularity x condition) will provide a main effect of regularity to test your hypothesis, as outlined in your Stage 1, but it will also provide a main effect of condition and an interaction between condition and regularity. Please only report the main effect of regularity so you do not need to interpret the main effect of condition and the interaction without an initial hypothesis. These outputs should not be reported in the pre-registered result section.

As recommended, we will only report the main effect of Regularity. We have changed the Study Design table accordingly:

> *"The sample size of 120 was chosen to detect other small effects and is thus adequate to detect the main effect of Regularity in the frontoparallel conditions, which is likely to be large. Based on an unpublished study with the same stimuli and task (Karakashevska et al. forthcoming), we estimate the effect will be 1.14 microvolts, and effect size will be η_p^2 = 0.528. Even assuming true effect size is half this, power approaches 1."*

Hypothesis 2 is directly examined.

For each sub level of hypothesis 3 it seems you just need paired t-tests between each condition and baseline, and between the static and moving frame conditions – therefore no need for the main effect of the ANOVA as suggested in the study plan. Power analyses should be calculated for the smallest effect of interest across your study, which should consider each pairwise comparison, rather than the main effect.

We need to check we are interpreting this comment correctly. We assume that you are NOT recommending we exclude the repeated measures ANOVA from the future results section.

We assume you are instead recommending that we should power our study to find the predicted pairwise differences directly, rather than a main effect of Block.

If we understand correctly, we agree with this recommendation. We have thus changed our power analysis section to focus on pairwise comparisons exclusively (quoted in full below).

Finally, it seems for hypothesis 4 you are performing a one-sided equivalence test (or a non-superiority test; testing the effect is not greater than zero). Please explicitly state this in the Stage 1 text and study design table.

This is true. We are performing a one-sided equivalence test. Please see revised study design table.

Hypothesis 4 is the most intricate part of the planned analysis. We have added a new Figure 8 to the manuscript, because it is much easier to visualize the predictions of hypothesis 4 than explain them:

> *"Hypothesis 4 predicts that meaningful perspective cost will be eliminated in the Moving frame block. This is different from other hypotheses because we are predicting absence of an effect. We will use a one-sided equivalence testing approach. Predicted results are shown in Figure 8A. Perspective cost in the Moving Frame block is predicted to be significantly above -0.35 microvolts (our definition of a small negative ERP effect). The same conclusions would follow from results in B (despite*

significant difference from zero) and C (despite no significant difference from + 0.35). Figure 8D illustrates an alternative possible outcome where perspective cost is not significantly above -0.35 microvolts, and therefore hypothesis 4 would not be supported. In all cases, significance is established with one-tailed, one sample t tests."

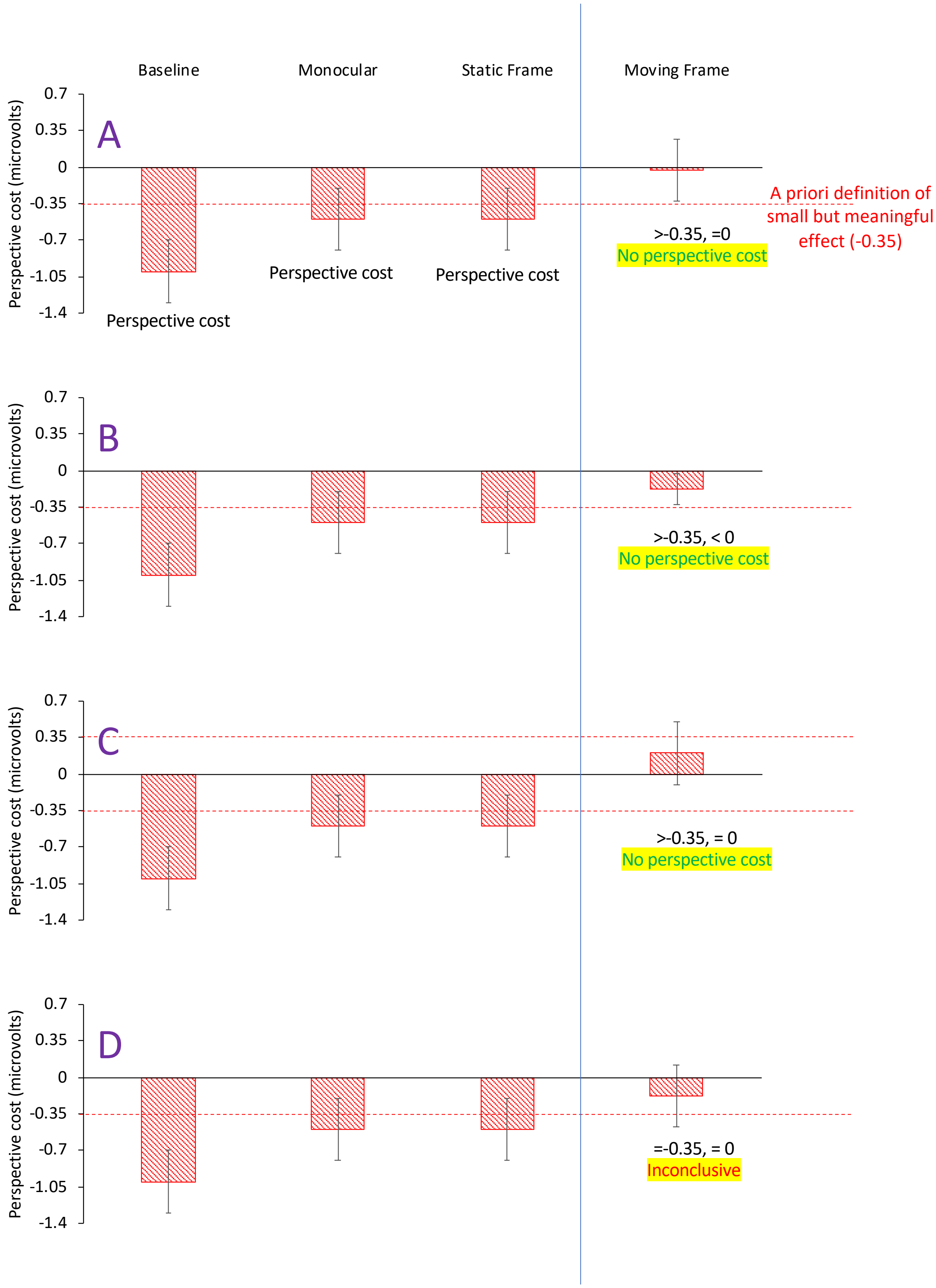


Figure 8. Different possibilities of results to illustrate the one-sided equivalence testing approach. In Baseline, Monocular and Static Frame blocks, we predict a perspective cost. In the Moving Frame block, we predict no perspective cost. The crucial threshold is -0.35 microvolts (our a priori definition of a small negative ERP effect). If confidence intervals do not cross the -0.35 line, we will conclude that perspective cost is absent. This is the case in panels A, B and C. In contrast, the results in panel D are

inconclusive, and do not establish the absence of perspective cost (despite non-significant difference from zero).

We discuss this further with Apthorp. We are happy to hear advice about the validity of our reasoning at this stage, in case the procedure can be improved.

2. Power analyses:

You mention "An SPN modulation effect size 0.34 SD corresponds to around 0.35 microvolts (Makin et al., 2022), which is smaller than nearly all reported SPN modulations". As the 0.35 is a crucial quantity in the power and equivalence testing analyses, the authors should evidence the claim by listing the papers they are referring to, and the range of effect sizes.

This crucial 0.35 microvolt quantity is based on analysis of the SPN catalogue (introduced in Makin et al. 2022), rather than a list of published papers. This catalogue includes all 249 SPN datasets from Liverpool, both published and unpublished (https://osf.io/2sncj/). This has the significant advantage of avoiding file drawer effects, which often distort aggregated estimates of effect size. We have now explained this more explicitly in a new power analysis section, which incorporates a new Figure 4:

"Power analysis

We powered our experiment to find relatively small ERP differences of 0.35 microvolts. This threshold is informed by analysis of the 249 SPNs in the SPN catalogue (https://osf.io/2sncj/), described in Makin et al. (2022). Figure 4 illustrates relevant SPN distributions. Each ridge in Figure 4A represents a distribution of participant SPNs around the mean (the largest, most negative, SPN is at the base). The scatterplot in Figure 4B shows all 249 SPNs as data points, with mean amplitude is on the X axis, and Cohen's d_z (Mean / SD) on the Y axis. The second order polynomial regression line indicates a plausible effect size d for an SPN of a given amplitude. This shows that -0.35 microvolt SPNs are likely to have Cohen's d_z of -0.34. This also applies to within-subject pairwise differences between SPNs.

Furthermore, as explained in Makin et al. (2022), 178 of the 249 SPNs in the catalogue are significant (p <0.05, one sample t test against zero, two-tailed). The smallest significant SPN in the catalogue is -0.342 microvolts. Our threshold of -0.35 microvolts is thus a reasonable empirical definition of a small but meaningful SPN or SPN modulation.

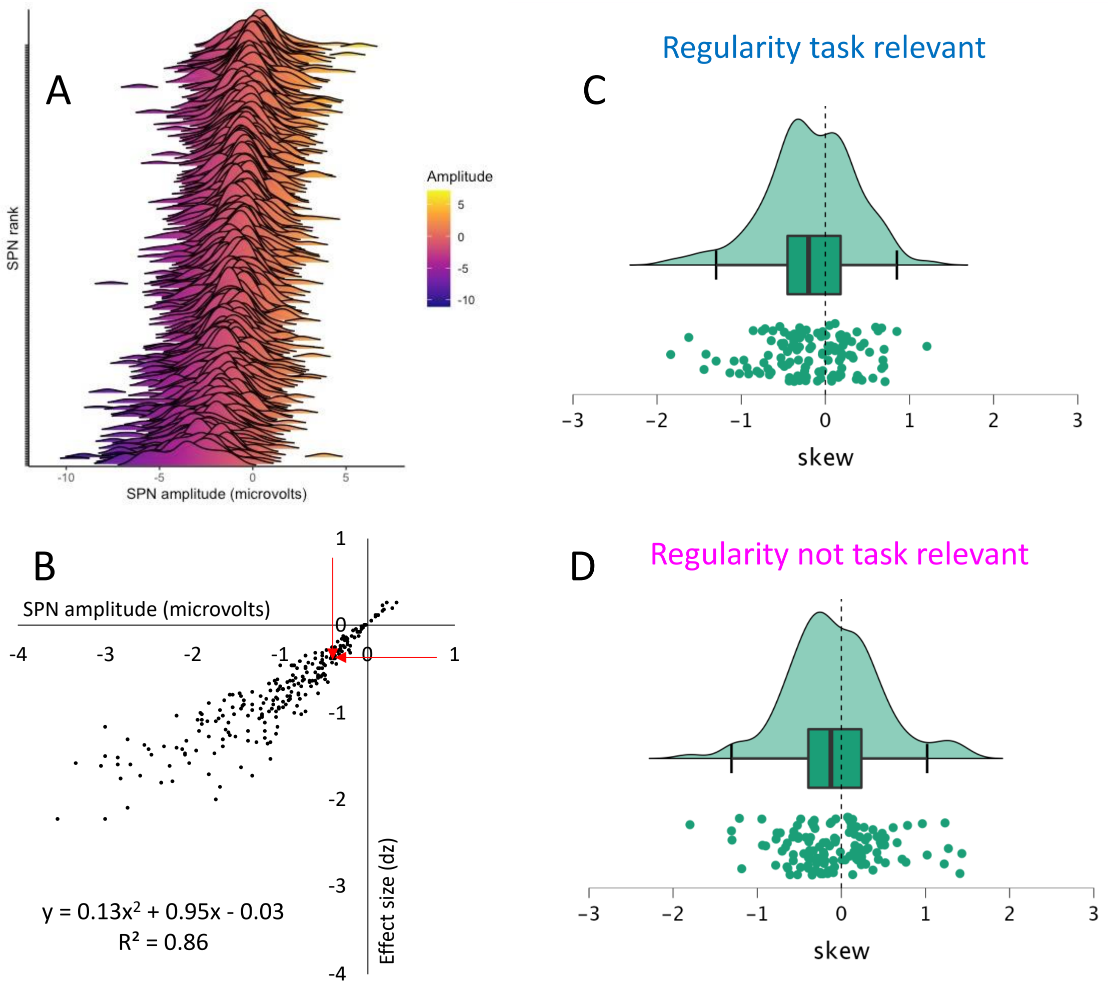


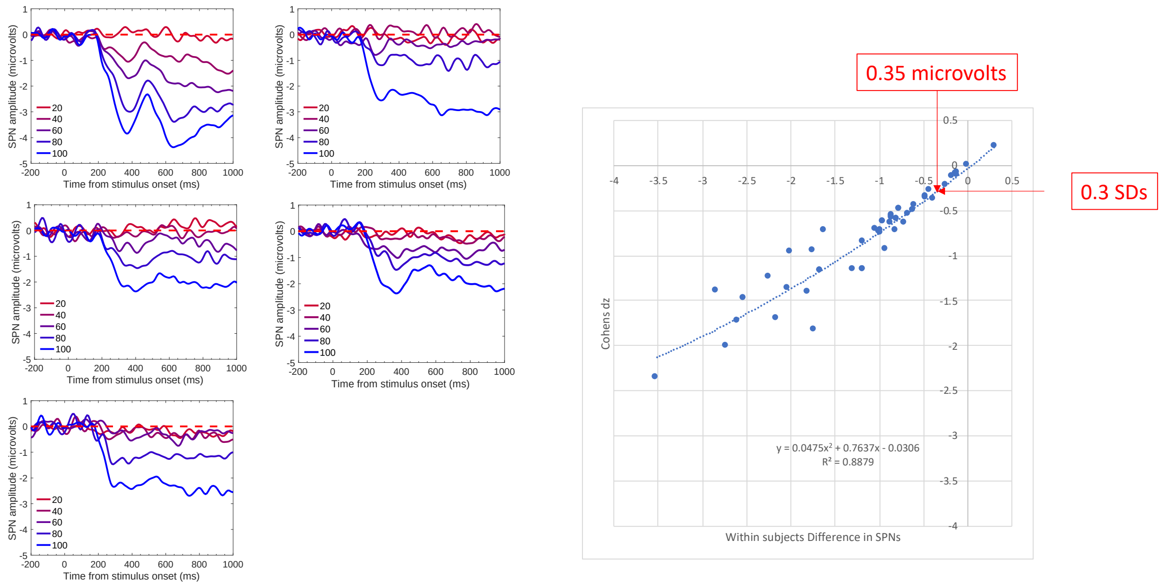
Figure 4. A) Distribution of 249 SPNs from the SPN catalogue (https://osf.io/2sncj/), shown as a ridgeplot. Each ridge is a distribution of individual participant SPNs around the mean. The largest (most negative) SPN is at the base. B) Scatterplot of 249 SPNs. The X axis is SPN amplitude in microvolts. The Y axis is standardized effects size (Cohen's d_z). The second order polynomial line suggests -0.35 microvolt SPNs have a typical effect size d of -0.34 (red arrows). C) Distribution of skewness statistics from experiments where regularity was task relevant. D) Distribution of skewness statistics from experiments where regularity was not task relevant.

Our planned sample of 120 provides 0.95 power for finding one tailed effect (d_z = 0.34, alpha = 0.02). A more conservative approach is to use two-tailed tests, even though we have a directional hypothesis. This reduces power to 0.91. Both criteria are more conservative than the conventional demands (power = 0.8, alpha = 0.05). We also note that the median sample size in previous SPN research was just 24. Our sample of 120 is thus more than twice as large as any published or unpublished within-subjects SPN experiment.

We verified these decisions with a power simulation approach. We computed a power analysis on 10,000 observations from a bivariate normal distribution with a specified correlation of 0.5 between conditions. This confirms we have 90% chance of finding a mean pairwise difference of 0.34 SDs with a sample of 120 (codes for the simulations can be found here: https://osf.io/utq8e).

Hypothesis 4 predicts an absence of perspective cost in the Moving frame block. Here we will use a one-sided equivalence testing approach (illustrated in Figure 8). If true perspective cost is -0.35 microvolts in given a block, we are likely to find that the effect is significantly below zero microvolts with one tailed one sample t test (power = 0.95, Cohen's d_z = 0.34, alpha = 0.02, one-tailed). Conversely, if true perspective cost is zero microvolts in a given block, we are likely to find that the effect is significantly above -0.35 microvolts (power = 0.95, Cohen's d_z = 0.34, alpha = 0.02, one tailed). "

The claim that '*This also applies to within-subject pairwise differences between SPNs*' requires additional supporting evidence (although this is beyond the scope of the manuscript). There are many possible within-subject pairwise comparisons that could inform this analysis, and some are arguably unrepresentative. We use analysis of pairwise differences in Makin et al. (2020). This suggests effect size of a 0.35 microvolt difference would be around 0.3 SDs (slightly lower, but comparable to the 0.34 estimate above) (Figure 5)



***Response to editors Figure 5**. There are 10 pairwise within subjects' differences in each of the SPN panels on the left. Amplitude and effect size of these pairwise differences* *are shown in the scatterplot on the right. Here an 0.35 microvolt SPN is associated with a typical effect size of 0.3.*

We can check particularly relevant pairwise SPN differences in existing data. In Karakashevska et al. (forthcoming), the difference between frontoparallel and perspective SPNs was 0.68 microvolts during a luminance discrimination task with Gabors, and d_z was 0.414. With 120 participants, we are very likely to find a significant effect of perspective (power = 0.98, alpha = 0.02, two-tailed).

In the Colour task of Makin et al., (2015), the pairwise difference between frontoparallel and perspective SPNs was 0.56 microvolts, and d_z was 0.746. With 120 participants, we are near-certain to find an effect of this size (power > 0.999, alpha = 0.02, two-tailed).

In contrast, in Karakashevska et al. (forthcoming), the perspective cost difference between Gabors and polygons was small (d_z = 0.183) and only borderline significant (p = 0.048). If true effect size is 0.183, power is just 0.511 (alpha 0.05, two-tailed). Our new experiment plan assumes that our cues will have more impact on perspective cost than polygons. In this experiment we have also eliminated the frame in our baseline condition, which may have served as a facilitating cue in Karakashevska et al. (forthcoming).

The Study design table has also been updated so it is consistent with the new power analysis section.

Minor comments:

- "exemple" fig 4.
- "bloc" Study Plan Table; Hypothesis 4 row.
- Please match hypotheses in main text with study plan table exactly to remove any confusion.
- You should also provide a "theory that could be shown wrong by the outcomes" for each hypothesis in the final column of the study design table.

We have now done minor things and revised the study design table.

Considering the positive comments from the reviewers, we believe your manuscript has potential for Stage 1 in-principle acceptance, therefore we request revision and resubmission. Please address each reviewers' and the recommenders comments and revise your manuscript accordingly.
Best,
Grace & Zoltan

We are pleased that the editors and reviewers feel that our Stage 1 manuscript has merit.

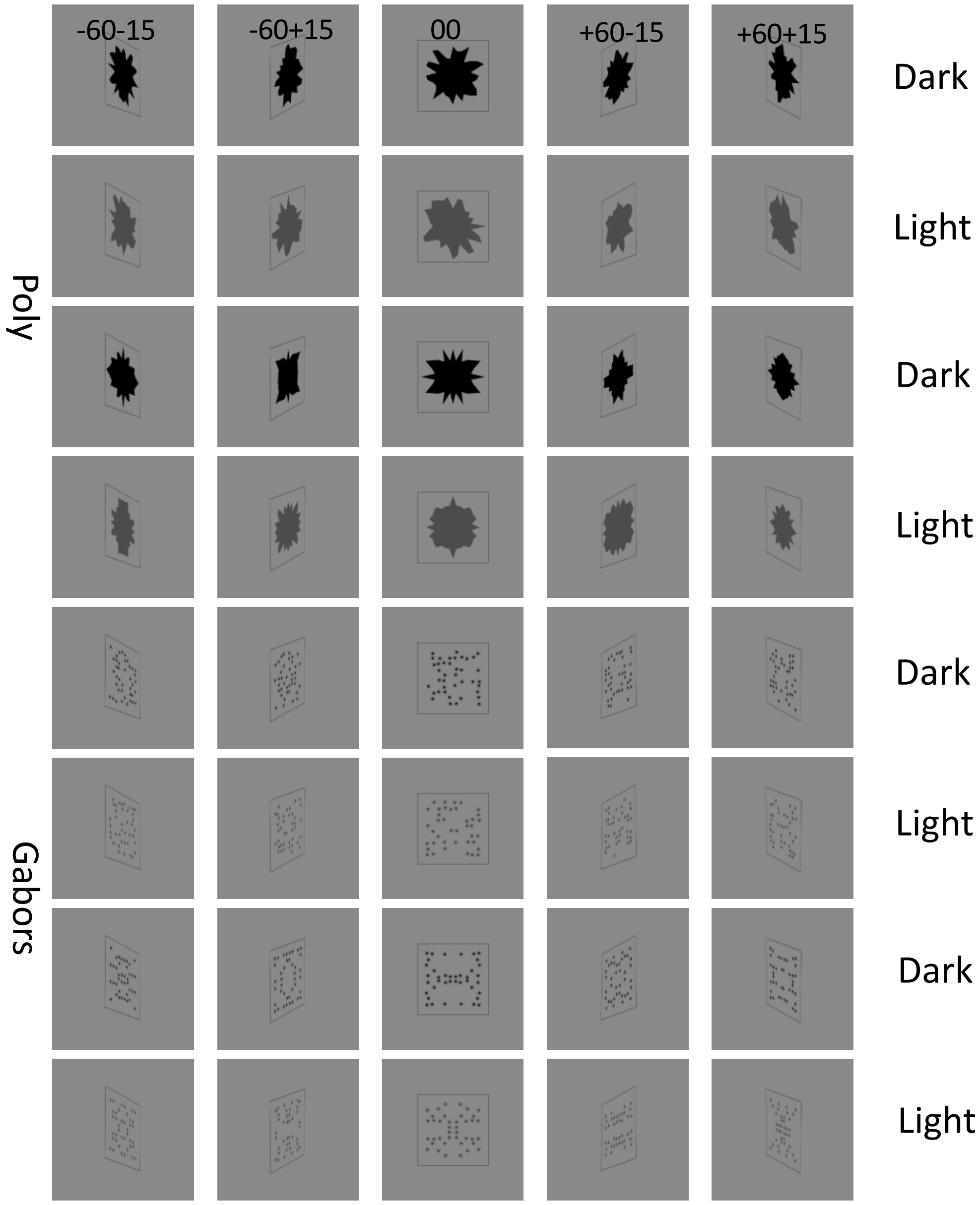
Review by anonymous reviewer 1, 28 Jul 2023 12:33

The manuscript, entitled "Putting things into perspective: Which visual cues facilitate automatic extraretinal symmetry representation?", by Elena Karakashevska, Marco Bertamini, and Alexis D.J. Makin is a registered report of a study measuring neurophysiological responses to mirror-symmetry in the frontoparallel plane and to mirror-symmetry of a planar figure that is slanted from the frontoparallel plane in a 3D scene. The current status of the study is Stage 1, so, the manuscript only includes the introduction, design, and methods sections. The study is well motivated and it is well designed, but it needs a little revision.

We are pleased with this positive review.

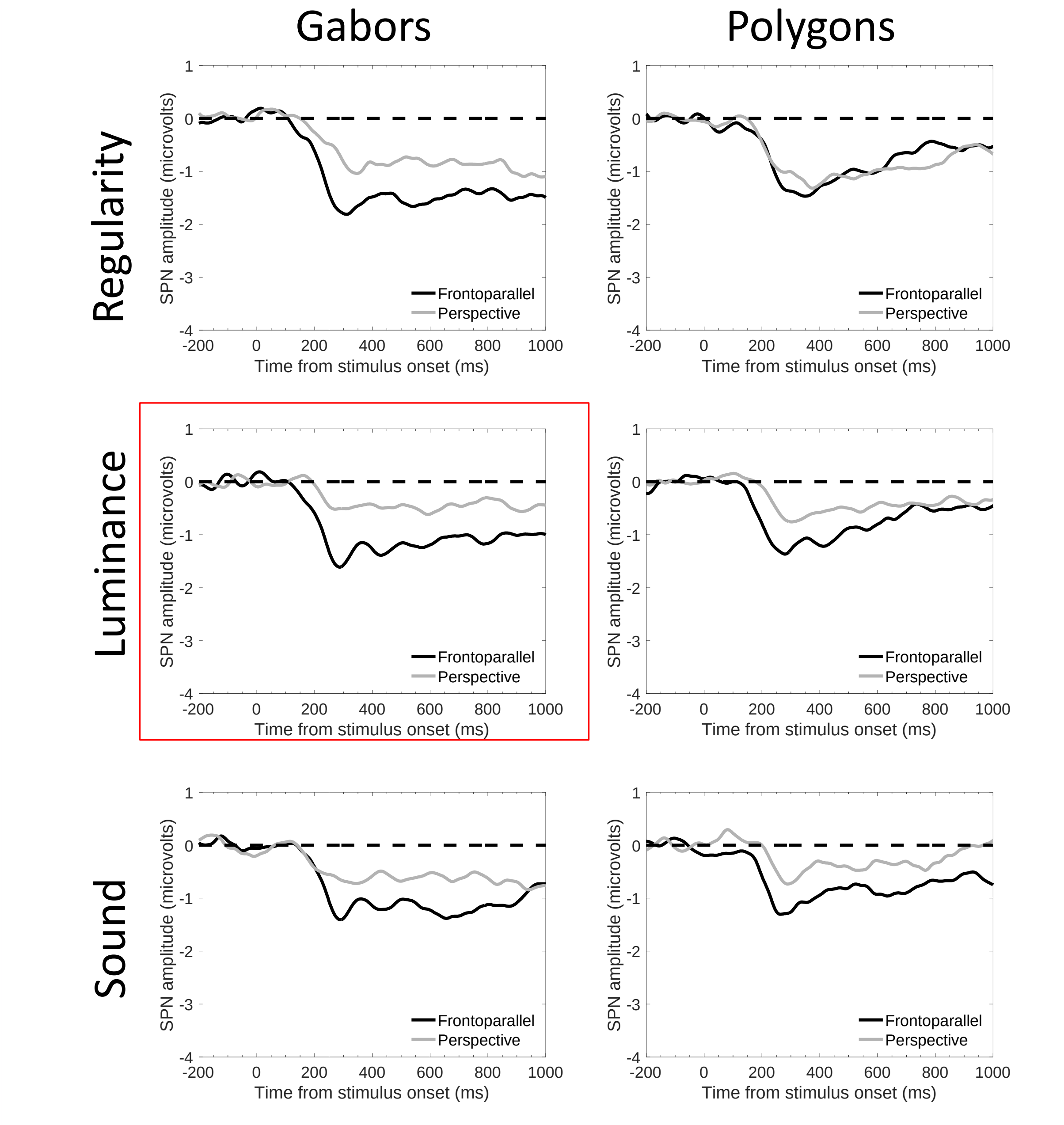
Sawada & Pizlo (2008), and several studies by Wagemans and his colleagues have shown that the mirror-symmetry of a slanted planar figure is hard to detect when only dot-stimuli are used. Reliable detection is possible with contours, so this will be a limitation of this study.

As explained in the response to editors above, we have recently finished analysing a new experiment comparing dot stimuli and polygons (we call this unpublished experiment Karakashevska et al., forthcoming). To be clear, Karakashevska et al. (forthcoming) will be a different publication, and it is not part of the registered report. However, the results validate many decisions about the planned work, so we discuss it in this response to reviewers and mention it in the paper. Stimuli are shown in Figure 1, and results in Figure 2.



Reviewer 1 Figure 1: Stimuli used in Karakashevska et al. (forthcoming). The Gabors are like those we intend to use in the Frame condition of the proposed research.

In *Karakashevska et al. (forthcoming)* there were three tasks. In the first task, participants discriminated regularity. In the second, participants discriminated luminance (the same as the task in our proposed new research). In the third, they judged congruence between luminance and the pitch of a concurrent sound. There were 40 participants in each task. As shown in Figure 2, perspective cost was present in all tasks ($F(1,117) = 18.459$, $p < 0.001$, $\eta p^2 = 0.136$). This was also significant in luminance task with Gabors specifically ($t(39) = 2.629$, $p = 0.012$, $dz = 0.416$).



Response to editors Figure 2. *SPN waves from Karakashevska et al. (forthcoming). The SPNs generated by Gabors during a luminance task (red frame) informs our planned research.*

The decision to use dots rather than polygons is partly informed by the results of Karakashevska et al. (forthcoming). We *want* a large perspective cost in the baseline block, so we can then we can measure substantial reductions in perspective cost in the other blocks. Karakashevska et al. (forthcoming) found that polygons reduce perspective cost (albeit not by much), so polygons are not the right stimuli for the new experiment.

We agree that the results of our experiment may not generalize to polygons, so we have added this point to the introduction:

“Previous behavioural studies have found that perspective cost may be reduced for polygons compared to dot patterns (Sawada & Pizlo, 2008, Wagemans, 1993). In a recent SPN study (Karakashevska et al., forthcoming), we found that polygons slightly perspective cost but do not eliminate it. In the current work, we will use dot patterns, and acknowledge that the results may not generalize to polygons.”

p. 1. Information about pictorial depth cues in the Static frame condition is missing in the manuscript.

We have added this to the abstract on page 1. We have also added more to the method:

“Several pictorial depth cues support 3D interpretation in the perspective conditions. The frame gives the impression that the elements are printed on a flat surface with salient edges. This is subject to salient foreshortening. The top and bottom edges of the frame converge on a vanishing point to the left or right, suggesting they are parallel in the object. The left and right edges also converge on a vanishing point far above or below, again suggesting they are parallel in the object. When there are horizontal and vertical symmetry lines, these converge on the same vanishing points as the frame. The size of the elements, and distance between them, also produces a mild texture gradient. Finally, elements are ovals, consistent with a circle seen in perspective.”

p. 8. Hypothesis 4 is a subset of Hypothesis 3. Hypothesis 3 is a composition of 4 sub-hypotheses and they are based on different factors in the visual stimuli. Hypothesis 3a is about cue conflict. The hypothesis, 3b, is based on an additional pictorial cue. Hypothesis 3c is based on an additional motion cue. Hypothesis 3d is concerned with the difference between these additional pictorial and motion cues. The authors do not explain why these sub-hypotheses were combined to make a single hypothesis.

Hypothesis 3 is indeed a composition of four sub-hypotheses: We predict a main effect of Block, driven by pairwise differences 3a, 3b, 3c and 3d. The project would not be fundamentally different if these were treated as 7 separate hypotheses. However, this it would impair concision, so we would rather nest them.

Hypothesis 4 is NOT a subset of hypothesis 3. Hypothesis 4 predicts that there will be no perspective cost in moving frame block. Hypothesis 4 predicts a null result and requires special statistical considerations. It is easier to treat hypothesis 4 as a sperate category throughout for this reason alone.

p. 5. > participants perform symmetry discrimination tasks (Karakashevska et al., 2022).

There is no Karakashevska et al. (2022) in the References section. Is it Karakashevska et al. (2021)?

Fixed.

p. 6. > stereo cues indicate that that it is flat (Allison & Howard, 2000). …

A plane is still flat even when it is slanted in a 3D scene. Perhaps, the authors want to say "frontoparallel" here.

We have changed this to frontoparallel.

pp. 9-10. > For hypothesis 4 we predict an absence of an effect in the moving frame condition.

What is the effect on?

For clarity, we should say 'absence of perspective cost' rather than 'absence of an effect'. We have changed this section.

> In stage 1, we will run 4 one sample t tests against zero.

What are these t-tests about?

These t tests are cruical for testing hypothesis 4. The analysis associated with hypothesis 4 is the most intricate part of our pre-registered report. We predict some perspective cost in Baseline, Monocular and Static Frame blocks**, but no perspective cost in the Moving Frame block**. We cannot use a non-significant difference from zero to confirm the absence of a perspective cost in the Moving Frame block (the classic mistake of using p > 0.05 to confirm the null).

Instead, we use a one-sided equivalence testing approach for Hypothesis 4. This is tricky to describe in our case, because terms such as 'above' and 'below', and 'larger' and 'smaller' are conceptually reversible, and particularly confusable when talking about differences in the negative range. We have thus added a new Figure 8 which disambiguates all this. We have tried to explain what the t tests are about more clearly with reference to this new Figure 8:

> *"Hypothesis 4 predicts that meaningful perspective cost will be eliminated in the Moving frame block. This is different from other hypotheses because we are predicting absence of an effect. We will use a one-sided equivalence testing approach. Predicted results are shown in Figure 8A. Perspective cost in the Moving Frame block is significantly above -0.35 microvolts (our definition of a small negative ERP effect). The same conclusions would follow from results in B (despite significant difference from zero) and C (despite no significant difference from + 0.35). Figure 8D illustrates an alternative possible outcome where perspective cost is not significantly above -0.35 microvolts, and therefore hypothesis 4 would not be supported. In all cases, significance is established with one-tailed, one sample t tests.*

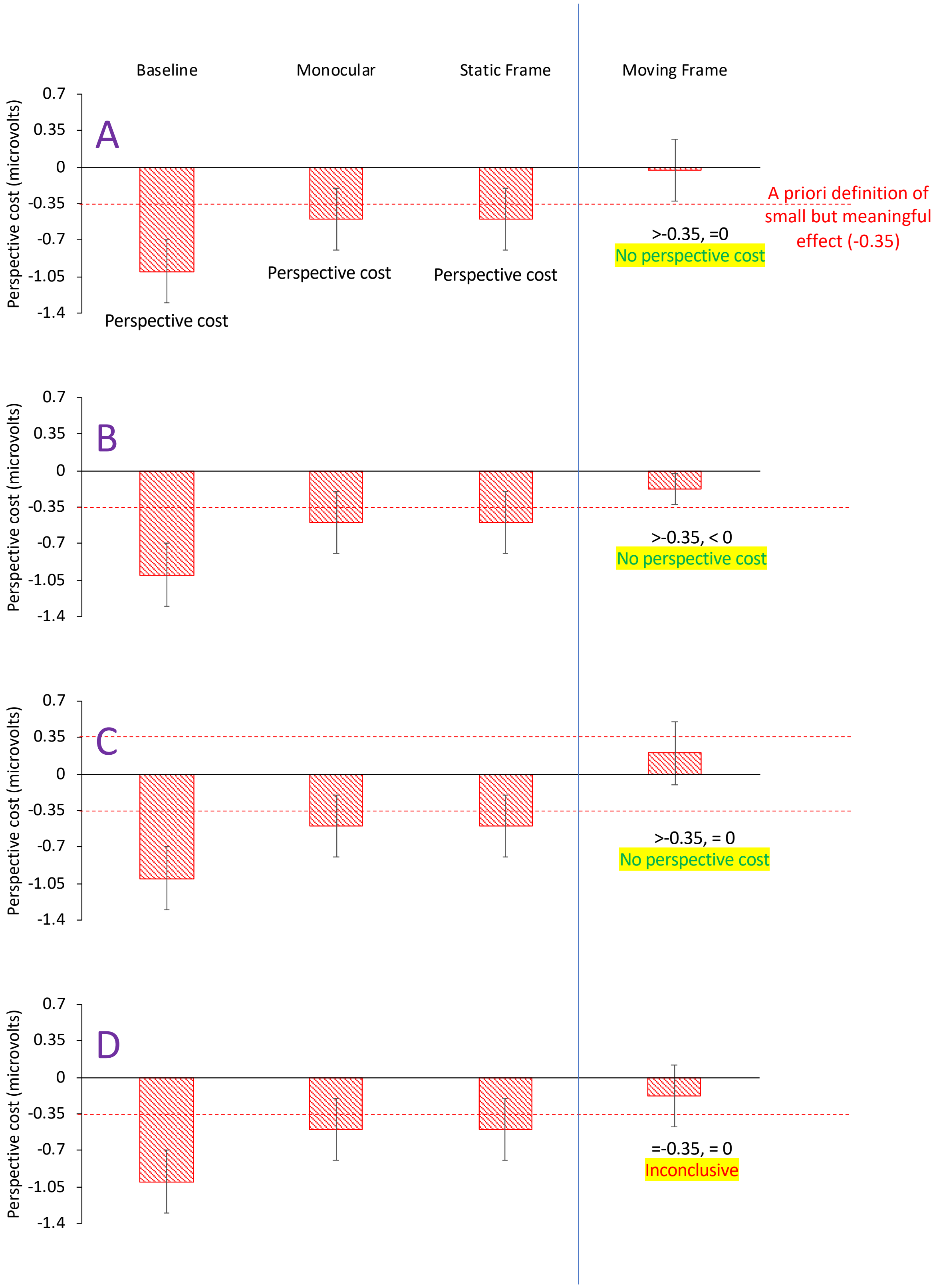


Figure 8. Different possibilities of results to illustrate the one-sided equivalence testing approach. In Baseline, Monocular and Static Frame blocks, we predict a perspective cost. In the Moving Frame block, we predict no perspective cost. The crucial threshold is -0.35 microvolts (our a priori definition of a small negative ERP effect). If confidence intervals do not cross the -0.35 line, we will conclude that perspective cost is absent. This is the case in panels, A, B and C. In contrast, the results in panel D are inconclusive,

and do not establish the absence of perspective cost (despite non-significant difference from zero)."

Are the authors referring to "power"? The authors will conduct the proposed experiment only once.

Yes, we were referring to power. Some of our power analysis sentences were an attempt to demystify the concepts. Rather than using shorthand statements that presume knowledge (e.g. power = 0.9, d = X, N=Y) we often elaborated (e.g. if true effect size is X, and sample size is Y, we would observe a significant result in 90% of experiments like ours). We have tried to improve the section on statistical power. We have also added new analysis of previous research to justify some of our chosen thresholds:

"We powered our experiment to find relatively small ERP differences of 0.35 microvolts. This threshold is informed by analysis of the 249 SPNs in the SPN catalogue (https://osf.io/2sncj/), described in Makin et al. (2022). Figure 4 illustrates relevant SPN distributions. Each ridge in Figure 4A represents a distribution of participant SPNs around the mean (the largest, most negative, SPN is at the base). The scatterplot in Figure 4B shows all 249 SPNs as data points, with mean amplitude is on the X axis, and Cohen's d_z (Mean / SD) on the Y axis. The second order polynomial regression line indicates a plausible effect size d for an SPN of a given amplitude. This shows that -0.35 microvolt SPNs are likely to have Cohen's d_z of -0.34. This also applies to within-subject pairwise differences between SPNs.

Furthermore, as explained in Makin et al. (2022), 178 of the 249 SPNs in the catalogue are significant ($p < 0.05$, one sample t test against zero, two-tailed). The smallest significant SPN in the catalogue is -0.342 microvolts. Our threshold of -0.35 microvolts is thus a reasonable a priori definition of a small but meaningful SPN or SPN modulation.

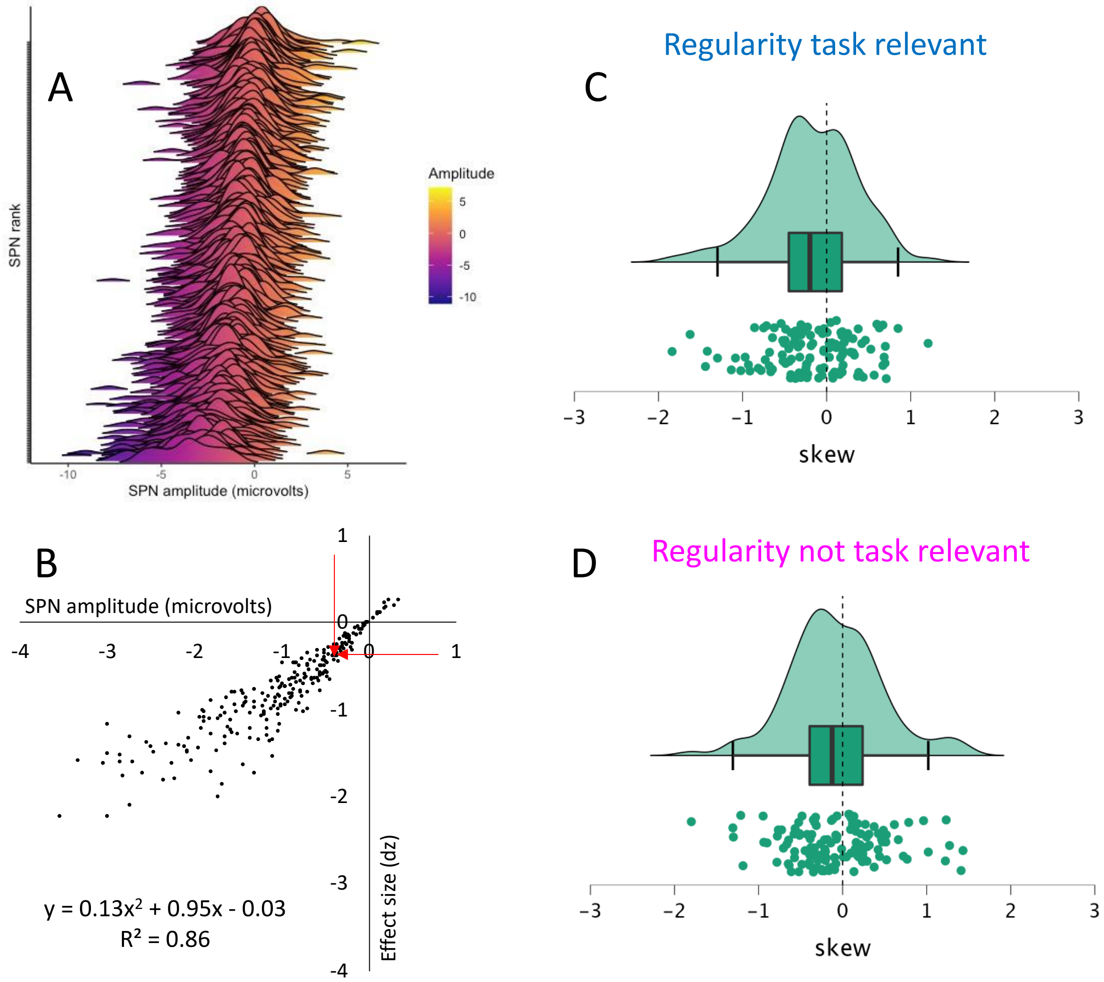


Figure 4. A) Distribution of 249 SPNs from the SPN catalogue (https://osf.io/2sncj/), shown as a ridgeplot. Each ridge is a distribution of individual participant SPNs around the mean. The largest (most negative) SPN is at the base. B) Scatterplot of 249 SPNs. The X axis is SPN amplitude in microvolts. The Y axis is standardized effects size (Cohen's d_z). The second order polynomial line suggests -0.35 microvolt SPNs have a typical effect size d of -0.34 (red arrows). C) Distribution of skewness statistics from experiments where regularity was task relevant. D) Distribution of skewness statistics from experiments where regularity was not task relevant.

Our planned sample of 120 provides 0.95 power for finding one tailed effect (d_z = 0.34, alpha = 0.02). A more conservative approach is to use two-tailed tests, even though we have a directional hypothesis. This reduces power to 0.91. Both criteria are more conservative than the convention demands (power = 0.8, alpha = 0.05). We also note that the median sample size in previous SPN research just 24. Our sample of 120 is thus more than twice as large as any published or unpublished within-subjects SPN experiment.

We verified these decisions with a power simulation approach. We computed a power analysis on 10,000 observations from a bivariate normal distribution with a specified correlation of 0.5 between conditions. This confirms we have 90% chance of *finding a mean pairwise difference of 0.34 SDs with a sample of 120 (codes for the simulations can be found here: https://osf.io/utq8e).*

Hypothesis 4 predicts an absence of perspective cost in the Moving frame block. Here we will use a one-sided equivalence testing approach (illustrated in Figure 8). If true perspective cost is -0.35 microvolts in given a block, we are likely to find that the effect is significantly below zero microvolts with one tailed one sample t test (power = 0.95, Cohen's d_z = 0.34, alpha = 0.02, one-tailed). Conversely, if true perspective cost is zero microvolts in given a block, we are likely to find the effect is significantly above -0.35 microvolts (power = 0.95, Cohen's d_z = 0.34, alpha = 0.02, one tailed).
"

> Stage 2 is required to establish…What is "stage 2"?

This was part of our treatment of hypothesis 4. We have tried to simplify this and avoided talking about stage 1 and 2 (see quote above) and new Figure 8.

p.11. > … with small Gabors (approximate 0.25 dva diameter, Figure 5A).

I do not see any Gabor patterns in Figure 5A. Perhaps, the authors are actually referring to Gaussian patterns.

The elements are a degenerated type of Gabors - but we could call them Gaussian filtered patches. However, we have now used 'dots' throughout the manuscript, and mentioned that the dots were generated using Gaussian filtering in the stimulus section.

> … asymmetrical patterns had accidental rows and columns …

What does "accidental" mean in this sentence?

During stimulus generation, random number generator chooses which cells to be occupied with elements, and 40% of cells are occupied. This sometimes accidentally results in long straight rows and columns of dots, even in the asymmetrical condition. The within-cell positional jitter counters this. The asymmetrical stimuli appear too grid like.

We have changed this sentence:

"Without jittering, asymmetrical patterns of have perfectly straight rows and columns of aligned elements."

> Perspective views were produced by changing the position of the virtual camera.

What is the virtual camera?

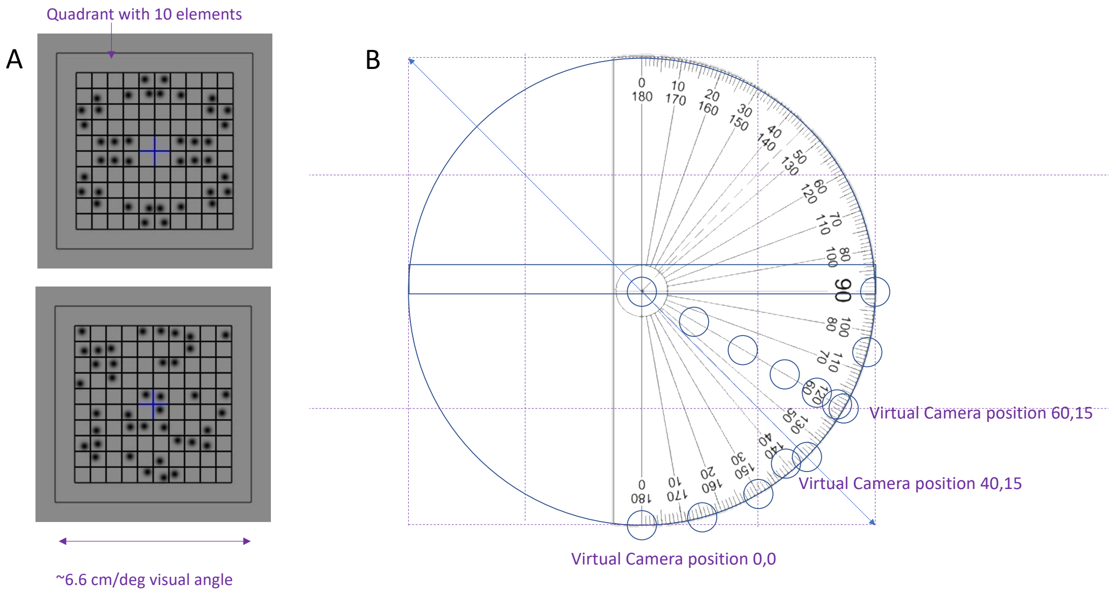
> For frontoparallel trials, the virtual camera was on the equator and vertical meridian.

I understand that the authors are trying to explain the process used to generate their visual stimuli by making use of an analogy between the sphere and the earth, but, the authors need to first explain the orientation of their "earth" relative to a virtual scene. If this is not explained, readers cannot understand how the equator, or the meridian is oriented. At this point, this paragraph does not clarify the process.

We have tried to improve our description of stimulus generation with its virtual camera, and equator and Greenwich meridian metaphors.

> *"Perspective views were produced by changing the position of a virtual camera on the surface of a virtual sphere, looking inwards towards the centre (Figure 6). The 'equator' of the sphere is horizontally aligned with the horizontal midline of the screen. The vertical 'meridian' of the sphere was aligned with the vertical midline of the screen. A stimulus in the middle of the screen has a centre point at the centre point of the virtual sphere."*

The new Figure 6 with larger protractor and expanded legend also helps:



> ***"Figure 6. Stimulus construction diagram. A**) Arrangement of dots in symmetrical and asymmetrical exemplars. The top panel shows construction of a symmetrical exemplar. A quadrant is populated with 10 small Dot elements. This quadrant was reflected across horizontal and vertical axes. The bottom panel shows construction of an asymmetrical exemplar. Here all 4 quadrants are independent. **B)** The protractor diagram represents virtual view angles used to generate perspective stimuli. This is a top-down view of a screen. The centre of the screen is in the centre of a virtual sphere. The protractor represents the equator of this sphere. Purple dots are virtual camera positions used in stimulus rendering. The camera is always focused on the centre of the screen/sphere."*

To validate this stimulus rendering procedure, we compared a photograph of one of our perspective stimuli, taken from the participant's chin rest, to a photograph of a frontoparallel frame as photographed by a real camera clamped in a real physical location. Crucially, these shapes are identical (see red lines in the stimulus below).

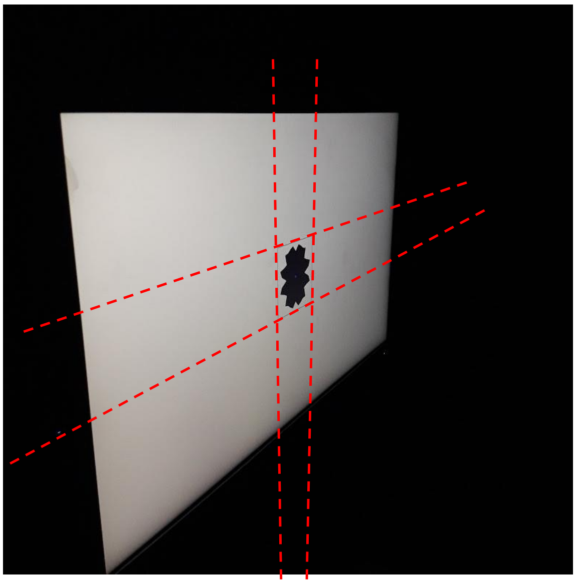


Photo of screen from position -60, 15. Stimulus is objectively square

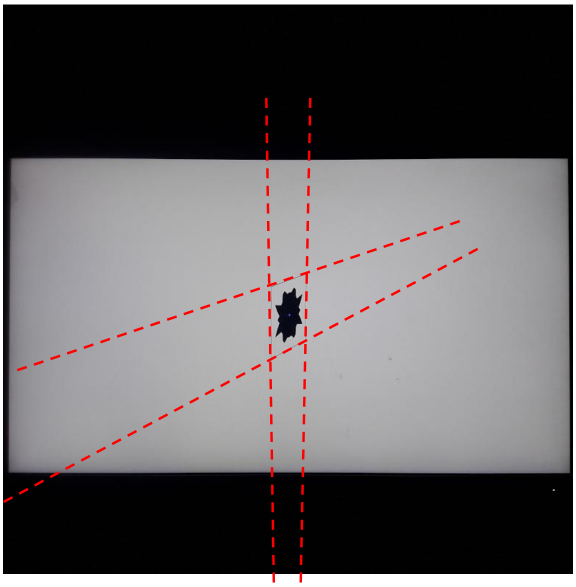


Photo of screen from position 0,0, with rendering of -60,15 stimulus.

Reviewer 1 Figure 3. Photographic validation of stimulus rendering – look at small black frame within the red lines.

p.12 > This study therefore involved projective transformation rather than a superior perspective transformation

The perspective transformation is not superior to the projective transformation. It is a sub-set of the projective transformation. A retinal image of a planar figure in a scene is a perspective transformation of the figure based on the pinhole camera model.

We are happy to hear advice regarding the correct terminology here. We tried to be consistent with a Journal of Vision paper by Sawada and Pizlo (2008) (doi: 10.1167/8.5.14). In their introduction, Sawada and Pizlo (2008) wrote:

> *"It is important to emphasize that the observer's eye must be placed at the center of the perspective projection (C_X, C_Y, C_Z) that was used to compute the perspective images. Only then will the retinal image in the observer's eye be a valid perspective image of the simulated 2D figure slanted in the 3D space."*

We ensured that this was achieved (see Figure 3 above).

To us, the quote from Sawada and Pizlo (2008) suggests that authors think eye-camera matching is superior, and without out this the simulation would not be valid.

The next sentence in Sawada and Pizlo (2008) says:

"Otherwise, the retinal image will be a composition of two perspective projections, which is a ***projective*** *not a perspective transformation of the simulated 3D figure (Coxeter, 1987; Pizlo, Rosenfeld, & Weiss, 1997a, 1997b; Wagemans, Lamote, & van Gool, 1997). "*

[Bold added]

This sentence suggests Sawada and Pizlo (2008) call the less desirable alternative (where eye and camera are not matched), a **projective** transformation.

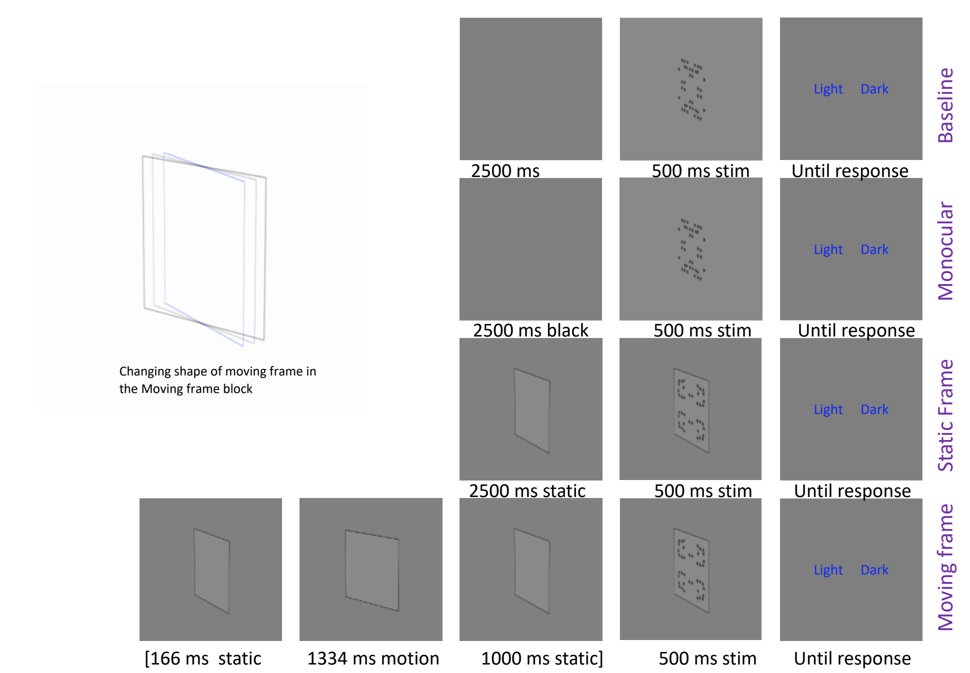
It is possible that we have misunderstood this.

We are happy to take advice here. For now, we have removed the potentially misleading word 'projective transformation' from this paragraph:

"These perspective stimuli have several advantages over those used by Makin et al. (2015). In Makin et al. (2015), the position of the participant's eye and virtual camera were not matched. This is a limitation, because the participants in Makin et al. (2015) had to do two visual transformations, first adopting the position of the virtual camera, and then correcting for perspective distortion (Sawada & Pizlo, 2008). Furthermore, symmetry around the vertical axis was not substantially disrupted by the perspective in Makin et al. (2015). This feature can be seen by inspecting the stimuli in Figure 2. Consequently, if participants focused spatial attention on the axis region, they would have near-perfect retinal symmetry to guide judgements. In the new study, slant and tilt were used to reduce retinal symmetry around the axis. The angles of 60 and 15 degrees were also chosen to follow recommendations in Sawada and Pizlo (2008)."

p.13. Figure 5B is unclear and is also noisy.

We have changed this to Figure 7 and improved:



"Figure 7. Trial structure in the 4 blocks. In the Baseline and Monocular blocks, each trial will begin with a 2500 ms blank screen. This will be followed by a 500ms stimulus presentation and a response screen. In the Static frame and Moving frame blocks, each trial begins with an empty frame. In the Moving frame block, the virtual camera changes location during the first part of the pre-stimulus interval, giving the impression of a frame rocking back and forth around the vertical axis (see inset)."

p.14. > … to classify Gabor element luminance.

Gaussian?

Yes.

p.15. > For the perspective conditions, the virtual camera will move from +/- 60 to +/-40 degrees and back again twice, …

This sentence is unclear. Does the camera oscillate between +60° and -60° first and then oscillate between +40° and -40°? Or, does it oscillate between +60° and +40° or between -60° and -40°?

It only oscillates from + 60 to +40 or -60 to -40. It does not cross zero (see inset in Figure 6).

We have clarified this:

> *"In the Moving frame block, the first part of the pre-stimulus interval will show a moving frame. For the perspective conditions, the virtual camera will move round 20 degrees, from its most extreme starting position at +60 (or -60) degrees to position nearer the meridian at +40 (or -40) degrees and back again. Vertical position never changes, remaining at + or - 15 degrees (see inset in Figure 7). The camera shift happens twice, giving the perceptual impression that the frame rocks back and forth (diagrammatised with purple dots on the protractor in Figure 6)."*

Review by Guillaume Rousselet, 27 Jul 2023 14:36

This RR is relatively clear and presents well conceived hypotheses and design. There are enough experimental conditions to allow a clear interpretation of the results. A lot of code and the stimuli are already shared online, which is brilliant. I'm not an expert in symmetry perception so my comments are mostly about the structure of the article and the analyses. I look forward to seeing the results!

We are pleased with this positive review.

##Abstract
There is a very abrupt transition from a general topic to a specific goal about computational resources. At least one extra sentence is needed to explain the problem. Third sentence also introduces a new topic abruptly, with the explanation only found in the next sentences: reverse order for better flow.

What is the meaning of "selectively reduced"? Would "reduce" suffice? Otherwise explain.

The key sentence "However, this perspective cost might be reduced when additional visual cues support extraretinal representation." is insufficient to understand the problem. How do the different blocks help answer the question?

"The task [...] they will". Rephrase to focus on task or participants, but not both in the same sentence.

"we will conclude that automatic extra-retinal symmetry representation occurs during luminance discrimination" -- luminance discrimination appears for the first time at the end of the abstract and should be explained earlier.

We have made these changes, and the abstract now reads:

> *“**Introduction:** Objects often project different images when viewed from different locations. Our visual system can correct for perspective distortion and identify objects from different viewpoints that change the retinal image. This study will determine the conditions under which the visual system spends computational resources to construct view-invariant, extraretinal representations. We focus on extraretinal representation of planar symmetry. Given a symmetrical pattern on a plane, symmetry in the retinal image is degraded by perspective. Visual symmetry activates the extrastriate visual cortex and generates an Event Related Potential (ERP) called Sustained Posterior Negativity (SPN), and previous studies have found that the SPN is reduced for perspective symmetry during secondary tasks. However, this perspective cost might be reduced when additional visual cues support extraretinal representation.*
> ***Method:** 120 participants will view symmetrical and asymmetrical stimuli presented in a frontoparallel or perspective view. The task will not involve symmetry, the task will be to discriminate luminance. Participants will complete four blocks. In the Baseline block there will be no cues supporting 3D interpretation. In the Monocular viewing*

block, participants will view the same stimuli with one eye. In the Static frame block, additional pictorial depth cues will be available - the elements appear printed on flat square surface with salient edges. In the Moving frame block, motion will enhance 3D interpretation before stimulus onset.
Expected results: *We will compute perspective cost as the difference between the frontoparallel SPN and the perspective SPN. We predict perspective cost will be reduced in all three blocks compared to baseline. If our predictions are confirmed, we will conclude that automatic extra-retinal symmetry representation occurs during luminance discrimination when sufficient visual cues are available.*

##Introduction
"Reflectional symmetry is everywhere in the universe." Even in black holes? Do you need this sentence?

We don't know! We have removed this.

"Both symmetrical and asymmetrical stimuli generate event related potentials (ERPs) at posterior electrodes." -- could you be more specific? Any brief visual presentation triggers ERPs.

"the symmetry wave" -- needs more explanation. Do you mean a sequence of ERPs following the presentation of a symmetric stimulus?

"This difference is called the 'Sustained Posterior Negativity' (SPN)" -- difference between what and what?

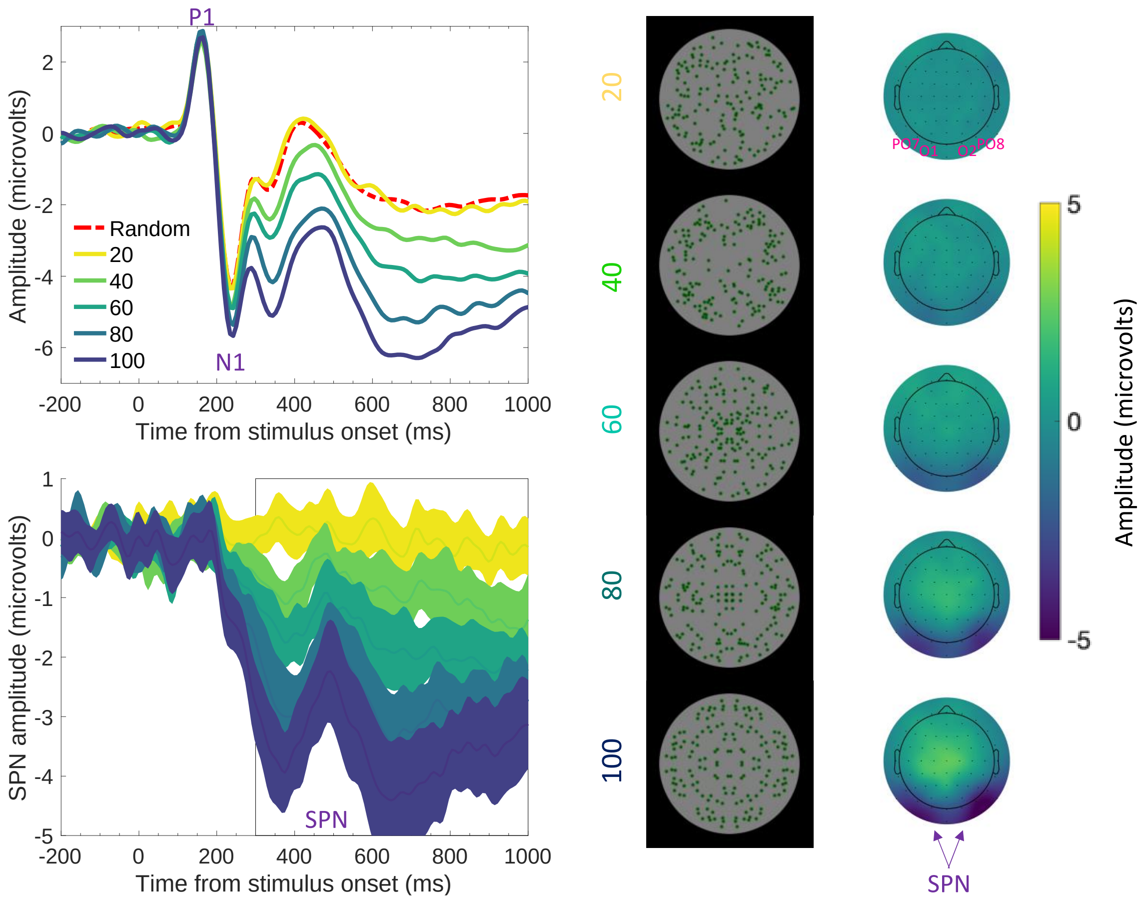
We have clarified this:

> *"Symmetrical and asymmetrical stimuli, like all visual stimuli, generate event related potentials (ERPs) at posterior electrodes. These ERP waves begin with the P1 and N1 components of the visual evoked potential (VEP). After the VEP, there is a persistent difference between the ERP generated by symmetry and the ERP generated by asymmetry. This late difference is called the 'Sustained Posterior Negativity' (SPN)."*

Figure 1: turn this image into grey levels and you will see the issue. The colourmap is not linear and colourblind friendly. Viridis and related colourmaps are linear and colourblind friendly. There are also better divergent colourmaps you could use in the topographic plots.

We have now used the Viridis colour map for Figure 1. We obtained the matlab RGB coordinates here.

https://uk.mathworks.com/matlabcentral/fileexchange/51986-perceptually-uniform-colormaps



"Figure 1. Results of Makin et al. (2020). The grand-average ERPs are shown in the upper left panel and difference waves (reflection-random) are shown in the lower left panel. A large SPN is a difference wave that falls a long way below zero. Topographic difference maps are shown on the right, aligned with the representative stimuli. The difference maps depict a head from above, and the SPN appears as dark blue at the back. Red labels indicate electrodes used for ERP waves [PO7, O1, O2 and PO8]. Note that SPN amplitude increases (that is, becomes more negative) with the proportion of symmetry in the image. In this experiment, the SPN increased from ~0 to –3.5 microvolts as symmetry increased from 20% to 100%. Figure adapted from Makin et al. (2022)."

Figure 2 B & C: in grey levels the contrast between conditions is poor. For accessibility, make one condition black, the other one grey.

Done.

"Stereo defined symmetry is another form of extraretinal symmetry..." -- unclear how that topic relates to the two studies mentioned in the previous sentence; be explicit.

We have tried to explain better:

"Stereo defined symmetry is another form of extraretinal symmetry: Here symmetry is not present in the retinal image in each eye, and the cyclopean contours can only be seen when images are fused in the visual cortex. However, SPN amplitude is equal for stereo and contrast defined symmetry when participants perform symmetry discrimination tasks (Karakashevska et al., 2021)."

"indicate that that it is flat" -- that x 2

Fixed

"perspective cost would be zero (as in Figure 2B)" -- the two conditions actually differ in that figure. Maybe phrase as close to zero or practically equivalent, which would prepare readers for the equivalence test that you present later on.

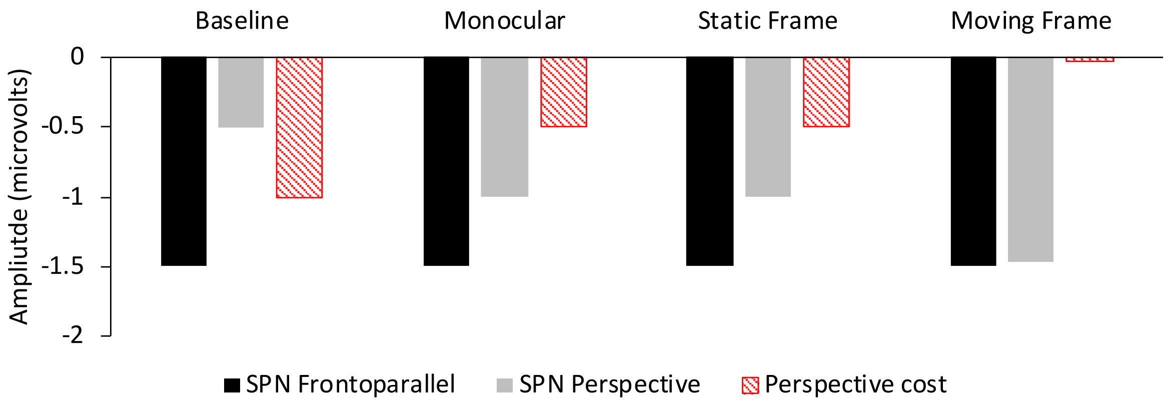
Done

"We predict that perspective cost will highest" -- be missing

Done

Figure 3: contrast could be increased in that figure too. A suggestion is to have one condition in black, one in grey and one in white with a black surround.

We have made Figure 3 consistent with Figure 2. Frontoparallel is black and Perspective is grey. Stripy red is perspective cost (this works when figure is in greyscale).



*"**Figure 3. Predicted results.** The SPN is the difference between symmetrical and asymmetrical conditions (negative bars represent a large SPN). The SPN may be larger (more negative) in frontoparallel (black) than perspective (grey) conditions. This*

difference is called perspective cost (red). We predict that perspective cost will be highest in the baseline block (left) and reduced in the other three blocks. Perspective cost may approach zero in the moving frame block (right). The predicted amplitude of these effects is more speculative than the rank order.

”

"by covering the one eye" -- delete the

Done

##Method
"A sample 120 participants" -- of missing

"All participants will have normal or corrected to normal vision and no history of neurological conditions." -- based on a self report?

Yes, this will be based on a self-report. We will however obtain participants' eye dominance in the lab. We will use the distance hole-in-the-card test (Rice et al., 2008; doi: 10.1016/j.jaapos.2008.01.017). We have now included this in the manuscript as:

“The preferred sighting eye will be determined using the hole-in-the card test. A red cross (3 × 3 cm) will be presented approximately 5 m in front of the participant. The participant will hold a card (13 × 20 cm) with both hands, at arm's length and move the card until the cross is visible through a hole in the centre of the card (1.5 cm in diameter), with both eyes open. The examiner will then cover the right eye of the participant and ask if the cross has remained in his/her line of view. The eye that allows the participant to maintain the view of the cross while the other eye is closed will be documented as the preferred sighting eye.”

"We thus powered our experiment" -- I know it is a shortcut but to be accurate power cannot be the property of an experiment; it is only define in the long run for a line of research. Also, power is not defined in a vacuum, it must be for a specific test.

"N=120 provides 92% chance of finding a significant..." This statement is inaccurate as there is no probability associated with one experiment. Also, you first need to explain what will be measured, how that quantity is distributed, and what test(s) will be used before you can address power. So a bit of reorganisation of that section is needed.

We agree that some of these sentences about power may have been ambiguous. We have now reworded the power analysis section, and made other improvements in response to the editor and other reviewers:

“Power analysis

We powered our experiment to find relatively small ERP differences of 0.35 microvolts. This threshold is informed by analysis of the 249 SPNs in the SPN catalogue (https://osf.io/2sncj/), described in Makin et al. (2022). Figure 4 illustrates relevant SPN distributions. Each ridge in Figure 4A represents a distribution of participant SPNs around the mean (the largest, most negative, SPN is at the base). The scatterplot in

Figure 4B shows all 249 SPNs as data points, with mean amplitude is on the X axis, and Cohen's d_z (Mean / SD) on the Y axis. The second order polynomial regression line indicates a plausible effect size d for an SPN of a given amplitude. This shows that -0.35 microvolt SPNs are likely to have Cohen's d_z of -0.34. This also applies to within-subject pairwise differences between SPNs.

Furthermore, as explained in Makin et al. (2022), 178 of the 249 SPNs in the catalogue are significant ($p < 0.05$, one sample t test against zero, two-tailed). The smallest significant SPN in the catalogue is -0.342 microvolts. Our threshold of -0.35 microvolts is thus a reasonable a priori definition of a small but meaningful SPN or SPN modulation.

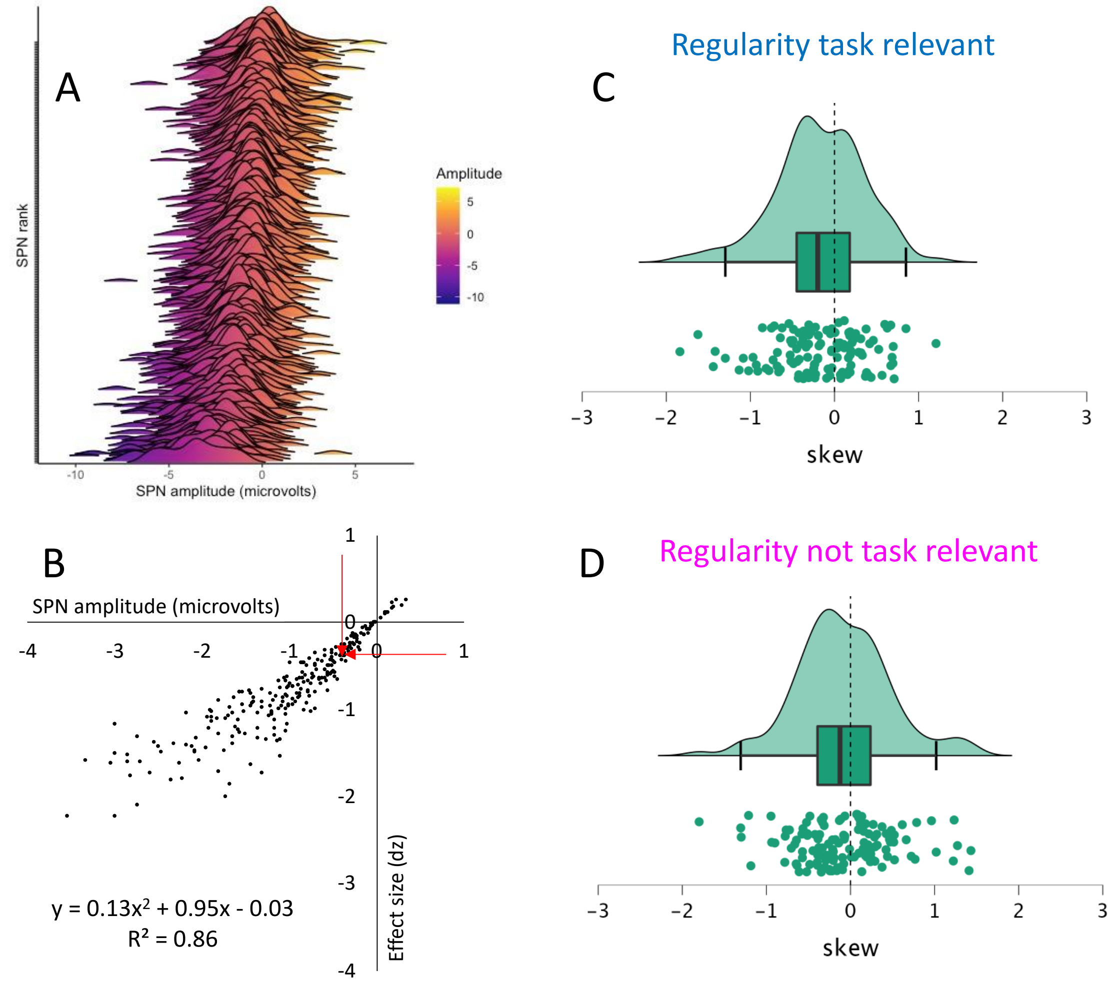


Figure 4. A) Distribution of 249 SPNs from the SPN catalogue (https://osf.io/2sncj/), shown as a ridgeplot. Each ridge is a distribution of individual participant SPNs around the mean. The largest (most negative) SPN is at the base. B) Scatterplot of 249 SPNs. The X axis is SPN amplitude in microvolts. The Y axis is standardized effects size (Cohen's d_z). The second order polynomial line suggests -0.35 microvolt SPNs have a typical effect size d of -0.34 (red arrows). C) Distribution of skewness statistics from experiments where regularity was task relevant. D) Distribution of skewness statistics from experiments where regularity was not task relevant.

Our planned sample of 120 provides 0.95 power for finding one tailed effect (d_z = 0.34, alpha = 0.02). A more conservative approach is to use two-tailed tests, even though we have a directional hypothesis. This reduces power to 0.91. Both criteria are more conservative than the convention demands (power = 0.8, alpha = 0.05). We also note that the median sample size in previous SPN research just 24. Our sample of 120 is thus more than twice as large as any published or unpublished within-subjects SPN experiment.

We verified these decisions with a power simulation approach. We computed a power analysis on 10,000 observations from a bivariate normal distribution with a specified correlation of 0.5 between conditions. This confirms we have 90% chance of finding a mean pairwise difference of 0.34 SDs with a sample of 120 (codes for the simulations can be found here: https://osf.io/utq8e).

Hypothesis 4 predicts an absence of perspective cost in the Moving frame block. Here we will use a one-sided equivalence testing approach (illustrated in Figure 8). If true perspective cost is -0.35 microvolts in given a block, we are likely to find that the effect is significantly below zero microvolts with one tailed one sample t test (power = 0.95, Cohen's d_z = 0.34, alpha = 0.02, one-tailed). Conversely, if true perspective cost is zero microvolts in given a block, we are likely to find the effect is significantly above -0.35 microvolts (power = 0.95, Cohen's d_z = 0.34, alpha = 0.02, one tailed)."

"strict thresholds" -- stricter? I agree that aiming for 90% power in the long run is a big improvement over the traditional 80% (another tradition with zero foundation). I wonder why people think it is ok to miss an effect in one out of 5 experiments.

Yes, low statistical power is a chronic problem is cognitive neuroscience, where researchers rarely even reach the conventional 80% level (Button et al. 2013). In our recent paper, we acknowledged the need to increase sample size in SPN research (Makin et al., 2022).

The power simulation is a great addition, but you need to justify the use of a normal population. Given that you have access to a large database of SPN, it would be very informative to illustrate a large n distribution. I see this is mentioned later on: "Analysis of the whole SPN catalogue suggests that individual participant SPNs are usually normally distributed around the grand average." So bring it all together, before the power section, ideally with an illustration.

We have added another new Figure before the power section. The ridgeplot in Figure 4A is an illustration of the 249 approximately normal distributions. See quote above. We have also added more on normality assumptions. This is elaborated in response to a later comment.

"a specified correlation of 0.5" -- correlation between what and what?

Between the participant perspective cost in Block A and Block B, etc. In within subject designs, the variables are usually correlated. 0.5 is quite a conservative, low correlation.

"LCD monitor" -- add specs.

We have changed this:

“Participants will be positioned 57 cm from a 51 X 29 cm (1920 X 1080 pixel) HP E233 LED backlit monitor, with 60Hz refresh rate.”

"The luminance of the light and dark elements" --report the values.

Done

"It will thus marginally darker" -- be missing

"on the Baseline and Monocular blocks" -- on -> in.

"This feature can be seen be inspecting" -- be -> by

Done

Figure 5 B: the disk with angles is presented under trial structure and is too small. I suggest to split this figure in two or to make it larger.

We have split this figure in two as recommended.

##EEG
matlab -> Matlab + add version number.

Done

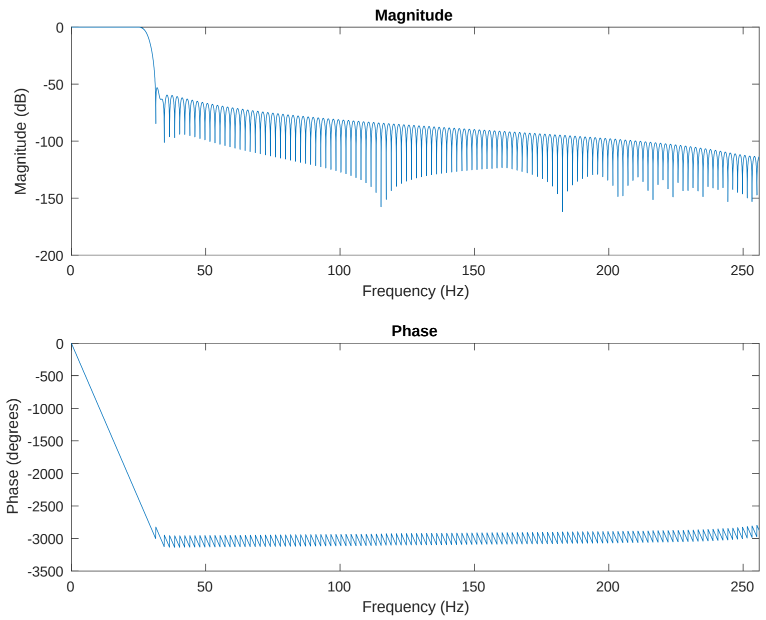
Add details about the filter characteristics. LP filter at 25 Hz seems a bit drastic, but that depends on the slope/order of the filter.

The filter we intend to use is implemented by this line:

```
EEG = pop_eegfiltnew(EEG,[],25,[],0,[],1);
```

In Matlab the command help pop_eegfiltnew gives more information

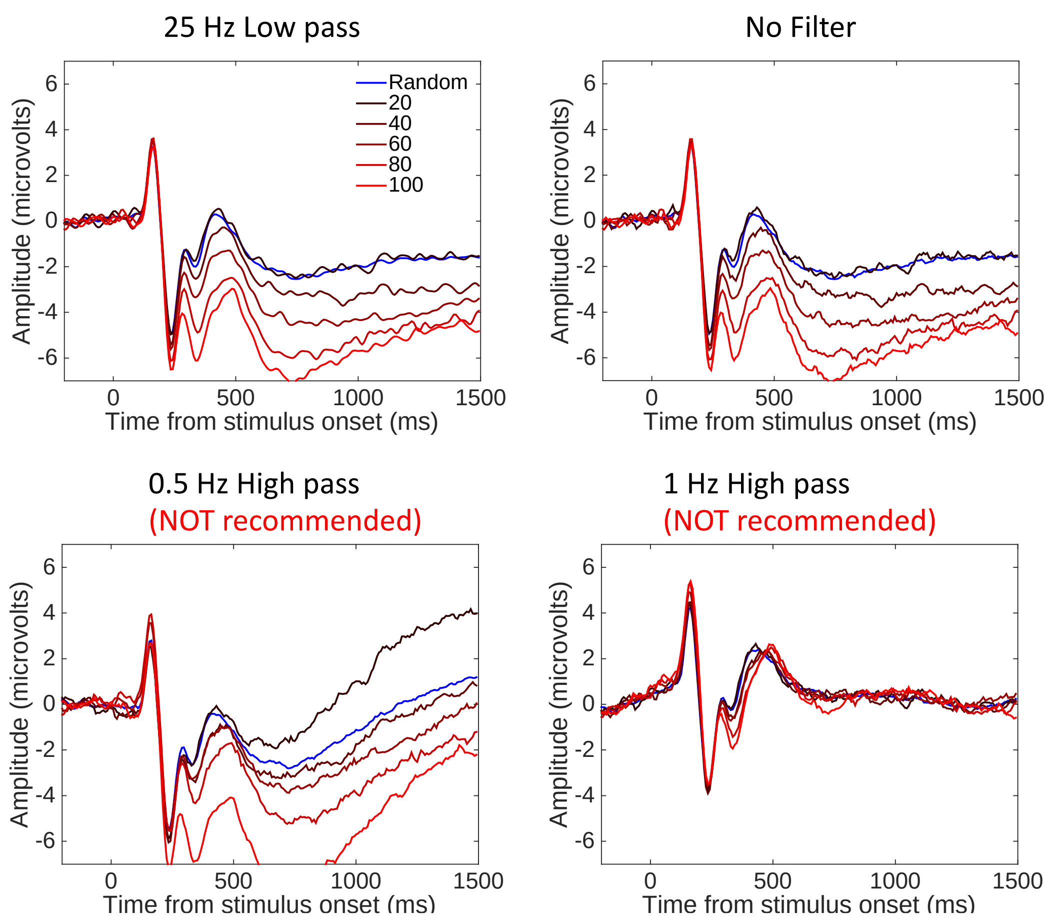
Visualization of the frequency and phase response looks like this:



Review 2 Figure 1. Filter properties.

We are not experts on EEG filter parameters, so any advice at this stage would be welcome.

In previous research we investigated the consequences of different filter parameters on SPN waves and found it makes very little difference. The Figure 2 below (from supplementary materials of Makin et al. (2020)) is instructive. 25 Hz low pass is very similar to no filter (Figure 2). In contrast, high pass filters are very disruptive for the SPN.



Reviewer 2 Figure 2. Consequences of high and low pass filtering. While 25Hz low pass filtering only removes minor roughness, high pass filtering dramatically alters the shape of ERPs. Although it removes most high amplitude artefacts and may avoid the need for ICA cleaning in some cases, it should never be used in SPN research.

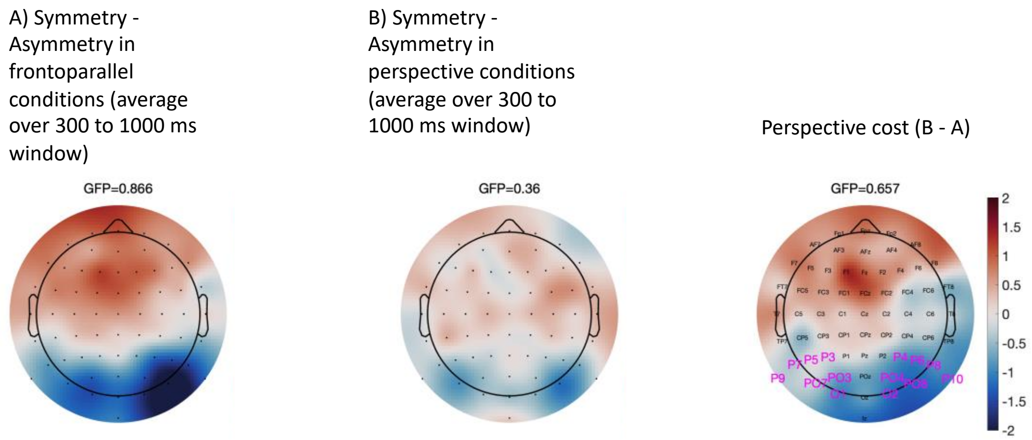
As a result of considering this point, we have made our epochs are slightly longer (-500 to 700 ms) to provide a buffer around the interval of interest.

"These channels will then be replaced with spherical interpolation." -- what is the point of interpolation? It doesn't add any information to the analyses. Do you plan to include an electrode factor in the ANOVAs? If not then interpolation is a waste of time.

We believe channel interpolation is a useful step in EEG pre-processing. Even though we will only analyse a cluster of posterior electrodes, all channels contribute to the average reference. Second, bad channels can dominate ICA, and make it difficult to removing blink artifacts. Third, our standard criteria are to remove trials where absolute amplitude exceeds 100 microvolts at any of the 64 scalp electrodes. Without channel interpolation, we are sometimes left with cases where one bad channel causes every trial to be removed.

Cluster of electrodes: do you plan to average the ERPs across these electrodes? If the SPN varies a lot across electrodes, it would be more powerful to use a localiser to identify the best electrode(s) in each participant. Otherwise averaging over so many electrodes will necessarily lower the effect.

Yes, we do plan to average across the electrode cluster and time window. We believe this is probably the best approach, especially considering that a previous study with the same stimuli can be used to make an a priori decision about the spatiotemporal parameters (Figure 3).



Reviewer 2 Figure 3. SPN difference between frontoparallel and perspective conditions (GFP = global field power, the SD of amplitudes across the 64 electrodes). Pink electrodes are those we plan to use in the new analysis.

Averaging over electrodes will necessarily lower the effect, but which effect are we talking about? It could be that the predicted pairwise difference 3a conditions is maximal at different electrodes to the pairwise difference 3b etc. All this complexity is eliminated if we average over an electrode cluster.

We could optimize for the overall SPN. That is, for each participant, we chose the electrodes where the overall symmetry-asymmetry difference is maximal. But then this might mask the more interesting SPN differences between frontoparallel and perspective conditions.

While we are not experts on localizer approaches, we believe there is some risk that they would overweight ‘lucky’ electrodes, where noise is going in the direction of the predicted effect.

##Analysis plan

"We will check for violations of the normality assumption using the Kolmogorov-Smirnov test." This is a bad idea for reasons explained here:

https://garstats.wordpress.com/2022/09/30/normtest/

The KS test is extremely poor at detecting deviation from normality. More importantly, you mentioned above that you have good reasons to believe that the SPN population is normally distributed, so that makes a test of this assumption superfluous. Also, non-parametric tests are not equivalent to the parametric ones: they do not test the same hypotheses.

We have improved our treatment of normality we will not base decisions on the KS test. We have now added a new section called 'Normality assumptions', which refers to Figure 4C and D:

> *"Normality assumptions*
>
> *The ridgeplot in Figure 4A shows that individual participant SPNs are often normally distributed around the mean. Indeed, only 8-9% of the 249 SPNs violate the assumption of normality according to Shapiro-Wilk and Kolmogorov-Smirnov tests ($p < 0.05$). However, the validity of these tests is questionable. We therefore analysed the distribution of 249 skewness statistics associated with the 249 SPNs. There is a small but significant mean negative skew when regularity is task relevant (Figure 4C, mean = -0.174 microvolts, SD = 0.529, t (124) -3.665, $p < 0.001$). However, this is less pronounced when regularity is not task relevant (Figure 4D, mean – 0.081 microvolts, SD = 0.561 t (123) = -1.609, $p = 0.110$). We expect the SPN data will be normally distributed in new project, where regularity is not task relevant."*

"If we violate the assumption of Sphericity..." -- just use the GG correction by default. Otherwise, if your tests are conditional on other tests, you need to redo the power analyses to include the extra decisional step.

We will use the GG correction by default, even if the assumption of sphericity is not violated according to Mauchly's test. This is an interesting point about power analysis. However, based on recommendations of the editor, we justify our power analysis by considering pairwise differences, rather than the main effect of Block.

##Results
For stage 2, consider how you will represent the results in sufficient detail. I would expect an article free of bar graphs and with clear representation of individual results. Here are some

guidelines:
https://onlinelibrary.wiley.com/doi/full/10.1111/ejn.13400

We will certainly consult these guidelines. While bar graphs are useful for illustrating predicted results (as in Figure 3), we will add violin plots with individual data points for real results.

Review by Benoit Cottereau, 27 Jul 2023 03:58

This submission focusses on the mechanisms underlying symmetry perception in humans using scalp EEG recordings. The authors propose to characterize how different visual cues support extraretinal symmetry representation. To this aim, they will measure the sustained posterior negativity (SPN) in different viewing conditions and compute a perspective cost which corresponds to the difference between frontoparallel and perspective SPNs. They hypothesize that this perspective cost will be diminished (compared to a baseline condition) under monocular viewing (i.e., when the cue conflict between perspective and binocular disparity is removed) and when additional perspective cues (either static or moving frames) are added. In my opinion, this submission could pass stage 1 of the review process as the research question is scientifically valid and the proposed hypotheses are plausible. In addition, the experiments sound feasible and the methodology is well developed and can thus be replicated. I provide more detailed comments below.

1A. Scientific validity of the research question

The proposed scientific question stems from numerous psychophysical and neuroimaging (EEG and also fMRI) studies which suggested that extraretinal symmetry representations are not constructed automatically when attention is focused on another task (e.g., when participants are instructed to report non-symmetrical features of the stimuli). In event-related EEG recordings ('ERPs'), these mechanisms can be reflected by a perspective cost corresponding to the difference between frontoparallel and perspective SPNs. Here, the authors wish to question whether this perspective cost is removed under more naturalistic viewing conditions, when sufficient cues are available to support 3D interpretation. This is a valid and scientifically justifiable question which was not addressed in previous works. This question is answerable through quantitative research and does not suffer from ethical issues.

1B. Logic, rationale, and plausibility of the proposed hypotheses

As preliminary hypotheses, the authors propose that their experimental protocol will permit to measure sustained posterior negativities (SPNs) at posterior electrodes between 300 and 600 ms post stimulus onset. They also propose that in the baseline condition (symmetric/asymmetric stimuli without additional cues), these SPNs will be significantly larger for frontoparallel than for perspective stimuli, leading to measurable perspective costs. These preliminary hypotheses are supported by the results of previous studies performed by the authors using a similar experimental protocol (Makin 2022; 2015).

The main hypothesis of the study is that perspective costs will be reduced (as compared to a baseline) under more realistic viewing conditions, i.e. under monocular viewing (when the conflict between perspective and binocular disparity is removed) and when additional perspective cues (either static or moving frames) are added. The authors also hypothesize that the perspective cost will be lower with moving frames than with static frames (because the cues supporting the extraretinal representation of symmetry are weaker in this latter case). In addition to this main hypothesis, the authors also propose that the perspective cost will approximate zero with moving frames (although in this case, a conflict with binocular disparity is still present). All the hypotheses are precisely stated and follow directly from the research question.

1C. Soundness and feasibility of the methodology and analysis pipeline (including statistical power analysis or alternative sampling plans where applicable)

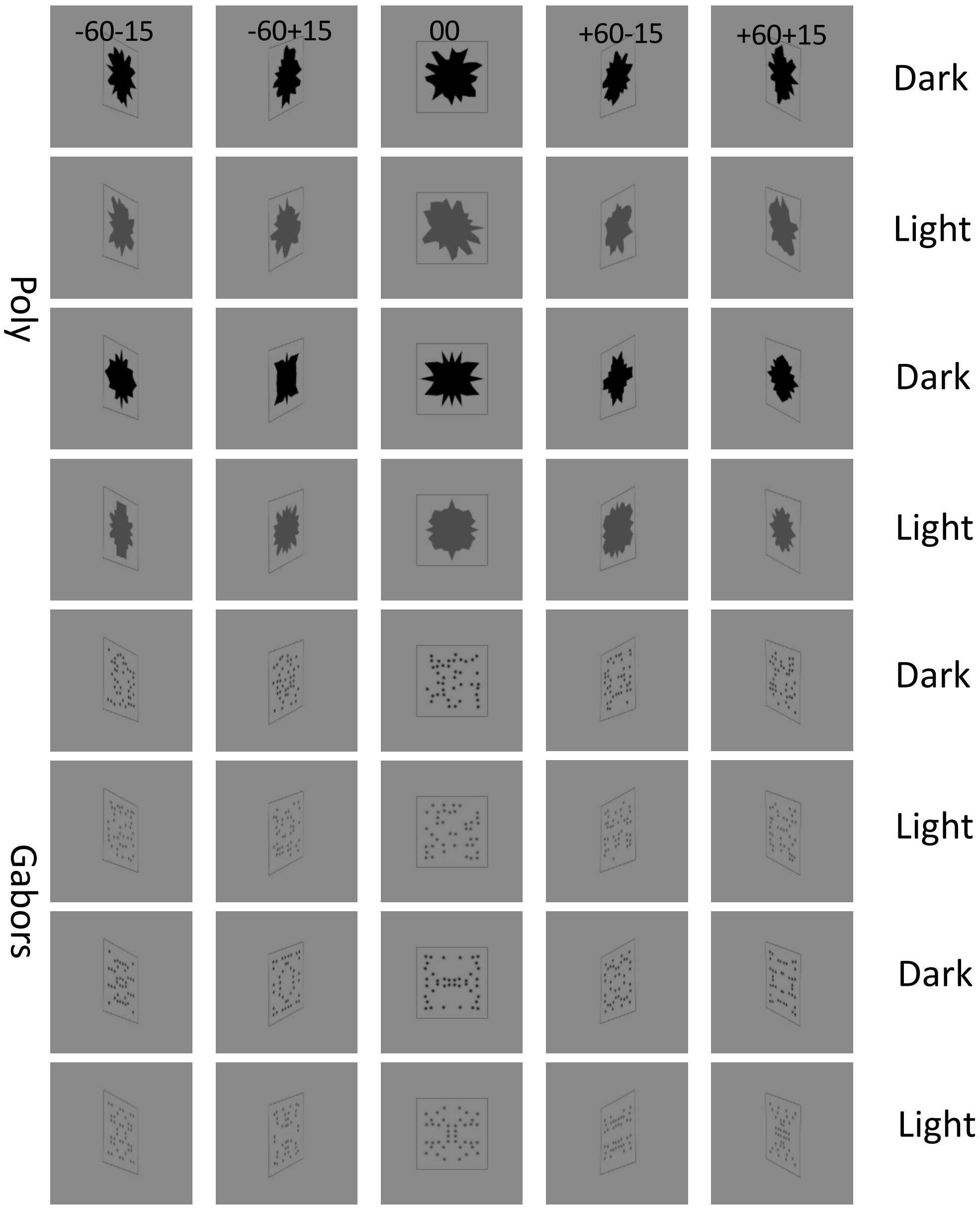
The methodology developed in this submission is based on previous EEG studies from the same group which demonstrated the feasibility and soundness of the proposed experiment. The authors already measured significant SPNs (see e.g., Makin, 2022) and perspective costs (Makin, 2015) using a similar analysis pipeline. This pipeline is based on a classical pre-processing of the EEG data (data are re-referenced to scalp average, filtered and segmented into epochs, an independent component analysis is used to remove artefacts such as eye blinks, for each condition, event-related potentials are computed on a pre-defined posterior electrode cluster and between 300 and 600 ms after stimulus onset). The authors provide a convincing statistical justification of their sample size (n = 120) which should permit to properly test the different proposed hypotheses. It has to be noted that this sample size is much larger than in previous EEG experiments which measured SPNs.

I nonetheless noted a few points that might deserve some attention:

- The authors chose the luminance values used in their task based on a pilot experiment and in order to get more than 90% of correct responses. This value is rather high (chance level is 50% in this case). Is it possible that for some participants, the task is very easy and they can also attend to the symmetry of the stimuli, thereby reducing the perspective cost, even in the baseline condition?

This is a possibility. Fortunately, we can learn from another recent unpublished experiment, with the same luminance task. We call this Karakashevska et al. (forthcoming).

We did not discuss Karakashevska et al. (forthcoming) in the Stage 1 because it has not yet been peer reviewed, and we had not finalized our analysis at that time of submission. However, this analysis is now finished, and it is relevant to addressing the task difficultly point. The stimuli used by Karakashevska et al. (forthcoming) are shown in Figure 1 below:



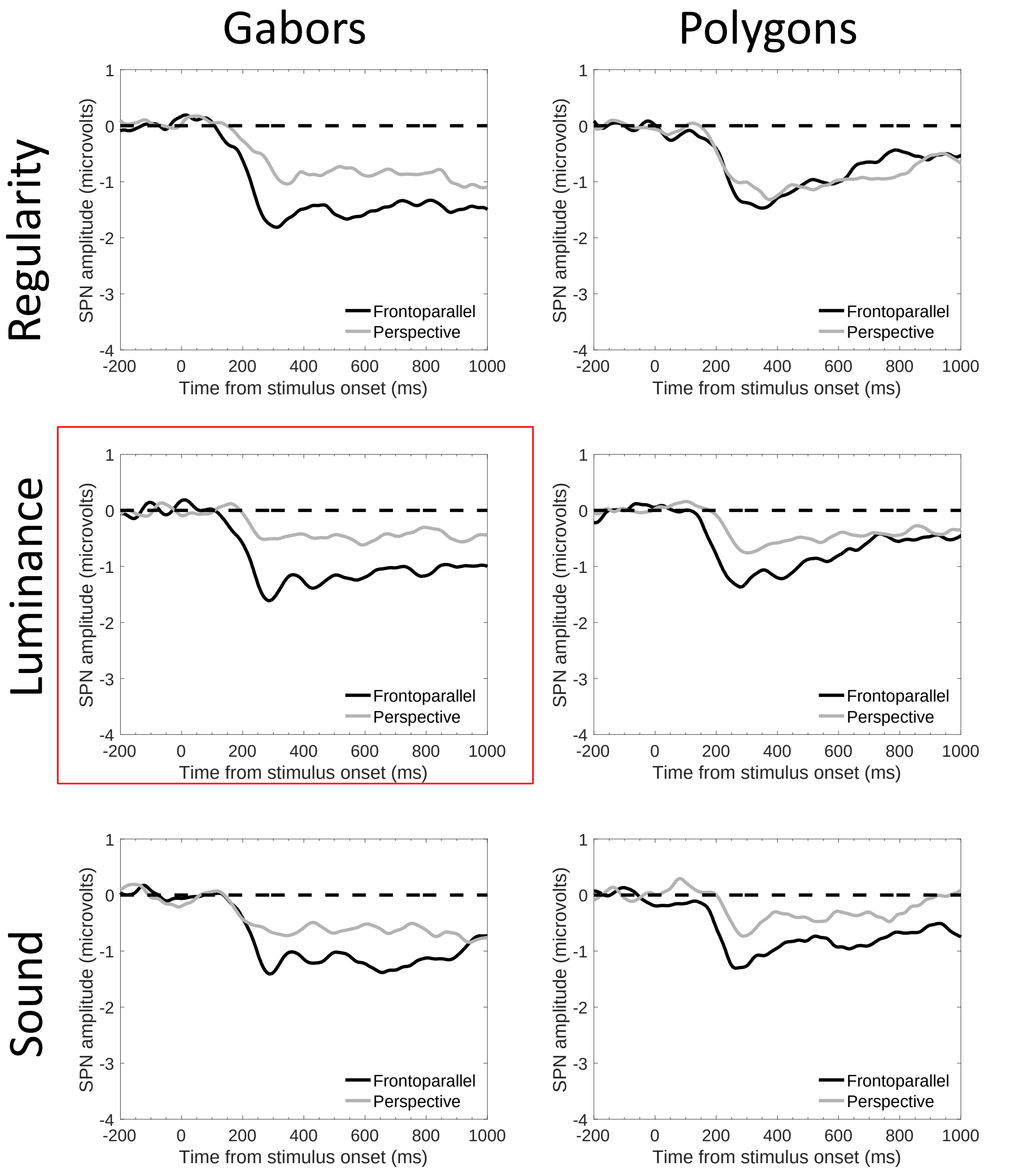
Reviewer 3 Figure 1: Stimuli used in Karakashevska et al. (forthcoming). The Gabors are very like those we intend to use in proposed research.

In *Karakashevska et al. (forthcoming)* there were three tasks. In the first task, participants discriminated regularity. In the second, participants discriminated luminance (the same as the task in our proposed new research) in the third, they judge congruence between luminance and the pitch of a concurrent sound. There were 40 participants in each task, so 120 in total.

Results of Karakashevska et al. (forthcoming) are show in figure 2. Crucially there was a perspective cost during the Luminance task with Gabors (see second row of Figure 2). (t (39)

= 2.629, p = 0.012, dz = -0.416). This suggests such tasks are not too easy, and perspective cost is not eliminated for that reason.

Furthermore, stimulus presentation duration will be reduced to 500 ms in our new experiment (as compared to 1000 ms in Karakashevska et al. forthcoming). This reduces time available for spontaneous task switches and gives a purer measure of perspective cost.



Reviewer 3 Figure 2. SPN waves from Karakashevska et al. (forthcoming). The SPNs generated by Gabors during a luminance task (red frame) informs our planned research.

The Karakashevska et al. (forthcoming) results greatly reduce uncertainty about the value of the stimuli and task in the planned experiment.

- The authors are intending to replace any participants whose performance is below 80% in any block. This criterion may be a little harsh. Is there any justification for it?

In the Karakashevska et al. (forthcoming) luminance task, all 40 participants were above 90% correct (except one who was at 50%, and obviously responding randomly). It is very unlikely that someone with normal vision would fall below 80% due to perceptual limits.

- In the moving frame condition, the frame motion will stop 1000 ms before stimulus onset (in this time interval, only a static frame will be displayed). Did the authors wonder whether the neural responses triggered by the motion will still be observable after 1000 ms (i.e., after stimulus onset). These possible late ERPs will be removed in the computations of the SPN (with respect to baseline) and of the perspective costs but should be taken into account if the authors are intending to show raw ERPs.

This is possible, although motion evoked potentials happen within 250 ms (e.g., Heinrich 2007, 0.1007/s10633-006-9043-8). The motion processing in the moving frame condition will most likely be finished by the time the dots appear.

The analysis pipeline includes -200 to 0 ms baseline before dot pattern onset, will subtract away any amplitude differences resulting from the earlier motion.

We have now clarified this:

> *"It is likely that motion evoked potentials generated by the moving frame will be complete long before the baseline period."*

1D. Clarity and degree of methodological details, replicability of the proposed study procedures and analysis pipeline

The manuscript gives sufficient methodological details for the experimental protocol to be reproduced. The authors notably provide weblinks (osf) to their codes for power analysis simulations, generating the stimuli and running the experiments and processing the EEG data in Matlab. The proposed methodology is clearly structured and easy to follow.

1E. Consideration of outcome-neutral conditions (e.g. absence of floor or ceiling effects; positive controls; other quality checks) for ensuring that the obtained results are able to test the stated hypotheses

The main hypothesis of the study will be tested by comparing the perspective costs in different viewing conditions with a baseline. The baseline condition was already used in previous study and is likely to lead to a significant perspective cost. The chances that the obtained measurements permit to test the main hypothesis are thus very high.

We are pleased with the overall positive evaluation of our paper. To be clear, the Static Frame condition is most like the previous study (Karakashevska et al., forthcoming). We predict that perspective cost will be even higher than this in the Baseline condition.

Review by Deborah Apthorp, 28 Jul 2023 04:45

This report aims to investigate human processing of visual symmetry. The authors are experienced in this field. Previous research shows there is an automatic pre-attentive response to visual symmetry which is seen in EEG recordings, the Sustained Posterior Negativity (SPN). This effect is diminished when displays are shifted from the frontoparallel plane, but only when participants are not actively detecting symmetry in the displays. The authors suggest that this may be because visual displays used in experiments provide few (or no) perspective cues to indicate that the display is shifted from the frontoparallel plane, and that this "perspective cost" may be reduced if participants are given more perspective cues. The Introduction explains the gap in the literature and sets out the research question well.

The hypotheses are logically set out, especially the first three. For Hypothesis 4, I think we need to see more justification of why motion cues are expected to eliminate perspective cost. This seems to be a very central hypothesis for the study, and yet it is only justified in a very small paragraph on Page 6 (last paragraph before Study Aims and Hypotheses). Why should this cue be so much stronger than the other two?

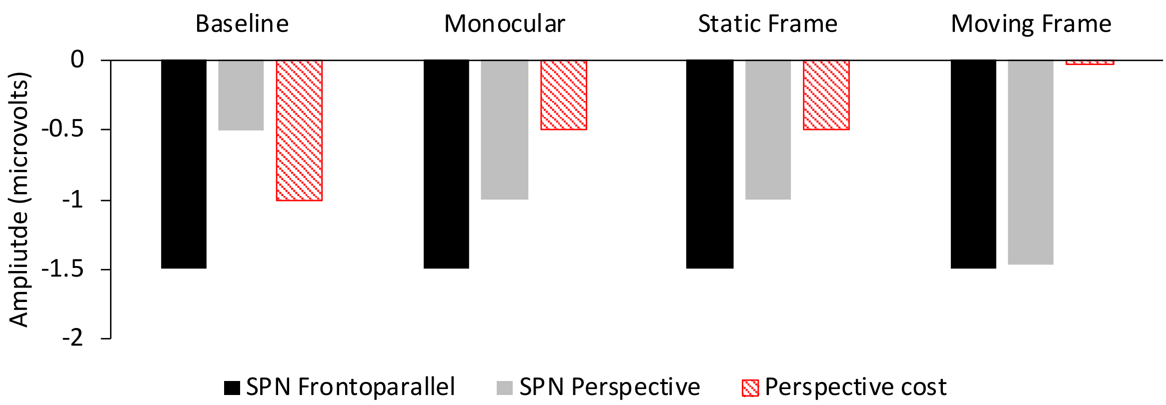
This is a very interesting consideration.

The Moving Frame cue is likely to be stronger than the Static frame cue because it contains the Static frame cue, and then the supplements it with an additional structure-from-motion cue.

We are not *extremely* confident that perspective cost will approximate zero in moving frame condition, but it is a reasonable assumption based on other results. We have added this reasoning:

> *"Given the results of Makin et al. (2015) and Karakashevska et al. (forthcoming), is possible that perspective cost will be around 1 microvolt in the Baseline block, and then reduce by 0.5 microvolts with each additional cue. It would thus reach zero in the Moving frame block, where frame and motion cues summate."*

We have also changed Figure 3 to make this aspect of the predictions clear:



"Figure 3. Predicted results. The SPN is the difference between symmetrical and asymmetrical conditions (negative bars represent a large SPN). The SPN may be larger (more negative) in frontoparallel (black) than perspective (grey) conditions. This difference is called perspective cost (red). We predict that perspective cost will be highest in the baseline block (left) and reduced in the other three blocks. Perspective cost may approach zero in the moving frame block (right). The predicted amplitude of these effects is more speculative than the rank order."

There is also another, more subtle motivation for Hypothesis 4.

Frontoparallel-perspective SPN equivalence (e.g. a perspective cost of zero) would be a very interesting observation in any Block. It would suggest that complete view invariance had been achieved, even though participants are performing a luminance task.

It would be a shame if the most interesting observation in our project were statistically confirmed with an analysis that was not pre-registered!

The sampling plan is a little vague (voluntary sampling isn't a thing - do the authors mean convenience sampling? How will participants be recruited? How much will they be compensated? Will their visual acuity and stereo vision be tested - and if so, how?). The power analysis seems sound and 120 participants is certainly a high number for an EEG study. The justification for the effect size of interest is good and is conservative, based on previous studies.

We have now changed the section about sampling:

> *"Participants will be recruited using convenience sampling and will be compensated for their time with vouchers or course credit."*

It isn't quite clear whether 120 is the initial number and more participants will be recruited on top of this if a participant's data need to be excluded (e.g. for poor performance on the behavioural measures). I assume this is the case but if could be clearer.

N=120 will be final number. Any participants excluded due to poor performance or EEG data problems will be replaced. We have now explained this in the participant section.

Hypotheses 1-3 can all be tested in a single analysis (the 2X4 RM ANOVA suggested in the analysis for H1). H1 can be tested by the main effect of Symmetry, while 2 and 3 can be tested by the suggested pairwise comparisons for block. (Since the hypotheses are directional, a 1-tailed test is most appropriate for each.)

This is very reasonable; however, we defend our current analysis plan as follows:

First, Hypothesis 1 is that there will be an SPN in the **Frontoparallel conditions**. As stated:

> *"For hypothesis 1 we will test for presence of an SPN in the frontoparallel conditions. We will run a 2 Regularity (Symmetry, Asymmetry) X 4 Block (Baseline, Monocular, Static Frame, Moving Frame) repeated measures ANOVA. We expect a strong main effect of Regularity."*

The factor Angle is not included in this analysis.

This comment may be more relevant to Hypothesis 3. There are three statistically equivalent approaches that one could take to Hypothesis 3:

1) We could run a 2 Regularity X 2 Angle X 4 Block RM ANOVA. The predicted Perspective cost effect would manifest as a significant 3-way interaction (which might be verbally described: *The effect of Regularity on amplitude was reduced in the perspective conditions, and more so in the baseline block than the other blocks, and more so in the Moving Frame Block than the Static Frame Block*).

2) We could compute the SPN (Symmetry-Asymmetry), run a 2 Angle X 4 Block RM ANOVA. The predicted perspective cost effect would manifest as a significant two-way interaction (which might be verbally described: *The effect Angle on SPN amplitude was greater in the Baseline block than the other blocks*).

3) We could compute perspective cost (frontoparallel SPN – perspective SPN). The predicted effect is now just a main effect of Block (which might be verbally described: *Perspective cost was greater in baseline block than the other blocks*).

Approaches 1, 2 and 3 would all have the same effect sizes and p value. However, 3 is the easiest to understand – it is the most direct reflection of our predictions. It also concisely segues into pairwise comparisons and equivalence testing. We therefore prefer approach 3.

Regarding the one-tailed t tests. These are of course justifiable when pre-registering a directional hypothesis, and one tailed testing presents a slightly more optimistic power analysis.

However, our impression is that some researchers are suspicious of one tailed t tests because they are associated with a crude form of p hacking (p = 0.08? So close! I'll have to do a one tailed test. Okay, p = 0.04. Good! Its significant now. I'll just pretend I predicted a directional effect…).

We are also mindful that the other reviewers and editors did not request one tailed tests.

We now discuss both one and two tailed approaches for maximum transparency without sacrificing brevity:

> *"Our planned sample of 120 provides 0.95 power for finding one tailed effect (d_z = 0.34, alpha = 0.02). A more conservative approach is to use two-tailed tests, even though we have a directional hypothesis. This reduces power to 0.91. Both criteria are more conservative than the convention demands (power = 0.8, alpha = 0.05)."*

For Hypothesis 4, I would suggest an equivalence testing approach. See Lakens, School & Isager (2018) for a tutorial on this. I really think this would be a much simpler approach than the currently suggested approach of testing against a specific value.

Our plan is indeed to use a one-sided equivalence test. In needs to be a one-sided equivalence, because we are interested in establishing the absence of substantial perspective cost (*not* the absence of substantial perspective benefits). We are not experts on equivalence testing, but our understanding is it does require a priori setting of a specific value (?). This is now explained with help of a new Figure 8:

> *"Hypothesis 4 predicts that meaningful perspective cost will be eliminated in the Moving frame block. This is different from other hypotheses because we are predicting absence of an effect. We will use a one-sided equivalence testing approach. Predicted results are shown in Figure 8A. Perspective cost in the Moving Frame block is predicted to be significantly above -0.35 microvolts (our definition of a small negative ERP effect). The same conclusions would follow from results in B (despite significant difference from zero) and C (despite no significant difference from + 0.35). Figure 8D illustrates an alternative possible outcome where perspective cost is not significantly above -0.35 microvolts, and therefore hypothesis 4 would not be supported. In all cases, significance is established with one-tailed, one sample t tests.*

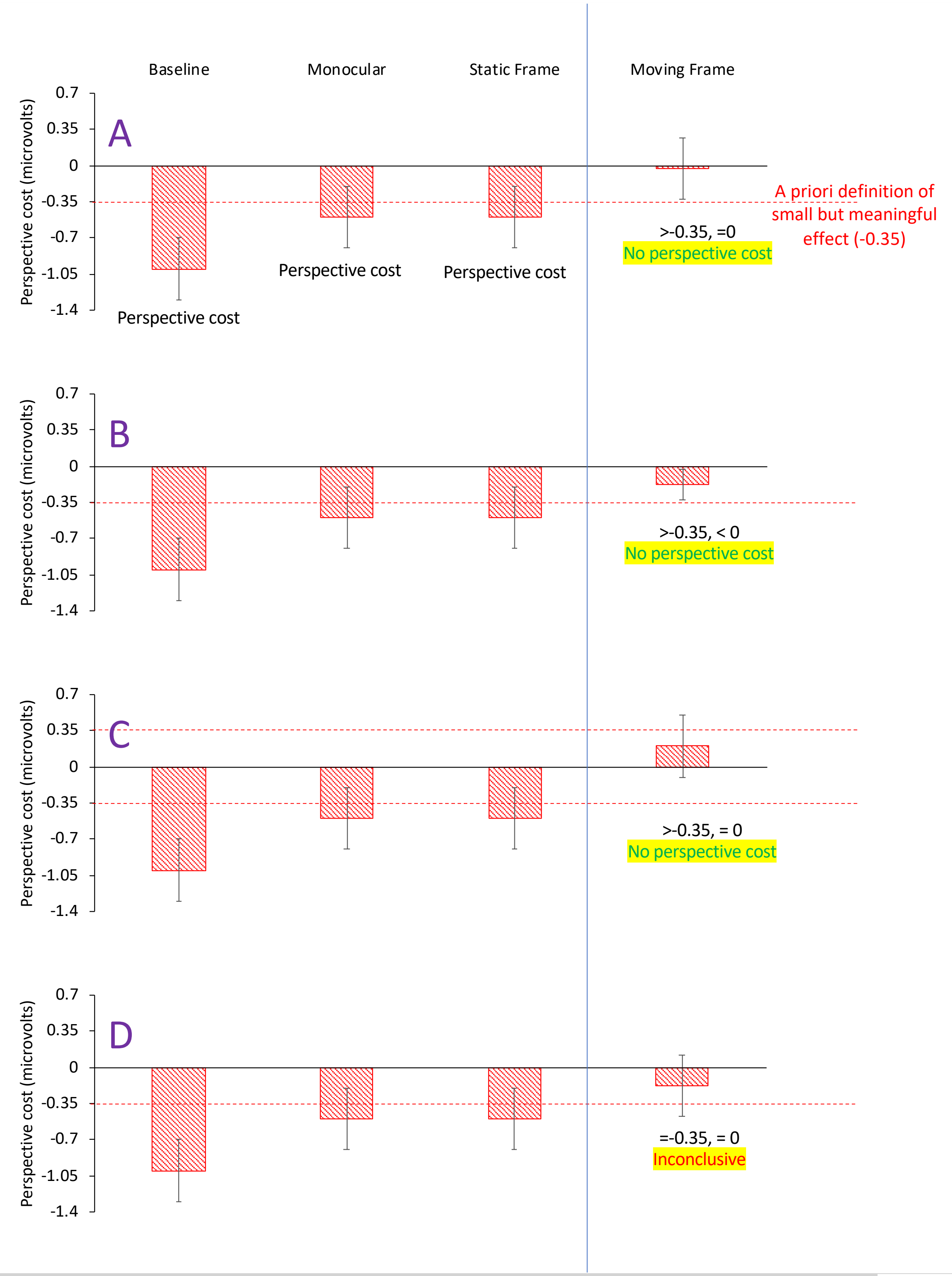


Figure 8. Different imaginary results to illustrate the one-sided equivalence testing approach. In Baseline, Monocular and Static Frame blocks, we predict a perspective cost. In the Moving Frame block, we predict no perspective cost. The crucial threshold is -0.35 microvolts (our a priori definition of a small negative ERP effect). If confidence intervals do not cross the -0.35 line, we will conclude that perspective cost is absent. This is the case in panels, A, B and C. In contrast, the results in panel D are inconclusive,

and do not establish the absence of perspective cost (despite non-significant difference from zero).”

Why do we rely so much on an apparently arbitrary -0.35 microvolt threshold throughout the paper? Why is this the correct definition of a small perspective cost? This is now justified at length in the manuscript, when talking about power analysis:

> *“We powered our experiment to find relatively small ERP differences of 0.35 microvolts. This threshold is informed by analysis of the 249 SPNs in the SPN catalogue (https://osf.io/2sncj/), described in Makin et al. (2022). Figure 4 illustrates relevant SPN distributions. Each ridge in Figure 4A represents a distribution of participant SPNs around the mean (the largest, most negative, SPN is at the base). The scatterplot in Figure 4B shows all 249 SPNs as data points, with mean amplitude is on the X axis, and Cohen’s d_z (Mean / SD) on the Y axis. The second order polynomial regression line indicates a plausible effect size d for an SPN of a given amplitude. This shows that -0.35 microvolt SPNs are likely to have Cohen’s d_z of -0.34. This also applies to within-subject pairwise differences between SPNs.*
>
> *Furthermore, as explained in Makin et al. (2022), 178 of the 249 SPNs in the catalogue are significant ($p < 0.05$, one sample t test against zero, two-tailed). The smallest significant SPN in the catalogue is -0.342 microvolts. Our threshold of -0.35 microvolts is thus a reasonable a priori definition of a small but meaningful SPN or SPN modulation.*

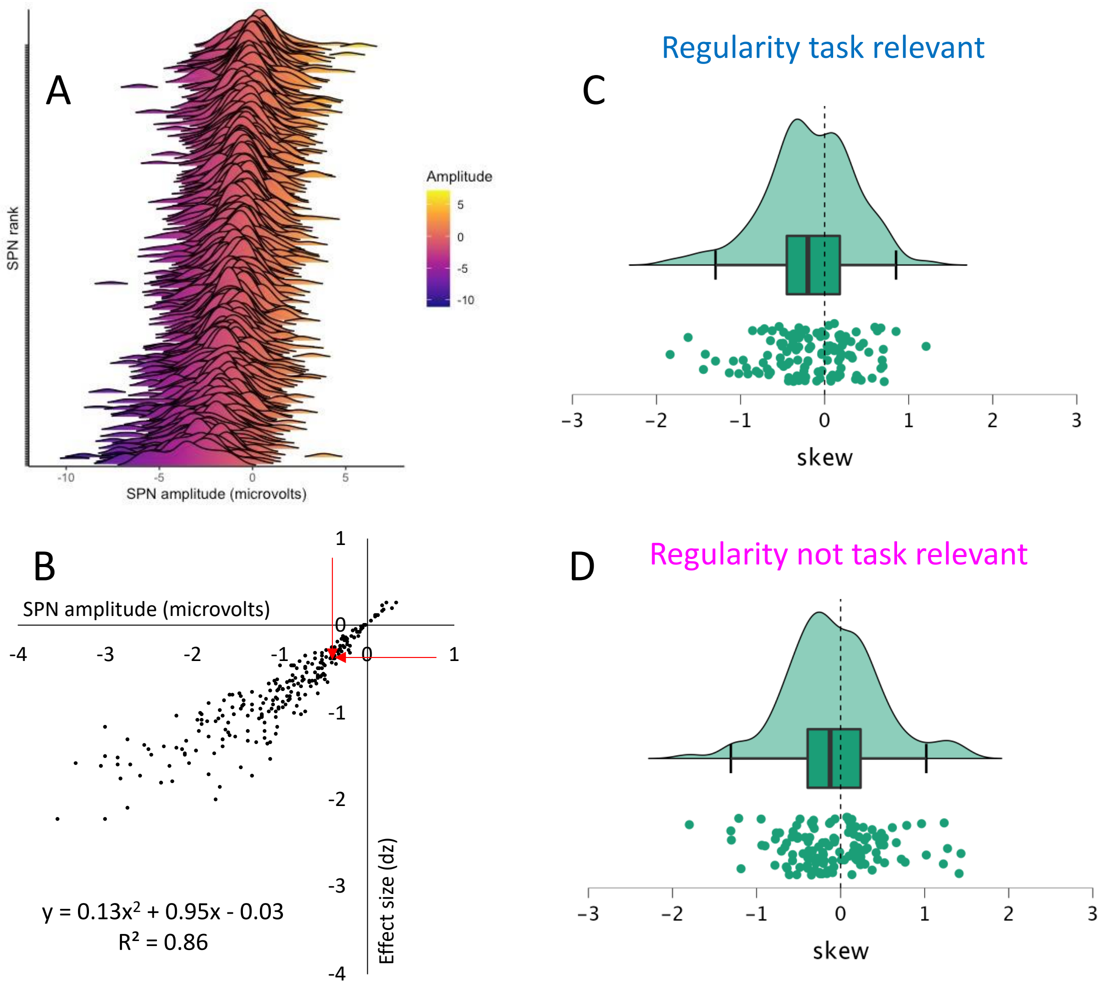


Figure 4. A) Distribution of 249 SPNs from the SPN catalogue (https://osf.io/2sncj/), shown as a ridgeplot. Each ridge is a distribution of individual participant SPNs around the mean. The largest (most negative) SPN is at the base. B) Scatterplot of 249 SPNs. The X axis is SPN amplitude in microvolts. The Y axis is standardized effects size (Cohen's d_z). The second order polynomial line suggests -0.35 microvolt SPNs have a typical effect size d of -0.34 (red arrows). C) Distribution of skewness statistics from experiments where regularity was task relevant. D) Distribution of skewness statistics from experiments where regularity was not task relevant.

Our planned sample of 120 provides 0.95 power for finding one tailed effect (d_z = 0.34, alpha = 0.02). A more conservative approach is to use two-tailed tests, even though we have a directional hypothesis. This reduces power to 0.91. Both criteria are more conservative than the convention demands (power = 0.8, alpha = 0.05). We also note that the median sample size in previous SPN research just 24. Our sample of 120 is thus more than twice as large as any published or unpublished within-subjects SPN experiment.

We verified these decisions with a power simulation approach. We computed a power analysis on 10,000 observations from a bivariate normal distribution with a specified correlation of 0.5 between conditions. This confirms we have 90% chance of finding a *mean pairwise difference of 0.34 SDs with a sample of 120 (codes for the simulations can be found here: https://osf.io/utq8e).*

Hypothesis 4 predicts an absence of perspective cost in the Moving frame block. Here we will use a one-sided equivalence testing approach (illustrated in Figure 8). If true perspective cost is -0.35 microvolts in given a block, we are likely to find that the effect is significantly below zero microvolts with one tailed one sample t test (power = 0.95, Cohen's d_z = 0.34, alpha = 0.02, one-tailed). Conversely, if true perspective cost is zero microvolts in given a block, we are likely to find the effect is significantly above -0.35 microvolts (power = 0.95, Cohen's d_z = 0.34, alpha = 0.02, one tailed)."

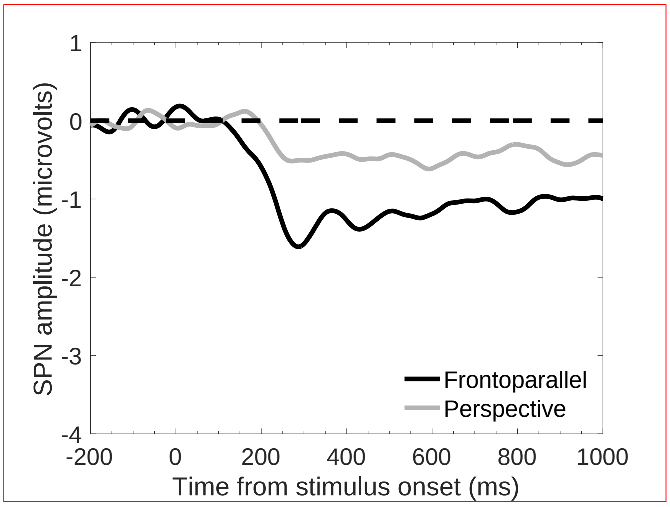
As a minor point, Figure 5b is almost impossible to read - specifically the diagram of the degrees of rotation. Perhaps this could be a separate figure?

We have now separated Figure 5 (Now Figure 6 and 7)

The task that participants will perform could be more clearly set out, perhaps in a figure. If it is designed to be relatively easy, how do we know that participants are not consciously attending to symmetry in the stimuli? It seems that the task (illustrated in figure 5B) is to say whether the stimulus is light or dark - but compared to what? Is there a standard? This isn't clear from the figure.

The task is to categorize the dots. They are either in the light category or the dark category. We have used this task with the stimuli before, and as you say, it is trivially easy (over 95% correct for most participants). Given this, it could well be that participants start spontaneously attending to symmetry, even though this is not their primary task. This could abolish perspective cost.

As explained in response to the editors, we have now completed a similar experiment with a similar luminance task. We call this Karakashevska et al. (forthcoming). There was a substantial perspective cost. The results are shown here:



Reviewer 4 Figure 2. Perspective cost in the luminance task of Karakashevska et al. (forthcoming).

Crucially there was a perspective cost during the Luminance task with Gabors (t (39) = 2.629, p = 0.012, dz = -0.416). This suggests our proposed task does not eliminate perspective cost because it is too easy.

Furthermore, stimulus presentation duration will be reduced to 500 ms in our new experiment (compared to 1000 ms in Karakashevska et al. forthcoming). This reduces time available for spontaneous task switches and gives a purer measure of perspective cost.

In general, the methodological detail here is excellent, and I particularly commend the authors for sharing all the study code on OSF, as well as the EEG processing pipeline. One thing that was not very clear to me was why the horizontal and vertical EOG channels are recorded when they are not used in the analysis (eye movements seem to be subtracted via ICA analysis, which doesn't include those channels). Also, the authors state that a "semi-automatic" process will be used to remove bad channels (p.15) - what is this process?

We do clean the data from eye movements and artefacts using ICA. However, we still use the external electrodes when collecting data, as they provide the strongest blink and eye movement signals. We can then ask participants to be mindful and avoid blinking when the critical stimulus is on the screen.

The semi-automatic process presents a visualization of all channels and makes it clear when one is far nosier than the average. The decision about whether the interpolate or not is make and entered by data analyst using a GUI. We have added this to the EEG analysis section:

"This presents a GUI where anomalous channels can be identified visually through variance distributions."

Another thing that isn't immediately clear from the report is how the SPN will be computed. The time frame and electrodes are given, which is great, but is the SPN the mean voltage difference across all electrodes of interest for the entire time period? Or are the peak values calculated? How will the numbers which go into the repeated measures ANOVA for each participant be calculated? I could not determine this from the report or, indeed, from the pipeline code provided. It is important that this is clearly set out in Stage 1, because as well all know, EEG analysis provides a large number of researcher degrees of freedom in this area!

The initial Biosemi_Pipeline.m Matlab script provides electrode X time data for each participant and condition (averaged over trial).

We will then use another script called 'stats extractor 2020.m'. This allows the analyst to list conditions, enter a time window (e.g. 300 to 600 ms) electrode cluster [e.g. P3 P5 P7 P9 PO7 PO3 O1 P4 P6 P8 P10 PO8 PO4 O2]. This script *averages* amplitudes over electrodes in the cluster, and then *averages* amplitudes over timepoints in the window. Amplitude in a spatiotemporal cluster becomes a single number.

This script outputs 8 columns (condition) X 120 rows (participant).

Subsequent stages are not implemented in a MATLAB pipeline.

Eight SPNs will be computed for each participant. This is the difference between symmetry and asymmetry conditions (black and grey bars in Figure 3). The four perspective costs will be computed as the difference between frontoparallel SPN and perspective SPN (red bars in Figure 3).

We have now clarified in the paper.

"We will average amplitudes across time windows, and then across electrodes, using the stats extractor 2020.m script (https://osf.io/vu2m7)."

Overall this is a very interesting study and the report is very well set out.

We are pleased with this positive review.