Peer review of Stage 1 Registered report titled “A multilab investigation into the N2pc as an indicator of attentional selectivity: Direct replication of Eimer (1996)”

**Review template**

For this review, I am using a modified version of a review template that was original developed by Dr Warrick Roseboom with the help of other colleagues in the School of Informatics and the School of Psychology in the University of Sussex (Dr Maxine Sherman, Dr Peter Lush, Dr Zoltan Dienes).

**Reviewer details**

Name: Dr Reny Baykova

Affiliation: School of Informatics and School of Psychology, University of Sussex

Relationship to the authors: None

**Are you an expert on this topic/method?** Please specify aspects of this study that you feel you are expert in and those which you are not (e.g. you know a lot about the topic, but maybe not so much about the precise methods used here, or vice versa):

|  |
| --- |
| I have experience with different types of EEG analyses and different pre-processing methods. I also have experience with pre-registering behavioural studies. I have not done meta-analyses before, and I am not deeply familiar with the literature around the N2pc. |

**Summary of paper (a few sentences):**

|  |
| --- |
| Constant et al. are proposing to conduct a pre-registered replication of the second EEG experiment reported in Eimer (1996) which investigates the N2pc component of the visual event-related potential. The N2pc can be observed in parieto-occipital electrodes contralateral to the presentation location of a target stimulus which is present along several non-targets in different locations on the display. In Experiment 2, Eimer (1996) found the N2pc can be elicited when the target stimulus is presented along with just one non-target stimulus shown on the other side of the display. In addition, Eimer (1996) showed that the N2pc can be elicited using different types of stimuli – forms and colour, with forms eliciting a larger N2pc than colours. Constant et al. are planning to investigate where they can replicate the three key findings reported by Eimer (1996): (1) the N2pc can be elicited in a form discrimination task; (2) the N2pc can be elicited in a colour discrimination task; (3) the amplitude of the N2pc elicited in the form discrimination task is greater than the amplitude of the N2pc elicited in the colour discrimination task. The project is part of the EEGManyLabs initiative, and currently the experiment will be conducted across 7 different labs. |

**Summary of opinion (a few sentences):**

|  |
| --- |
| I think this is a thorough pre-registration of the experimental design and analyses of the proposed project. In particular, the authors describe their analyses plans in a lot of detail. The authors have already provided skeleton analysis scripts in the online repository, which to me showcases their dedication to transparency and open science. In addition, replication studies of prominent effects and multilab studies are both amazing endeavour which I fully support. Below I have added some suggestions/ask for some clarifications that I think could make the project even better. |

**Transparency**

|  |  |
| --- | --- |
| 1. Will the data be made available on a public repository?  (make sure this is noted in “availability statement” or similar) | yes/no/comment |
| On page 7 of the manuscript, the authors have included a data sharing statement, stating that they will make raw data and analysis scripts publicly available. Just to confirm, will you also share the data of participants who have been excluded from the analysis?  On page 8 of the manuscript, the text states that some pilot data has already been collected. Would it be possible to share this data, so that I can run through the appropriate scripts that are on the online repository? | |
| 2. Is the code to reproduce the results available with the data?  (make sure this is noted in “availability statement” or similar) | yes/no/comment |
| Skeleton scripts for EEG pre-processing, the analysis conducted by individual labs, and the meta-analysis have already been shared on osf and github. Would it be possible to add the github repository in the manuscript as it is a bit easier to interact with the files on github than on osf? I am now not sure how I found the github page. | |
| 3. Have you checked the code/data? | yes/no/comment |
| I have only visually inspected the code. If pilot data is made available, I will run the code and amend my responses to the following questions. | |
| 4. Does the code run without requiring unreasonable amendments?  (e.g. changing directories, downloading repositories, fixing versioning is fine. Broken code or code not producing same results not fine). | yes/no/NA |
| - | |
| 5. Will the proposed analysis use inferential statistics?  (e.g. any statistical test/model/method against a criterion) | yes/no |
| The proposed analysis will NHST inferential tests to analyses the data of each lab individually, and the results of all labs together in a meta-analysis. | |
| 6. Is there a justification provided for the sample size in relation to this test(s)? | yes/no/comment |
| On page 7 of the manuscript, under “sample size and inclusion criteria”, the authors provide a power analysis to justify the sample size that will be collected by each participating lab.  Bearing in mind I have never done meta-analyses before,I was wondering whether it would be appropriate to do a power analysis for the meta-analysis as well? Is this something you have looked into? | |
| 7. If yes, is the provided justification reasonable?  (because “someone used this sample before” is not reasonable) | yes/no/comment |
| The justification for the required sample size is reasonable. The code for the power analysis is also available on the online repository as R and python code. I ran the R code and got the same results. I have not run the python code but it looks like it will give the same results as well.  When computing the sample size to be collected by each lab, the authors report two of the three numerical results in Eimer (1996) that they are attempting to replicate. For completion, I would suggest reporting the original results regarding the third hypothesis as well (the comparison of N2pc amplitude between forms and colours). It won’t change anything in the power analysis itself because the associated F-value is larger than the one used to compute the required sample size, but I think it will be good to have it there. | |
| 8. Are the primary hypotheses/analysis methods pre-registered? | yes/no/comment |
| 9. Does the stage-1 registered report robustly reduce researcher degrees of freedom? (ideally, for pre-registered tests there should be no discernable researcher dof) | yes/no/comment |
| I think the stage-1 registered report robustly reduces research degrees of freedom. The authors describe all aspects of the analysis in great detail – the hypotheses, the exclusion criteria for individual epochs, the exclusion criteria for participants, and the exclusion criteria for participating labs, the EEG pre-processing and analysis.  I have a few questions listed below:   1. Would all participating labs use the same instructions script, and could you share the instructions (apologies if this is already in the experimental files that have been shared, I didn’t manage to get to those)? 2. For completion, could you add the viewing distance that participants will sit at? 3. On page 4 of the manuscript, it says that the practice block will run “until the experimenter judges from the HEOG waves that participants are holding their eyes sufficiently still”. Does that mean until the absolute magnitude of the HEOG is below 25 µV for a particular amount of time? Does the practice block have a minimum length in terms of numbers of trials to ensure that participants have been presented with all target configurations? Will the practice block be repeated with participants start the second condition? 4. On page 4 of the manuscript, it states that “more laboratories might join after the in-principle acceptance”. Could you justify the decision to include additional labs after the in-principle acceptance? Could you also add more details, for example: (1) how many labs you will be looking to include; (2) when would be the deadline for additional labs to join – I imagine ideally that would be before data collection starts? 5. On page 4 of the manuscript, it states that “The first “Original” pipeline is the direct replication attempt, and the alternative pipelines will be used to cross-validate the results with more modern processing techniques”. What would this cross-validation consist of in the paper? | |
| 10. Do the presented analyses deviate from the preregistration in any way? (if yes, please identify deviations) | yes/no/NA |
| 11. If yes, are the deviations justified/reasonable? | yes/no/comment |
| 12. If present, are exploratory analyses clearly marked as such? | yes/no/NA |

**Content**

**Introduction**

|  |  |
| --- | --- |
| 13. Are the presented hypotheses well-motivated by presented literature? | yes/no/comment |
| 14. Are the presented hypotheses based on theories or assumptions that are reasonable or supported by the literature?[[1]](#footnote-1) | yes/no/comment |
| The first two hypotheses are based on the assumption that N2pc reflects selective attention allocation towards a target stimulus, and I think this is reasonable. The authors mention that these hypotheses are also consistent with other theoretical interpretations of the N2pc (more on this below).  It would be good to add a theoretical justification for the third hypotheses as well – the expected difference in N2pc between forms and colours is only justified. Unless I am missing something, I think the only justification for this hypothesis is that Eimer (1996) report a difference between the two conditions. Apologies in case this is already there and I have missed it.  Related to this, I think it would be worth explaining why the last column in Table 1A, “**Theory that could be shown wrong by the outcomes**”, is filled with NAs. If the study cannot inform theory, I think it would require further justification. | |
| 15. Is the presented literature review a fair reflection of the existing literature?[[2]](#footnote-2) | yes/no/comment |
| I am not too familiar with the literature around the N2pc so I cannot provide a detailed review on this aspect of the paper. To me, the introduction provided a short but clear overview of the relevant literature and the competing interpretations of the N2pc. Maybe one thing I was left wondering after reading the introduction is why replicating Eimer (1996) is important if it can’t distinguish between the different potential interpretation of the N2pc. Could you elaborate on what it would mean for the field if the effect doesn’t replicate? | |
| 16. Is previous work cited appropriately?[[3]](#footnote-3) | yes/no/comment |
| I am not too familiar with the literature around the N2pc so I cannot provide a detailed review on this aspect of the paper. I did read Eimer (1996) and compared approach of the original study with the approach of the proposed replication, and I think that Constant et al. cite it appropriately. | |

**Results/Analysis**

|  |  |
| --- | --- |
| 17. Have the analysis methods been applied correctly (as specified) as far as you can tell?[[4]](#footnote-4) | NA |
| Not applicable as this is a stage 1 registered report. | |
| 18. Is the experimental paradigm a close enough replication of the experimental paradigm used in the original study? Are deviations from the original design noted?  (this question is not part of the original template) | yes/no/comment |
| Looking at the original experimental design, I think the Constant et al. follow the procedure detailed by Eimer (1996) closely. I think all differences between the two studies have been noted clearly by the authors and I don’t expect these differences to have an effect on the results. | |
| 19. Is the paradigm appropriate for testing the hypotheses?[[5]](#footnote-5) | yes/no/comment |
| I think the experimental paradigm is appropriate for testing the stated hypotheses. If prior to the in-principle acceptance I come up with a reason why I think it isn’t, I will get in touch with the authors. | |
| 20. Is the proposed analysis a close enough replication of the analysis conducted in the original study? Are deviations from the original analysis noted?  (this question is not part of the original template) | yes/no/comment |
| Constant et al. propose using 2 different EEG pre-processing pipelines and 2 different statistical analyses, resulting in a total of 4 EEG analysis pipelines – “Original”, “ICA”, “Non-parametric bootstrapping”, and “ICA and non-parametric bootstrapping”. Of these, the “Original” pipeline is attempted as direct replication of the pre-processing pipeline of the original study.  Comparing the pre-processing steps in the “Original” pipelines and the pre-processing steps in Eimer (1996), I think Constant et al. have followed the original paper closely. I think all differences in the pre-processing between the two studies have been noted clearly by the authors, and I don’t expect these differences to have an effect on the results. I have a few questions on the statistical analysis. Could you describe how the statistical analysis in the original paper was conducted? From reading the description of the analysis on page 227 in Eimer (1996) and the ERP results on page 230, I can’t determine with great confidence what they did. It looks like they computed several ANOVAs – one with all the data, and then two more with the two task conditions separately. Could you also say why you decided to conduct the statistical analysis in the “Original” pipeline differently? Could you also say why you didn’t consider analysing the data using a 2X2 ANOVA with experimental task added as a predictor in addition to electrode laterality? Finally on the EEG analysis, could you say why you decided to only replicate the results for the occipital electrodes, and exclude the analysis of the parietal electrodes?  Are you planning on replicating the behavioural analysis in the original paper? | |
| 21. Are there any clear methodological confounds/errors in the results?  (provide a full explanation of why, citing references where necessary). | yes/no/comment |
| As with many neuroimaging studies, the results would be confounded by some amount of motor preparation. I think motor preparation would be of concern in this task also because the response keys – left and right – correspond to the different targets rather than the target location. Could you discuss this? Also, would responses be given with different fingers in one hand (e.g. right hand only, index finger presses the left button and middle finger presses the button) or with two hands (e.g. index finger of left hand presses left button, index finger of right hand presses right button).  I am not sure whether changing the types of responses would be appropriate in a direct replication.  Have you considered the possibility of using an eye-tracker to confirm that participants remain fixated at the central cross?  What is the reason for doing all 6 blocks associated with 1 condition followed by the 6 blocks associated with the other condition, rather than intermixing the order of the blocks? | |
| 22. Are there any analyses that you can think of that have not been considered and would potentially reflect alternative interpretations? | yes/no/comment |
| Before going through the repository, I was going to suggest including Bayesian t-tests for each of NHST t-tests. After going through the repository, I saw that the paired t-test script is set up to compute Bayes factors using the JZS prior, but if I am reading the code correctly it seems you have currently decided not to calculate Bayes factors (there are also not listed in the manuscript amongst the statistics you are planning to report). Therefore, I am going to proceed with my recommendation to include Bayes factors and to compute them using the approach detailed in Dienes (2014). Since your hypotheses are directional, my suggestion is to define the prior as a half-normal distribution with a mode of 0 and a standard deviation equal to half the prior-expected effect size (the prior effect being the difference in amplitude found by Eimer). think this would be very useful in case some of the effects don’t replicate because you will be able to see if you can provide evidence in favour of the null.  The paper on Bayesian analysis I reference is available here and it details the approach I described above: Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in psychology*, *5*, 781. | |
| 23. Do the data figures make sense and have clear labelling and explanation? (give suggestions of how the figures could be made more intuitive for you). | NA |

**Discussion**

**This section of the review template is not applicable as this is a stage 1 registered report, but I have left the questions here for reference.**

|  |  |
| --- | --- |
| 22. Does the Discussion provide a fair natural language account of what happened in the experiments in relation to stated hypotheses? | NA |
| 23. Does the Discussion provide a fair account of the links between the obtained results and motivating results outlined in the Introduction? | NA |
| 24. Are there any other relevant relationships between the results observed in this study and other literature that have been overlooked by the authors?[[6]](#footnote-6) | NA |

**Wildcard**

|  |  |
| --- | --- |
| Does the paper develop/deploy a method in a new and/or interesting way? | yes/no/comment |
| This is a replication, so the methods are limited to the original study. | |
| Does the paper provide a link between known results in an interesting way? | yes/no/comment |
| This is a replication so links between known results will be limited. | |
| Is there generally something about this paper that makes you go “Wow, I wish I had thought of that”? | yes/no/comment |
| Doing a replication in a multilab study is an admirable endeavour and I really want to be involved in a project like this in the future. | |

**Minor comments**

Please identify basic mistakes that you notice like typos, grammatical errors, or sections of text that could otherwise be expressed more clearly. Provide suggested alternative phrasing if you wish.

|  |
| --- |
| 1. In Table 1A, could you add the refresh rate of the monitor that will be used by the lab in Málaga, and the resolutions of all the monitors? 2. What is the reasoning for the differences in the impedance thresholds between labs? And aren’t these thresholds a bit high? 3. At the top of page 7, the last bullet point above “sample size and inclusion criteria” references with: “… forms or letters…” . Did you mean to say “forms or colours”? 4. On page 7, under “sample size and inclusion criteria”, the text says: “The most representative result are the effects of contralaterality in Study 2 (which is the replicated study) for electrode pair OL (corresponding to PO7/8 in the 10-10 system)”. Did you mean to say “electrode pair OL-OR”, or am I misinterpreting something? 5. Can you explain the existence of the target-only trials if they are excluded from the analysis? 6. The next comment is entirely subjective, so feel very free to ignore it. If I was making Figure 1, I would extend the shaded area which represents the time window of interest to cover the full height of the graph. There is something about the off-centred placement of the time window indicator in 1A that I find very unnerving. |

1. For example, hypotheses that are based on homuncular accounts of cognition (e.g. “we track time using an internal stopwatch, therefore if X then Y”) are unreasonable. It is also common to misinterpret measures as theories, e.g. intentional binding is an implicit measure of sense of agency, therefore when I measure temporal binding I am measuring implicit agency. Are any caveats/limitations like this acknowledged, either in the introduction or the discussion? [↑](#footnote-ref-1)
2. Note, this doesn’t mean comprehensive. Not all papers need to be review papers. Make sure to provide full references to papers you believe to be relevant but missing. [↑](#footnote-ref-2)
3. Cited work not relevant, too general, misinterpreted, retracted, unreliable or debunked, citing secondary sources, overuse of review papers. If no, provide explanation. [↑](#footnote-ref-3)
4. e.g. no optional stopping in frequentist test; no impossible degrees of freedom/t values/p values. See Appendix A in doi:10.1177/2515245918806489 for comprehensive list of things to look out for. [↑](#footnote-ref-4)
5. e.g. when making claims about how participants’ perception of a stimulus differs between conditions, does the task measure performance (sensitivity) or just propensity to report (bias); if you don’t know what this means refer to Kingdom and Prins, Psychophysics: A Practical Introduction, Chapter 2. Another common problem is the experimental manipulation not being precise enough, e.g. assuming that TMS can selectively target particular cells; TMS to the neck would not be an appropriate way of testing the causal role of baroreceptors in interoception, because TMS would stimulate the whole neck, not just baroreceptors. [↑](#footnote-ref-5)
6. Note, this is not meant for you to request references to your own work unless actually related. [↑](#footnote-ref-6)