Decision for round #1 : *Revision needed*

Understanding oscillatory correlates of pain expectation

Dear authors,

Many thanks for submitting your Stage 2 report for review. Please find attached the 3 reviewer opinions of your Stage 2 report, which are generally favourable.

Reviewers 1 & 3 have made a few minor points that should be reflected in your re-submission.

Reviewer 2 makes a stronger point around the interpretation of your non-significant EEG results. After consideration, and re-review of your Stage 1 plan, I think the difference lies in the precise terminology that was accepted at Stage 1. You had stated in the pre-registration that "no definitive conclusions will be drawn from a non-significant result", and I think on balance that this does allow for some interpretation of the observed result compared to if the pre-registration had stated that no conclusions would be drawn at all. Whether this should have, in hindsight, been tightened up at Stage 1 is an open question, but I do believe that your interpretations fit within the previously accepted plan.

Best wishes,

Gemma Learmonth

Dear Prof. Learmonth,

We are grateful for the positive evaluation of our manuscript and are confident that the adaptation made based on the feedback of the reviewers has improved our manuscript. Regarding the interpretation of our results, we share your opinion that the pre-registration did leave us room for some cautious interpretation of the results, as long as it is transparently discussed for which effect size we did not have sufficient power to detect a change. We would also like to highlight that within our field of experimental neuroscience, the estimated effect size is the standard criteria for sample size calculation and interpretation of results. At the same time, we do not want to negate the importance of the concept of “the smallest effect one would possibly still be interested in” proposed by Prof. Dienes. We hope that with the most recent changes in the manuscript, we were able to paint a nuanced picture of these sample size considerations as well as of the “interpretability” of our results that does the topic justice.

Best wishes,

Chiara Leu

Review by Markus Ploner, 20 Feb 2024 07:57

The stage 2 manuscript accurately follows the outlines of the stage 1 manuscript. The results have shown primarily negative findings, indicating a lack of a relationship between phase-locked responses and ongoing oscillations on the one hand and pain perception on the other hand. This conclusion is hampered by the limited sample size and failure to induce pain in most stimulation conditions.

We thank the Reviewer for their careful consideration of our manuscript. To be more transparent on the rationale of the sample size and to further explain the use of the "smallest possible effect one would be interested in" (Dienes, 2021)[1], we moved the chapter on the sample size justification which was written for the Stage I manuscript (in the later rounds of reviews) from the Supplementary Materials back to the main manuscript. Given our results and their limitations, it is indeed imperative to give the reader a detailed understanding of our sample size rationale.

The manuscript presents the results clearly and discusses the findings appropriately. Some clarifications and added details might further improve the manuscript:

1. *P.6, first line. Remove "be".*
2. *P. 6, third paragraph. Why were the stimulation temperatures changed compared to the temperatures specified at stage 1?*
3. *P. 6, third paragraph. Replace "will be" with "were"*
4. *P. 8, line 6. Remove "be"*

We thank the Reviewer for pointing out these typographical errors and have adjusted them accordingly. The temperature mentioned on p.6 was indeed also an error from a previous version of the manuscript, which had been written before the finalization of the protocol in our last pilot (changes of +0.5°C and +1°C for medium and high intensity stimulation, respectively). Nevertheless, we would like to apologize for this error, as we should have noticed this error prior to submission of the Stage I manuscript.

5. *P. 10, second paragraph. "the resulting amplitude was multiplied by the number of averaged chunks." Multiplied or divided by the number of chunks?*

The chunks resulting from cutting the EEG signal around the peaks at the frequency of stimulation and its harmonics are first averaged (therefore divided) and then multiplied by the number of chunks used for the average. We adapted this sentence to make this clearer.

6. *P. 24, first paragraph. "neither the expectation of a similar stimulus nor the mismatch in perception for condition HM seemed to have an influence on the recorded amplitude." This sentence was not clear to me. Please rephrase.*

We agree that this sentence was somewhat convoluted. We rephrased it to state more clearly that the mismatch between expected and perceived stimulus intensity did not seem to have an influence on the recorded amplitude, as the magnitude of the amplitude seemed to have been primarily driven by the actual temperature of stimulation.

[1] Dienes, Z. (2021). Obtaining Evidence for No Effect. *Collabra: Psychology*, *7*(1). https://doi.org/10.1525/collabra.28202

7. P. 25, last paragraph. "the smallest possible effect size that we would still be interested in." What would be the smallest effect size the authors would be interested in? And what were the criteria for defining this effect size?

The concept of the need to power for the smallest possible effect size that would still be interesting rather than for using a more conventional sample size that that power for an expected effect size is based on a recent publication of Dienes (2021)[1] (also a reviewer of this Registered Report). We would like to refer you to this publication for an in-depth discussion of the topic. Generally, the concept is that by relying on effect sizes for sample size calculations that is estimated from previous publications or general power calculation tools such as G*Power (ref), one risks missing out on effects that are smaller than predicted/expected. Not finding an effect based on this calculation does not necessarily mean the true absence of an effect, but rather that it could be smaller than anticipated. To avoid this risk, one can calculate the "smallest effect one would still be interested in", which is reflected by calculating the 80% confidence interval of the statistical model that was used to calculate the estimated effect, and using the lower bound of this confidence interval as the new model estimates (also described in our sample size section that was added in the Stage II Supplementary Materials). We understand that this distinction is not commonly done, and that clarification is important to understand the steps in our manuscript. We hope to clarify this for the reader by adding the section on the sample size which was previously in the supplementary material to the main manuscript.

8. P. 26, first paragraph. The study's main result is a negative finding that cannot be conclusively interpreted. This somehow disappointing outcome is mainly due to the limited sample size. The authors might discuss their sample size calculation and lessons learned for future studies more critically and openly.

We thank the Reviewer for this remark. As mentioned previously, we agree that the discussion of the sample size is a relevant part of this study. Importantly, we would like to highlight that while we did not have a sufficient sample size to detect the "smallest possible effect one would be interested in", we did have almost twice the sample size that would have been required by more conventional sample size calculations (reaching a power of 90% with an alpha level of 0.02). The detailed description of the simulation-based sample size calculation approach has previously been outlined in the Supplementary Materials of the Stage II submission and can now be found in section 2.2, pp. 6-8. We further adapted the wording in this paragraph, to state more clearly for which effect sizes we did and did not have a sufficient sample size.

Nevertheless, given our negative results it cannot be denied that a study with more participants could reach a different result. As the recorded EEG amplitudes related to the medium intensity stimulation were much lower than expected, we also have to consider that we overestimated the initial effect size in our simulation of the sample.

We would like to highlight that while the "smallest effect one would still be interested in" - a concept still rather unknown in our community - is theoretically interesting and valid, our experiment would be considered to have a sufficient sample size by the general publication standards in our field.

Review by Zoltan Dienes, 20 Feb 2024 16:46

The authors have conducted the analyses they said they would; though I now note that it slipped us all by, that the precise analyses were not absolutely nailed down beforehand; for example, exactly how post hoc tests would be performed appears not to be pre-registered. However, whatever extra flexibility snuck through, the key theoretical finding was non-significant. So nothing need be done about this point.

My main point is that the pre-registration declares no conclusion follows from nonsignficant results for the EEG. The way the authors deal with this is make conclusions but add a paragraph saying do not take them seriously. That is to write in contradictions. The correct thing to do is draw no conclusions in the first place - a point which applies to the abstract as well. The discussion and summary of results in the abstract need a major re-write therefore. I realise the authors may wonder what to write about. They could say "whatever the difference is between HM and LM it lies in this interval" and give a confidence interval, and declare no conclusion can be made yet as to whether or not there is a difference of scientific relevance.

We agree with the Reviewer that we pre-registered to not make any *definitive* conclusions on the results of our experiment. In our opinion, this still does allow for a tentative discussion of possible conclusions, as long as there are transparent statements about the limitations of the interpretation of the results. Still, we understand that the abstract might have not been phrased cautiously enough, and we have adapted it accordingly.

Additionally, we calculated the 95% confidence intervals of our post-hoc calculated effect sizes and added them to Table 1. This allows for a clearer interpretation of the results and contextualizes how precise the effects of our results are (Lee, 2016)[2]. A brief discussion of these intervals in relationship to the effect sizes has been added to the discussion.

We hope that you can understand our choice to retain a tentative discussion of possible of conclusions and find the changes made to the abstract and the discussion satisfactory.

[2] Lee, D. K. (2016). Alternatives to P value: confidence interval and effect size. Korean J Anesthesiol, 69(6), 555-562. https://doi.org/10.4097/kjae.2016.69.6.555

Review by Chris Chambers, 06 Feb 2024 16:00

I think the authors have done a great job with this Stage 2 submission. The study remained impressively close to the approved protocol; I found the reporting of results to be clear and the Discussion insightful in considering the implications and limitations of the research. Broadly, the manuscript in my view meets the Stage 2 criteria and I have only a few very minor comments for consideration.

We are grateful for the positive feedback and the acknowledgment of our efforts to adhere to the Registered Report process. We have adapted the language and wording where applicable. Finally, we are thankful for the feedback on the repository, as this was our first Stage II PCI RR submission. We hope to fulfill all criteria after extending the README file and adding any digital study material in the Harvard Dataverse repository.

*In a few places, the language implies evidence of absence from statistically non-significant results. e.g. (with my suggested modifications highlighted in **underlined bold**):*

- *p22 "The conditions of interest HM and LM did not differ **significantly** in their modulation at the frequency of interest"*
- *p26: "Despite a strong effect of the visual cues on stimulus perception, no **significant** differences were found in the modulation of ongoing oscillations at the frequency of interest between the conditions of interest (medium intensity stimulation preceded by either a cue for a high or a low stimulation intensity)."*

I suggest checking carefully throughout for other instances and adjusting accordingly.

pp25-26 "Frequently, the targeted effect size is the observed effect in previous literature; yet, this approach might lead to the rejection of a hypotheses only because the effect might have been smaller than in previous investigations and not because there was truly no effect". I would suggest replacing "the rejection of a hypotheses" with "lack of support for a hypothesis" as I initially read "rejection" to imply rejection of the null (which would convey the opposite intended meaning).

We agree that the language in some places was not precise enough and have adapted the manuscript accordingly.

Q5 in the Stage 2 submission checklist asks: ***5. Have all digital materials that are necessary and sufficient to reproduce all data acquisition procedures been made freely and publicly available? Such materials can include, but are not limited to, software code associated with data acquisition hardware, stimuli (e.g. images, videos), survey text, and digital or digitized questionnaires.***

The authors answer YES and linked to the Harvard Dataverse repository at https://doi.org/10.7910/DVN/40ZRQR However as far as I can tell, this repository contains only data, not the digital study materials. Please either add the digital study materials to this repository (and add their mention in the README file), or alternatively add them to the study's OSF repository (https://osf.io/9ud7x/) and add a mention of this repository to the Stage 2 manuscript (as it is not currently stated anywhere). Please also augment the README file in the Harvard repository to provide a complete list (inventory) of files with an accompanying description that defines the file content.

We thank the Reviewer for taking the time to look at the repositories. We apologize for not being more diligent when uploading the data and EEG analysis pipeline before Stage II submission. All digital study

materials such as R scripts for statistical analysis, MATLAB scripts used to run the experiment, as well as the EEG processing pipeline and all processed data sets that were used for the analysis have now been added to the Harvard Dataverse repository associated with this RR. The README file has been updated accordingly. We hope that these files complete the repository to make the study fully reproducible. We additionally added a link to the OSF project repository containing all versions (Stage I and II) of this RR (p.5).