Recommender's comments = black
Authors' comments = blue

PCI-RR Stage 1 Triage

Dear Dr. McIntosh,

Thank you for your second round of triage comments. We have updated the manuscript according to your comments and address each point below.

Yours sincerely,

Grace Edwards, Mica Carroll, and Chris Baker

The guidance from the Cortex RR guide for authors, states: "For inference by Bayes factors, authors must be able to guarantee data collection until the Bayes factor is at least 6 times in favour of the experimental hypothesis over the null hypothesis (or vice versa). Authors with resource limitations are permitted to specify a maximum feasible sample size at which data collection must cease regardless of the Bayes factor; however to be eligible for advance acceptance this number must be sufficiently large that inconclusive results at this sample size would nevertheless be an important message for the field." I recommend that you read the whole relevant section of that guidelines document.
Often, in these cases, we would suggest a formal Bayes Factor Design Analysis (BFDA) (see this paper by Schönbrodt and Wagenmakers (2018): https://link.springer.com/article/10.3758/s13423-017-1230-y). In the BFDA method, you run simulations to determine the probability that your experiment will return evidence in favour of H1 or H0. This method allows you to estimate--given your chosen BF threshold, your smallest effect size of interest, and your maximum sample size--what proportion of studies will stop because the evidential threshold has been crossed, and what proportion because max n has been reached. With BFDAs for each hypothesis, you can plot the simulation results to show the sensitivity of your design given your assumptions (see Figs 2 and 3 of the following paper for an example of how the simulation results can be presented): http://dx.doi.org/10.1037/bne0000345).
It is entirely up to you whether you want to go down a Bayesian or frequentist route for hypothesis testing, but the sample size calculation should be appropriate to the chosen method.

The concern about the unusual mix of the frequentist power analysis and Bayesian inferential approach to the statistics is well taken. Due to the original design of the study we plan to replicate (Ghin et al., 2018), we do not have an estimate of the within-subjects variance to inform a full power analysis approach, or a stopping rule for the Bayes Factors. Thus, we used the between-subjects variance to power our study, but this is a conservative estimate for the maximum participants we need. We included the

Bayes Factors to test H1 and H0 with our within-subjects design to potentially avoid running more participants than necessary and wasting resources. However, we do understand the concern about mixing frequentist and Bayesian approaches. Due to the lack of an estimate of within-subjects variance, we will continue with the frequentist approach. We have updated *Participants, Statistical Analysis, Expected Results, Study Design Table* and *Appendix B* accordingly.

2) In addition to your critical hypothesis tests, you propose some exploratory analyses to inform your interpretation. You can state an intention to add these exploratory analyses, as you do at the bottom of p7, but it is not necessary (or appropriate) to then describe in any detail how these will be conducted, or to include them in the core design table, because they are exploratory. The Stage 1 RR should specify the registered experiment, and any further exploratory analyses should be added at Stage 2. (On the other hand, if these exploratory analyses are essential to your purpose, they should be made into registered hypotheses, and specified as such.) Note that your Stage 2 conclusions must not be inappropriately guided by exploratory parts of your analyses,

We will remove the mention of our exploratory analysis from the *Methods* section of our manuscript, leaving just the intention to add the exploratory analysis in the *Introduction* (pg.7)

3) I previously suggested that you should be more precise in the distinction between replication and reproduction. On p6 you now state, “In order to implement some additional controls, we propose to reproduce the contralateral impact of hf-tRNS over left hMT+, rather than perform an exact replication. In contrast to an exact replication, we include only a selection of the original conditions, and have introduced extra within-subject stimulation controls in our design.”
Actually, in most widespread usage, what you describe is a form of replication (just not a direct or exact replication). In my understanding, the term ‘reproduction’ is generally taken to imply computational reproducibility of the same outcomes from exactly the same dataset.

Following the Recommender’s suggestion, we will propose our study as a replication with the specification that it is not a direct or exact replication in the *Introduction.*