Cerebral laterality as assessed by functional transcranial Doppler ultrasound in left-and right-handers: A comparison between handwriting and writing using a smartphone.

We are most grateful for the constructive comments by the recommender. The manuscript has now been revised and all of the comments were carefully addressed and changes made accordingly. The specific ways in which the manuscript has been modified to address the issues raised are discussed below. The titles of recommender's comments are in italics and the titles of our own responses in normal font.

Recommender's suggestion 1. Statistical approach / Power analysis:

You propose to use a default Bayes Factor with a Cauchy prior scaled with r=0.2 for Hypothesis 1. This hypothesis is set to confirm the null. By using a small scaling factor you are concentrating probability density for the alternative hypothesis near small values and this may very well make sense for what you are testing here - but it also limits your ability to actually confirm the null. It is important to ensure sufficient sensitivity to reliably detect hypothesised effects - or the lack thereof.

Using the original Rouder implementation of the Bayesian t-test with your minimum sample size of 32 your BF10 would be 0.4891 even for a perfect null effect (d=0). Even with your maximum sample size of 48 it is still 0.4332, nowhere near your criterion of 1/6 of confirming the null. While this calculation may be different in whatever implementation of Bayesian inference you're using (e.g. my understanding is that JASP has undergone some changes since Rouder's paper was published), this is the kind of information that is required to ensure adequate power.

If experiment is not sensitive enough to obtain conclusive results, you either need to adjust the prior or increase your planned sample sizes.

Author's response:

We thank the recommender for his observation that using a Bayes Factor with a Cauchy prior scaled with $r = 0.2$ for Hypothesis 1 would limit our ability to confirm the null. To address that, we have changed our Hypothesis 1 analysis plan. Specifically, we opted to go for an $r = 0.707$ (sqrt(2)/2) which has been proposed by Morey et al (2011) as an appropriate prior for detecting moderate evidence. We have also opted to change our BF10 to being greater than 1/3 for both Hypothesis 1 and 2 (page 8, line 13). With these criteria, there is an above 70% chance of detecting evidence in favor of the null for our initial sample size of 32 (Phylactou & Konstantinou., 2022). We have also opted to increase the maximum sample size from 48 to 50 (page 8, line 17), which corresponds to an 80% chance of detecting evidence in favor of the null for a BF10 greater than 1/3 (Phylactou & Konstantinou., 2022). The analysis plan for Hypothesis 1 now reads as follows (page 20, line 10):

> "Hypothesis 1: There will be no difference in the cerebral laterality during handwriting vs. writing using a smartphone after hand motion correction (linguistic component). To test Hypothesis 1, we will perform a Bayesian dependent samples *t*-test for the LI_handwriting_corrected and LI_typing_corrected. The prior chosen for the *t*-test will be a Cauchy distribution with a 0.707 width parameter, corresponding to a "medium" effect size, according to Morey et al., (2011)."

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*(4), 406–419. https://doi.org/10.1037/a0024377

Phylactou, P., & Konstantinou, N. (2022). Bayesian *t*-Test sample size determination: Reference tables for various Bayes factor thresholds, effect sizes, sample sizes, and variance assumptions. https://psyarxiv.com/jnp8c/

Recommender's suggestion 2: Hypothesis 2:
This hypothesis strikes me as a "main effect" of handedness. This should be a 1-df test that ignores the writing mode. Rather than two independent t-tests you effectively only need one for the sum (or average) of handwriting and typing. However, since you are interested in whether or not left-handers show weaker lateralisation than right-handers, isn't the crucial contrast here an interaction? Specifically you want to test whether the difference between left- and right-handers is weaker for typing than handwriting, and you hypothesise that this is a null effect. If this is incorrect, please clarify.

Author's response:
We thank the recommender for his comment and for his suggestion. We have now re-thought our original hypothesis and we have reformed it. We have decided to keep only the hypothesis relating to the potential differences in lateralization between right- and left-handers regarding typing because the differences regarding handwriting vs. typing will have already been examined in Hypothesis 1. Because of this change, examining the Hypothesis 2 using a *t*-test is the most appropriate approach. The updated Hypothesis 2 is as follows (page 8, line 5):

> "Right-handers will exhibit stronger left-hemispheric lateralization compared to left-handers for the linguistic part of writing using a smartphone (after hand motion correction)."

We have updated parts of the "Abstract" (page2, line 13) "Analysis Plan" (page 20, lines 16-18) and the "Hypothesis Table" (page 23) in accordance to the aforementioned changes.

Recommender's suggestion 3: Exclusion criteria:
Please provide more detail/information about how exclusions/rejections of participants will work.How will you detemine "Inadequate ultrasound penetration of the skull"? A similar question applies to the second criterion about noisy data: what determines whether epochs are "accepted"?

Author's response:
We thank the recommender for his suggestion. We have updated our text in order to address his concerns. The text now reads (page 14, line 19):

> "1. Inadequate ultrasound penetration of the skull (determined as inability to secure a clear ultrasonographic signal, making the ultrasonography impossible),
> 2. Noisy data (in the cases where less than 10 out of the 20 epochs in each condition are accepted; i.e., less than 10 epochs that have cerebral blood flow volume values in the

range of 70% to 130% of the mean velocity or an absolute left-minus-right channel difference less than 20% multiplied by the inter-quartile range of the individual)."

Recommender's suggestion 4: Hypothesis table:
This is perhaps a matter of taste. Certainly any journal you may eventually submit this to will probably want to copyedit this in any case. However, I found that table very difficult to digest. A lot of text is awkwardly squeezed into narrow columns. Words are broken across lines. Generally it is very hard to understand. I realise that such tables are often used in RR manuscripts but I would reduce the amount of text here considerably. Move most of the prose into the manuscript itself (where it should be read) and keep this to the essentials and/or abbreviated statements. Instead of a table, it might perhaps in fact be better to have a box diagram outlining the various possible outcomes? This will make it a lot easier for reaers (and reviewers!) to understand. **[Note added by Managing Board: please keep the design table in the manuscript as it is a required component, but abbreviate it as much as possible to improve readability. The idea of adding a box diagram or flowchart is a good idea]**

Author's response:
We thank the reviewer for his suggestion. We have changed the orientation of the page for the "Hypothesis Table" and have reduced some of the text in order to make the table more readable (pages 22-24). No further major change was made as per the note added by the Managing Board.

Recommender's suggestion 5: Quantifying handedness with the pegboard:
You have a thorough design for quantifying participants' handedness. However, the way you calculate the handedness index from the pegboard task is different from the other two measures (Edinburgh Handedness Inventory and the Hand Preference task), and even of the opposite sign. This is definitely a matter of taste, but for clarity I wonder if you could consider changing how the pegboard index is calculated to make it directly comparable to the other two tasks (at least by changing the sign so (LH-RH)+(LH+RH)). This would make it easier for readers to understand and might also reduce potential for confusions and errors for experimenters. But as I said, this is just a suggestion.

Author's response:
We thank the recommender for his suggestion. We have accepted and implemented that. The formula is now as follows (page 16, line 18):

"LI = [(LH-RH) / (LH+RH)]*100"

We also changed the following sentence to reflect the interpretation of the results. The sentence reads:

"A positive score will signify superior right hand skill, while a negative score will signify superior left hand skill."