Research question

I can see the paper has already been through several rounds of revision; I’m not sure whether the earlier rounds are visible – am new to PCI and didn’t see them, but that may be as well as I can take a completely fresh look.

I’m sorry to raise a pretty fundamental question at this point, but I think we need a better rationale for the study. Currently it has rather a so-what feel to it. I feel it could be improved to make it more compelling, and so will share my specific thoughts on how it might be improved.

We are told: “Studying writing using a smartphone is important as more than 6 billion people owned a smartphone in 2021 and spent 2.5 hours using it per day.”

But I can’t grasp why the lateralisation of smartphone writing matters. There’s also an ancillary question about lateralisation in left- and right-handers, which I think has more merit, but it is not well integrated with the rest of the rationale.

Suppose the first hypothesis were disconfirmed and it was found that smartphone writing was either more or less lateralised than handwriting. Would that affect how we theorise about cerebral lateralisation? If the study found that lateralisation was greater or less for smartphone writing than for handwriting, I am not sure how we’d interpret that, given the particular features of smartphone writing that differentiate the tasks.

Task analysis

The potential for confounds is emphasised by the way in which ‘smartphone writing’ is defined in the study: using the thumb of one hand to select letters. First, this means you have an inevitable confound with left- and right-handers using different thumbs. Second, I wondered how many people would be familiar with typing that way. Do you have to turn the phone to landscape direction or change settings to use it effectively? Don’t many people who use thumbs use both thumbs? Essentially, if this is an unfamiliar approach to writing then it seems to introduce a confound, in terms of the need to learn how to do it. And, if the subjects are used to writing on a phone with a different method, then they would have to suppress their natural tendencies.

Another factor is that in conventional oral Word Generation, the words are generated during a silent interval, during which laterality is measured, whereas it seems like with writing tasks envisaged here, the writing starts immediately after the cue. So it’s hard to know how far any difference seen in a written condition is just because the words are written immediately.

Pondering on these issues made me wonder why, if the main interest is in studying laterality in nonvocal word generation/spelling, the authors hadn’t gone for the option of testing people who could type on a regular keyboard. The advantage of typing (assuming you select subjects who can touch type using all fingers) is that both left- and right-handers use both hands – presumably to a similar extent, given keyboard design

Essentially, the possible tasks can be broken down into stages as follows:

**Conventional word generation:** generate words, speak them after a delay.

**Handwriting**: generate words, convert word into letters in correct sequence (spelling), activate motor program to create written representations. This will inevitably introduce confound with handedness, given the movement of one hand will increase blood flow to the contralateral side. And indeed, if I’ve understood it correctly, in their preprint, Papadatou-Pastou did find R-sided laterality in left-handers on this task – though less extreme than for a nonverbal motor control task.

**Smartphone with one thumb**: generate words, convert word into letters in correct sequence (spelling), use one thumb to select letters from keyboard. This will also introduce confound with handedness. It may also introduce confound with task novelty.

**Typing on keyboard**: generate words, convert word into letters in correct sequence (spelling), use both hands to type response on keyboard. This avoids confound with handedness. If touch typists are recruited, task novelty should not be a factor.

(You could also use both thumbs used for smartphone keyboard in landscape orientation, but that may not be familiar for all users, so again there is a novelty confound).

So if you wanted to see how far word generation in a non-spoken modality was similar to conventional spoken word generation, then typing would provide a method that did not need a subtracted condition to control for lateralised hand movement. I realise that is not the question the authors have started with, but I’d personally be intrigued to see if those tasks were equivalent, and whether there was greater dissociation between tasks in left-handers. Compared with what is proposed by the authors, there would be far fewer confounds and so it could be easier to interpret in terms of brain organisational differences in left- and right-handers – which I do think is an important and tractable question.

So I think there would be value in a direct comparison of oral word generation and typed words on a keyboard, though you would need to modify the oral task to start immediately after the cue so that the timing of the two tasks is comparable. There is precedent for this: Gutierrez-Sigut, E., Payne, H., & MacSweeney, M. (2015). Investigating language lateralization during phonological and semantic fluency tasks using functional transcranial Doppler sonography. Laterality, 20(1), 49–68. https://doi.org/10.1080/1357650X.2014.914950

Reliability and positive controls

There’s one key problem with the design as it stands. There is a prediction of a null result. There are two reasons why two measures may not differ statistically: one is because they really are equivalent, and the other is because they are so unreliable that genuine differences are masked. I’m not all that familiar with Bayes factors, but I don’t think they really solve that issue. Minimally, I think we need to have some planned analyses that consider the reliability of tasks. A classic approach to this is Bland-Altman analysis (see references). Although I have used that in some studies, I find the visual representation of information confusing, and think the same point can be made using scatterplots. We had exactly this criticism In our work on bilingualism, where a reviewer argued that a lack of difference between Lis in two languages could be just down to poor reliability, and we were able to counter that argument by reporting the data as scatterplots with individuals showing standard errors. See figure 3 here: <https://wellcomeopenresearch.org/articles/1-15> for an example. I also recommend plotting the time course of the blood flow velocity (see figure 2 from the same paper), as this can establish whether the maximum differences between L and R in different tasks are occurring at a similar time point.

In fact, if you were to move away from the approach that requires control tasks, that provides an opportunity for an even stronger test, because you could just have two blocks for each of two tasks, and so you’d have an index of within-task test-retest reliability. The correlation between block 1 and block 2 for the same task establishes a lower bound for the cross-task agreement. And in effect this acts as a positive control – i.e. if you had low correlations for two runs of the same task, this would indicate that the methods were not suitable to show the effects of interest. This is a realistic concern, because there will be motor movements during the interval of word generation, and these could make the signal noisier than it would be with conventional oral WG.

Predictions re handedness groups

I think there is real interest in questions about whether the pattern of lateralisation differs between left- and right-handers, which was one question looked at in a recent study by Parker et al (2022). The extension to a task involving writing does give this extra interest; I don’t think that has been properly looked at previously. However, the sample size of 16 L handers and 16 R handers does seem too small to demonstrate a handedness effect on Word Generation, let alone a differential handedness effect on another task.

This is output from R’s pwr package, using means from Parker et al on Word Generation to estimate the true effect:

n = 16

delta = 0.86

sd = 1.8

sig.level = 0.05

power = 0.3729685

alternative = one.sided

NOTE: n is number in \*each\* group

So there is only a 1/3 chance of showing a handedness effect on the well-established WG task.

Following Mazoyer et al, it may be possible to increase power by focusing on very strong left-handers – according to Mazoyer et al, they are more likely to have atypical laterality, so this approach could amplify the effect of interest.

This brings us to the next topic, the analysis.

Statistical analysis

In effect, it looks as if the predictions would fit into a 2 x 2 within x between Anova design, with task as a within-subject variable and handedness between-subjects. This may be a rather old-fashioned approach these days when everyone is using linear mixed models. But I think it’s worth considering something along these lines, whether implemented as Anova or something more sophisticated, because the significance of the interaction term would be testing the key prediction of more dissociation between lateralities for left-handers.

A practical problem for casting the analysis this way is that if we focus on the interaction between handedness and task, then the issues regarding power become even more serious: unfortunately interaction terms tend to require much larger sample sizes than main effects.

Overview

I think there is the core of a good project in this paper, but the research question needs tightening up, and, unfortunately, I think that a much larger sample size will be needed to answer it convincingly.

References

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, 8, 307–310.

Parker, A. J., Woodhead, Z. V. J., Carey, D. P., Groen, M. A., Gutierrez-Sigut, E., Hodgson, J., Hudson, J., Karlsson, E. M., MacSweeney, M., Payne, H., Simpson, N., Thompson, P. A., Watkins, K. E., Egan, C., Grant, J. H., Harte, S., Hudson, B. T., Sablik, M., Badcock, N. A., & Bishop, D. V. M. (2022). Inconsistent language lateralisation – Testing the dissociable language laterality hypothesis using behaviour and lateralised cerebral blood flow. *Cortex*, *154*, 105–134. <https://doi.org/10.1016/j.cortex.2022.05.013>