Review by April Shi Min Ching

Is CPP an ERP marker of evidence accumulation in perceptual decision-making? A multiverse study.

Yikang Liu, Chuyao Yan, and Hu Chuan-Peng

Overall comment: The authors propose an analysis of four existing EEG datasets, testing the fit of centro-parietal positivity (CPP) and behavioural data with a drift diffusion model, and also examining the effectiveness of different CPP quantification and pooling methods via multiverse analyses. The literature review and study rationale are presented clearly and succinctly, pointing to a need to examine the link between CPP and evidence accumulation across a series of more complex perceptual decision-making tasks. The chosen methods for CPP and behaviour modelling are state-of-the-art and appropriate. However, I am unsure about the multiverse analysis aspect and the main hypothesis/analysis plan, which in my opinion require more specification (I mainly discuss these issues in the rest of this review).

The point of a multiverse analysis is to examine and compare alternative analytic choices, and these alternatives ought to be sensible or defensible. In the context of this specific topic – the relationship between CPP and evidence accumulation – I think there is little reason to consider CPP quantification approaches (i.e., the other 8 pipelines) other than the combination of CPP build-up rate and trial-wise pooling. Unless the authors are able to justify the other inclusions, I think that the proposed multiverse analysis (or at least some of the pipeline choices, such as CPP amplitude and bin-wise pooling) is not well motivated.

I also think the analysis plan for Hypothesis 1 needs improvement. Here, the link between CPP buildup and drift rate is tested *separately* for each dataset – if the authors want to comment on whether evidence accumulation models are applicable to a wide variety of more complex perceptual tasks, the analysis should include a statistical comparison of equivalent measures from the different datasets.

As alternative analysis plans, I have two suggestions: (1) a "simpler" meta-analytical approach, i.e., comparison of measures of effect sizes (or equivalent) for each dataset/experiment. (2) A multiverse analysis in which the different datasets/tasks (and possibly alternative models) are "universes" – the question being asked here is whether the various paradigms produce an equivalent CPP that is an indicator of evidence accumulation.

Minor comments:

The third author's first/last name order is inconsistent.

Pg 13 "(6) Calculate CPP": Description could be more detailed – I assume "fitting a linear trend" is a least-squares linear regression, done for each participant and trial?

Pg 19, Fig 4A: I am assuming that this part described in the caption was not completed: "Each column represents different measurements, and each row corresponds to different pooling methods" – is the intended plot something like fig S1?

Pg 26-end: References for supplementary methods were not provided.