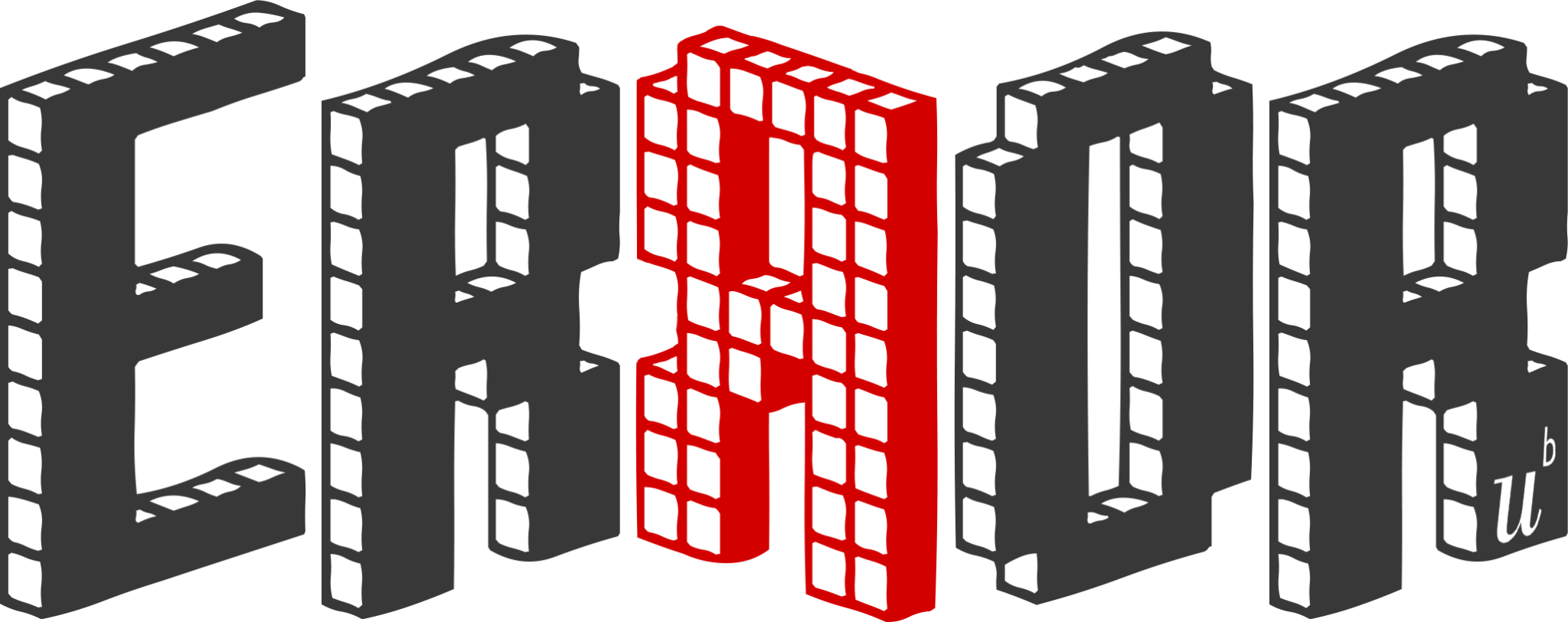
ESTIMATING THE RELIABILITY & ROBUSTNESS OF RESEARCH

ERROR RECOMMENDER REPORT

Wessel, J. (2017). Prepotent motor activity and inhibitory control demands in different variants of the go/no-go paradigm. *Psychophysiology*. doi: [10.1111/psyp.12871](https://doi.org/10.1111/psyp.12871)

DECISION: Minor errors

*Recommender*

**Ian Hussey**, University of Bern

29 Apr 2024

*Recommendation template version 1.0*

As the first recommendation issued by the ERROR project, I’d like to especially Prof Poldrack (reviewer) and Prof Wessel (author) for serving as the project’s first completed pair of error review and author response. These two documents have set an exceptionally high bar for all subsequent error reviews to follow. The substance and style of these reports embody everything we hope to see more of in academic research: acceptance of the possibility that errors occur; inspection for errors that is well-documented and verifiable; and acknowledgement and suitable correction when errors are found.

I should note that unlike editorial decision letters in pre-publication peer review, which are often private to the author and reviewers, ERROR recommendation letters are public documents whose function is to (1) communicate the presence or absence of any errors detected, (2) consider their severity, and (3) provide discussion of how similar errors elsewhere might be prevented or detected.

**Decision & recommendations**

Based on the reviewer’s report and the author’s response and their associated materials, I am returning the decision that the original article contains **Minor Errors**. That is, errors that have the benefit of being detectable thanks to the presence and sharing of research materials, but whose scope and implications are minor. The detected errors do not rise to the level where I would recommend that a correction be issued. However, I would like to thank Prof Wessel for indicating his willingness to issue such a correction. Authors can of course pursue a correction unilaterally if they choose to do so.

Following the ERROR project’s emergent guidelines, the recommendations associated with a minor errors decision are as follows:

* The report, author response, and recommendation will be posted on the ERROR website ([error.reviews](https://error.reviews/)).
* The author is asked to recognise these errors in future discussions of the article.

**Summary of errors detected & how they could be prevented in future**

Meta-science studies that extract details from published studies (e.g., meta-analyses and meta-methods studies) may be more prone to errors that is currently widely acknowledged. This has been shown to be the case in meta-analyses (e.g., Maassen et al., 2020: [10.1371/journal.pone.0233107](https://doi.org/10.1371/journal.pone.0233107)). The current review suggests that it may also be the case for meta-methods reviews. The author’s simulation suggests that even after this ERROR review and his research group’s careful re-extraction of data from all articles in the sample, the probability that errors remain in the dataset is 96%. From the author’s data in their response, errors by individual researchers seem difficult to avoid and are not clearly reduced or eliminated by reducing their individual workload or potentiating them towards finding errors.

Researchers should treat these extracted details as estimates of the original details that contain measurement error. As such, steps should be taken to quantify and reduce this error. At least two independent raters should be used to extract these details. The inter-rater reliability should be reported in the article along with the resolution strategy, and some consideration should be given to the prevalence of errors remaining in the data. Even with two raters, additional undetected errors will likely still be present (e.g. where the raters agree but are both wrong) and indeed prevalent. Naturally, the extracted data should be publicly available to allow for verifiability.

**Discussion of individual issues raised**

**Paper Selection for the literature analysis**

The reviewer and author agree about the existence and scope of this issue. I too agree that changes in the search terms may have produce a different set of results, therefore potentially impacting the generalisability of the results. I also agree that the likely magnitude of this is quite small, and that it does not rise to the level of an error in the original work.

**Extraction of the key parameters from the papers**

The author response notes that the original code was indeed available, contrary to what was stated in the reviewer’s report. Here, we the ERROR organising team, must acknowledge our own error: we did not sufficiently communicate the fact that the reviewer could directly ask the author for such materials as part of the ERROR review process, whether they are already publicly available online or not. We have updated our instructions for future reviewers to strengthen reviewers’ expectations that can and should ask the authors directly for materials, such as data and code, or aid with interpreting and understanding these materials, in the course of their review. Reviewers will be more explicitly instructed that the authors have consented to sharing such research materials beyond what is necessarily in the publication; that they can and should ask for these materials; and that they can ask appropriate questions about details where needed. We prefer that reviewers obtain and discuss materials directly with the authors (with the ERROR team in cc) rather than materials being transferred via us in order to (a) lower the internal administrative burden for the ERROR team, and (b) to make ERROR reviews a more collaborative interaction between authors and reviewers.

The reviewer and author agree about the existence and scope of this issue, and the author determined the reason for it (i.e., the adding of an offset value to aid plotting, whose offset values were then erroneously reported in text rather than the unadjusted values). I am in agreement that this represents a minor error. I would like to thank Prof Wessel for taking the time to retrace his steps to consider how the error occurred and indeed how errors like it may be prevented in future.

**Parameter Analysis**

Three of the 24 articles that Prof Poldrack re-extracted contained values different to those reported in Prof Wessel’s article (a fourth was identified too but is, as Prof Wessel points out, correct when rounded).

I was grateful that Prof Wessel and his research groups were willing to take the time to re-extract data from the other 90% of the articles. The discussion in the author’s response of comparable error rates being found even among new raters who were potentiated towards finding errors is a useful finding for future research. As noted in the author’s response, ratings by additional individuals appear to be key to detecting errors prior to analysis. The results of the author’s simulation study suggest that the probability that, even after being scored a second time, the data still contains one or more errors is 96%. This underscores the important point that error-reviewed articles cannot now be assumed to be error-free, only that they are likely to contain fewer errors, or that the degree to which errors can be checked for has been determined. I encourage readers of this recommendation to read Prof Wessel’s report in full.

**Unresolved issues**

It is reasonable to expect that ERROR reviews will leave some questions unresolved. It is useful to acknowledge some of these issues so that ERROR reviews do not artificially convey that they are the final word on issues of error detection and correction. In this case, it is a difference in the results between the author’s simulation study code and my attempt to reimplement it in R.

I have never worked in MatLab, therefore I attempted to convert the author’s simulation code to R with the help of ChatGPT. This was a useful exercise, as it allowed us to discover and fix an error in the original code. While my R code and the author’s MatLab code produce similar estimates of the probability of there being remaining errors in the dataset (both 96%), we get different estimates for the probability that reviewers make the same error (my R code: 72%; the author’s MatLab code: 51%). Neither of us can discern the reason for this difference, neither of us being experienced in the other language. I am happy to defer to the author’s estimate here.

I would like to thank both Prof Poldrack and Prof Wessel again for their efforts and error-acceptance here.

Ian Hussey