



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

Report version 1.0 (2024-05-07)

Recommendation template version 1.0

License: CC BY 4.0

As the first recommendation issued by the ERROR project, I'd like to especially thank 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. I would like to especially acknowledge the author, Prof Wessel, for going above and beyond in his author response by reextracting, with the aid of members of his research group, the whole dataset again to comprehensively check for errors not already found by the reviewer, and by constructing a simulation study to attempt to understand the probability of there still being additional undetected errors still in the data. 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](#)).
- 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

The original article conducted a meta-methods review of the task parameters used for go/no-go tasks and observed that only a minority of articles in the literature employ what the author argues to be the necessary parameters (ie rare no-go trials and short trial

durations). New data collection was then used to experimentally manipulate these task parameters, showing that go/no-go tasks only elicit reliable prepotent motor activity when the task is fast paced and when no-go trials are rare.

We asked the reviewer to focus their review on the first section – the meta-methods review of go/no-go task parameters employed in the literature.

Research elsewhere has shown that 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 raised the point regarding the Boolean search terms used for the review: while not erroneous, the choice of terms may not have been sufficient to identify all possibly relevant papers. Suggested alterations to the search query returned 14% more results than the original search. The reviewer noted however that it is unlikely that it is highly unlikely that additional articles would have changed the results systematically. The author agrees with the existence and scope of this issue, as do I. This is a useful question to have explored, but does not rise to the level of an error.

Extraction and analysis of go/no-go task parameters

The reviewer re-extracted the task parameters for 24 (10%) of the articles employed in the original review. I will focus on the P(nogo) parameter here, as the parameter argued

to be more important. Three of the 24 articles (12.5%) that the reviewer re-extracted contained values different to those reported in the article (a fourth was identified too but as the author points out, is correct when rounded).

I was very 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. This was done unprompted by us or the reviewer, out of a desire to understand the actual error rate in the data once initial errors had been found by the reviewer. This re-extraction uncovered eight (3.3%) additional errors in the values originally extracted. Additionally, 4 new errors were made where, on inspection, the original value appeared to be correct (Cohen's Kappa between original and new ratings = .958). The author provided a useful assessment of bias in the original ratings that suggests that errors were evenly distributed above and below the correct values, therefore these errors represented noise but not systematic bias.

The author's response provides a very useful discussion of the comparable error rates found even among new raters who were potentiated towards finding errors. This 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. I thank the author for taking the time to write a simulation study to estimate the probability that, even after being scored a second time, the data still contains one or more errors, and for being receptive to my feedback in finding and correcting bugs in this simulation. Like the author, I was surprised to learn that the probability of the dataset still containing at least one error 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 author response in full. The reviewer and author agree about the existence and scope of these data extraction issues. I also agree that this represents a minor error.

Reproduction of Figures 1 and 2

As part of attempting to reproduce the plots that illustrate the results of the meta-methods review of go/no-go task parameters, the reviewer noted that the original code was not available (although the data was linked in the article). The author's response notes that the original code was in fact available on their website. 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's reconstruction of the plots identified that the original figures included a small offset to the observed data to add plotting, and that this offset was erroneously brought into the reported results too. That is, the offset data used for plotting was also used to report results, rather than the actual data. I agree that this does constitute an error, and was well-spotted by the reviewer. The author agrees with the existence and scope of this error, and I agree that its magnitude is minor and it does not change the interpretation of the results. I sympathise with the author's frustration that this error was nonetheless made, and yet I think it provides a relatable and accessible example of how an erroneous workflow can give rise to erroneous results. The authors response provides a clear description of the workflow and how the error was likely made.

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 this or future ERROR recommendations do not artificially convey that they represent the final word on issues of error detection and correction for a given article. In this case, it is a difference in the results between the author's simulation study code, written as part of the author response, 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 could discern the reason for this difference in a reasonable amount of time, neither of us being experienced in the other language. I am happy to defer to the author's estimate here.

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

Ian Hussey