To Err is Human: An Empirical Investigation

Daniel Lakens & Lisa DeBruine

2024-06-21

This paper demonstrates some good and poor practices for use with the {metacheck} R package and Shiny app. All data are simulated. The paper shows examples of (1) open and closed OSF links; (2a) citation of retracted papers, (2b) citations without a doi, (2c) citations with Pubpeer comments, (2d) citations in the FORTT replication database, and (2e) missing/mismatched/incorrect citations and references; (3a) R files with code on GitHub that do not load libraries in one location, (3b) load files that are not shared in the repository, (3c) lack comments, and (3d) have hard-coded files, (4) imprecise reporting of non-significant p-values; (5) tests with and without effect sizes, and (6) use of "marginally significant" to describe non-significant findings.

Introduction

Although intentional dishonestly might be a successful way to boost creativity (Gino & Wiltermuth, 2014), it is safe to say most mistakes researchers make are unintentional. From a human factors perspective, human error is a symptom of a poor design (Smithy, 2020). Automation can be use to check for errors in scientific manuscripts, and inform authors about possible corrections. In this study we examine the usefulness of metacheck to improve best practices.

Method and Participants

In this study we examine whether automated checks reduce the amount of errors that researchers make in scientific manuscripts. This study was preregistered at osf.io/5tbm9. We randomly assigned 50 scientists to a condition where their manuscript was automatically checked for errors, an 50 scientists to a control condition with a checklist. Scientists had the opportunity to make changes to their manuscript based on the feedback of the tool. We subsequently coded all manuscripts for mistakes, and counted the total number of mistakes. We also measured the expertise of researchers (in years) to explore whether the automated tool would be more useful, the less research experience researchers had. We also asked researchers to rate how useful they found the checklist or app on a scale from 1 (not at all) to 7 (extremely useful). Data and analysis code is available from https://github.com/Lakens/to_err_is_human.
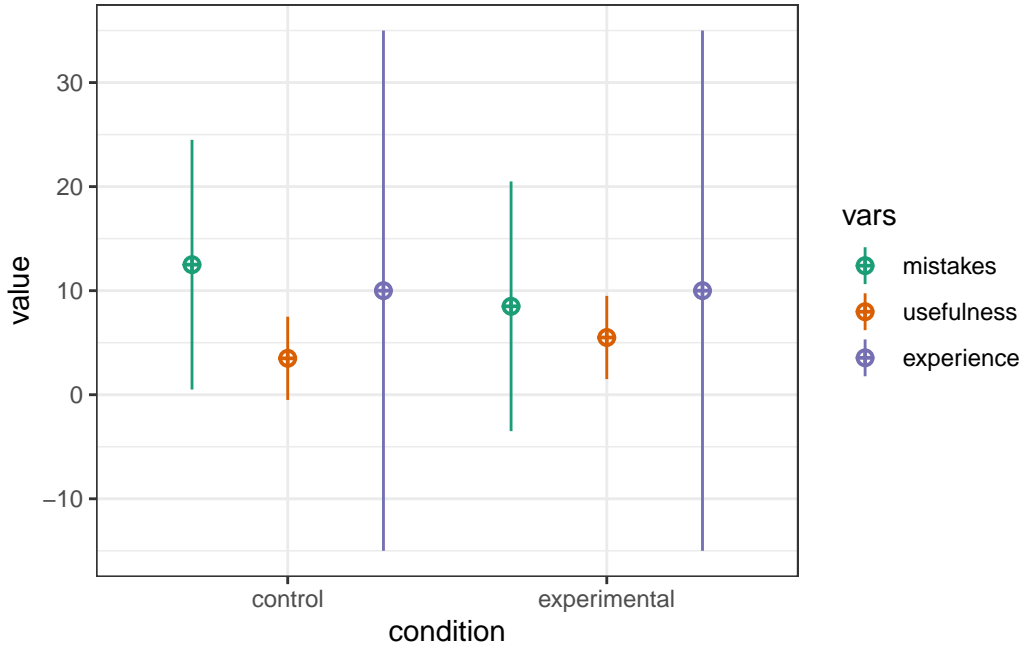
Results

Figure 1: The simulated data

All data needed to reproduce the analyses in Table 1 is available from https://osf.io/5tbm9 and code is available from the OSF.
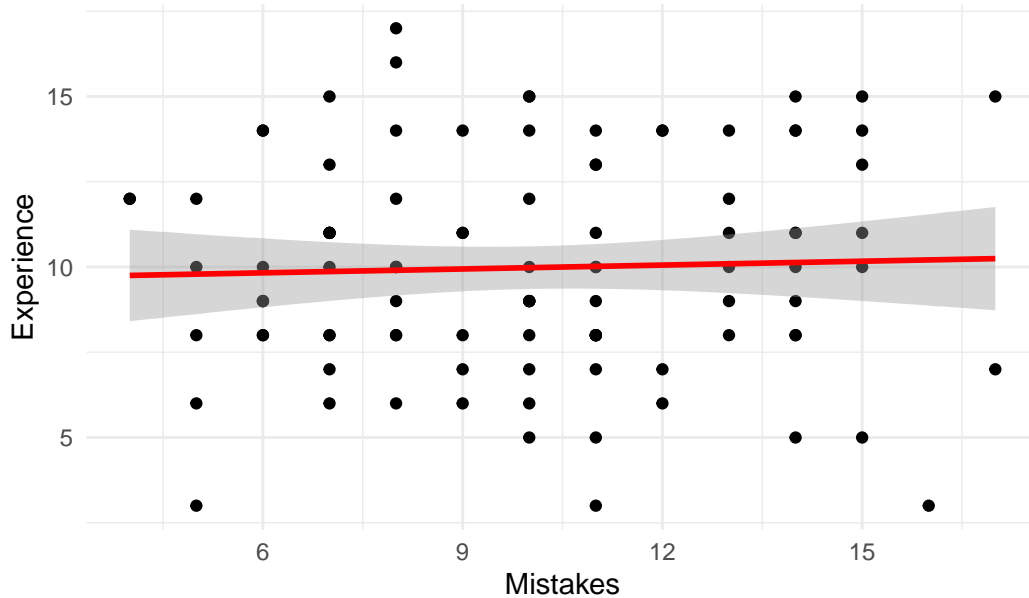
Table 1: The average number of mistakes and usefulness score for the control and experimental conditions.

Condition	Mistakes	Usefulness
control	10.90	4.50
experimental	9.12	5.06

On average researchers in the experimental (app) condition made fewer mistakes ($M = 9.12$) than researchers in the control (checklist) condition ($M = 10.9$), $t(97.7) = 2.9$, $p = 0.005$, d = 0.59.

On average researchers in the experimental condition found the app marginally significantly more useful ($M = 5.06$) than researchers in the control condition found the checklist ($M = 4.5$), $t(97.2) = -1.96$, $p = 0.152$.

Scatter Plot with Correlation: 0.04

There was no effect of experience on the reduction in errors when using the tool ($p > .05$), as the correlation was non-significant.

Discussion

It seems automated tools can help prevent errors by providing researchers with feedback about potential mistakes, and researchers feel the app is useful. We conclude the use of automated checks has potential to reduce the number of mistakes in scientific manuscripts.

References

Gangestad, S. W., & Thornhill, R. (1998). Menstrual cycle variation in women's preferences for the scent of symmetrical men. Proceedings Biological Sciences, 22, 927-933. doi: 10.1098/rspb.1998.0380.

Gino, F., & Wiltermuth, S. S. (2014). Evil Genius? How Dishonesty Can Lead to Greater Creativity. Psychological Science, 25(4), 973–981. https://doi.org/10.1177/0956797614520714

Smith, F. (2021). Human error is a symptom of a poor design. Journal of Journals, 0(0), 0. https://doi.org/10.0000/0123456789

Lakens, D. (2018). Equivalence testing for psychological research. Advances in Methods and Practices in Psychological Science, 1, 259-270.