Prioritizing computational reproducibility in behavioral science

Hauke Roggenkamp
Susanne Adler
Michael V. Reiss
Stefan Feuerriegel
Nicolas Pröllochs
Claire Robertson
Felix Holzmeister

July 3, 2025

Standfirst (in lieu of an abstract) goes here (up to 360 characters, including spaces).

The behavioral sciences have demonstrated remarkable progress in response to credibility challenges over the past decade (Open Science Collaboration 2015; Camerer et al. 2016, 2018; Klein et al. 2018). Specifically, the open science movement, pre-registration, and crowd-sourced science have transformed research practices and strengthened methodological rigor. One of these crowd-sourced collaborations is featured in this issue and reports that x percent of published findings replicate/are reproducible etc. (see the I4Rs meta-paper).

Whereas replicability describes the ability to obtain consistent results using new data, *reproducibility* is a more fundamental criterion that comes in two complementary forms (Dreber and Johannesson 2025). Computational reproducibility describes the extent to which one can recreate numerically identical results using the same data and code as the original authors. As such, it verifies the mapping between data, code, and reported results. Recreate reproducibility ignores the original authors' code and describes the extent to which one can recreate numerically identical results using the same data and only the published methodological descriptions. As such, it verifies the mapping between reported methods and reported results, catching gaps in methodological reporting or coding errors that computational reproducibility may miss.

More fundamentally, the critical standard is whether not whether the results but the *conclusions and claims* remain consistent. Rather than adopting a dichotomous pass/fail approach (analogous to the widely applied binary treatment of statistical significance) we advocate for viewing reproducibility as existing on a spectrum. This includes assessing the degree of preservation across: (1) the statistical significance and direction of primary results, (2) the substantive magnitude of estimated effects, and (3) the overall interpretation and policy implications.

Hauke Roggenkamp, PhD Candidate in Marketing at the University of St. Gallen, St. Gallen, Switzerland (email: Hauke.Roggenkamp@unisg.ch). Michael V. Reiss, Postdoctoral Researcher in Communication Science at the Leibniz Institute for Media Research, Hamburg, Germany. Susanne Adler, Postdoctoral Researcher in Marketing at the LMU Munich, Munich, Germany. Stefan Feuerriegel, Full Professor at the School of Management and the Faculty of Mathematics, Informatics, and Statistics at LMU Munich, Munich, Germany. Nicolas Pröllochs, Professor of Data Science and Digitization at the University of Gießen, Gießen, Germany. Claire Robertson, Postdoctoral Researcher in Social Psychology at NYU, New York, USA. Felix Holzmeister, Assistant Professor of Behavioral and Experimental Economics and Finance at the University of Innsbruck, Innsbruck, Austria.

When numerical discrepancies do arise, we recommend applying a principle of in *dubio pro reo*: giving authors the benefit