A tutorial on computationally reproducible research

Jordan W. Suchow (jws@stevens.edu)[1]
[1]School of Business, Stevens Institute of Technology

Keywords: reproducibility; computational methods; open science; scientific workflows

Introduction

Science is an exercise in seeing further by standing on the shoulders of giants—but what happens when those shoulders give way? The replication crisis has shaken confidence in published findings, with surveys revealing that more than 70% of researchers have failed to replicate another scientist's experiments (Baker, 2016). Even when researchers attempt the seemingly simpler task of *computational* reproducibility—applying the same analysis to the same data—success is far from guaranteed. A recent large-scale investigation found that only 52.6% of social and behavioral science papers could be precisely reproduced, even when original data were available (Miske et al., 2026). The costs are substantial: wasted resources, delayed discoveries, and eroded public trust in science.

Cognitive science faces particular challenges: neuroimaging pipelines with deep dependency stacks, behavioral experiments requiring millisecond timing precision, computational models with stochastic optimization, and statistical analyses spanning multiple languages. Yet as the National Academies note, while computation has transformed scientific research, "this revolution is not yet uniformly reflected in how scientists develop and use software and how scientific results are published and shared" (National Academies of Sciences, Engineering, and Medicine, 2019). The traditional methods section is insufficient to convey the information needed for computational reproducibility; the actual scholarship is the complete computational environment and the instructions that generated the results. And most researchers lack formal training in the practices that would make their work reproducible by default.

This tutorial equips cognitive scientists with both diagnostic skills and practical tools to make their research computationally reproducible. Participants will learn to identify the common failure modes (such as hidden randomness, numerical instability, dependency drift, implicit state, and fragile workflows) and match them to appropriate fixes, from minimal interventions like explicit seeding and dependency pinning to more robust solutions like containerization and workflow automation. Through hands-on exercises, attendees will diagnose reproducibility issues in example projects and audit their own research. They will leave with a reproducibility rubric, reusable templates, and a personal action plan for improving their computational practices.

Tutorial Overview

This half-day (4-hour) interactive workshop takes a diagnostic approach to computational reproducibility. Rather than prescribing a single toolkit, participants learn to identify failure modes, match appropriate fixes, and develop practical strategies through hands-on exercises. Participants leave with conceptual frameworks and practical checklists for their own workflows. All tutorial materials—slides, exercises, templates, and the reproducibility rubric—will be released openly so the tutorial can serve as a reusable resource beyond the in-person session.

Target Audience

This tutorial is designed for graduate students, postdocs, and faculty across all areas of cognitive science. The content is delivered at a level that is accessible to any researcher in cognitive science, regardless of specific computational background. No prior experience with specific reproducibility tools is required, only basic familiarity with computational research in any programming language. Participants are encouraged to bring a current project of their own to serve as the subject of the Part 4 audit; those who do not will work with a provided example project drawn from cognitive science research.

Learning Objectives

Participants will be able to: (1) distinguish reproducibility from replicability and understand the fuzziness of the boundary; (2) identify common failure modes; (3) evaluate the reproducibility of any research project; (4) navigate the ecosystem of tools including version control, containers, workflow managers, and archives; and (5) develop a practical improvement strategy for their research.

Tutorial Structure

Part 1: Foundations (30 min). Conceptual foundations: reproducibility vs. replicability, when results must be identical vs. meaningfully equivalent (e.g., stochastic models), the current state of reproducibility in the behavioral and social sciences (Hardwicke et al., 2018; Stodden et al., 2018), why failures happen, and why they remain invisible until reproduction is attempted.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

Part 2: Failure Modes & Guided Diagnostics Lab (65 min). The heart of the tutorial. We dissect the taxonomy of reproducibility failures with concrete examples across cognitive science domains:

Behavioral experiments: OS-level timing differences affecting millisecond precision; stimulus randomization without explicit seeds; software version incompatibilities.

Computational modeling: Seed control for randomization; nondeterministic CPU and GPU operations; tolerance-based equivalence for Monte Carlo methods; MCMC convergence diagnostics.

Cross-cutting issues: Environment drift (unpinned dependencies, dictionary ordering, R package updates changing defaults); implicit state (hardcoded paths, timezone artifacts); fragile workflows (undocumented manual steps).

Participants diagnose issues in example projects drawn from real (anonymized) cognitive science research.

Break (10 min).

Part 3: Tools as Interventions (65 min). Rather than surveying tools exhaustively, we present a decision framework matching fixes to failure modes:

Failure Mode	**Minimal Fix**	**Robust Fix**
Randomness	Explicit seeds	Container
Numerical instability	Pin versions	Tolerance tests
Dependency drift	requirements/renv	Docker
Implicit state	Relative paths	Workflow mgr
Fragile workflows	README steps	Automation

We also demonstrate some key tools (Git, Docker, conda/renv, Make/Snakemake, OSF/Zenodo) in the context of this framework.

Break (10 min).

Part 4: Reproducibility Audit & Personal Upgrade Plan (60 min). Participants apply the reproducibility rubric to their own projects, identify failure modes, and develop a concrete action plan with two prioritized improvements. We also distribute templates for environment management (e.g., environment.yml, renv.lock) and workflow automation (e.g., Makefile, one-command run scripts). Session concludes with Q&A and resource sharing.

Technical Requirements

Laptops strongly encouraged, but not required. No other specific technical requirements.

Presenter Information

Jordan W. Suchow is an Assistant Professor at Stevens Institute of Technology. He received his PhD in Psychology from Harvard University and completed postdoctoral training at UC Berkeley. He is a coauthor on recent large-scale reproducibility investigations (Breznau et al., 2025; Miske et al., 2026) and was a participant in the DARPA SCORE program on developing automated methods to assess replicability, reproducibility, and analytic robustness in the behavioral and social sciences.

Acknowledgments

We thank the Center for Open Science and the team behind the DARPA SCORE program.

References

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Breznau, N., et al. (2025). The reliability of replications: A study in computational reproductions. *Royal Society Open Science*, *12*(3), 241038. https://doi.org/10.1098/rsos.241038

Hardwicke, T. E., Mathur, M. B., MacDonald, K., et al. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*, *5*(8), 180448. https://doi.org/10.1098/rsos.180448

Miske, O., et al. (2026). Investigating the reproducibility of the social and behavioural sciences. *Nature*.

National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. The National Academies Press. https://doi.org/10.17226/25303

Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, *115*(11), 2584–2589. https://doi.org/10.1073/pnas.1708290115