

OCTOBER
2019

Is replication research the study of research or of researchers?

ANNETTE N. BROWN

Principal Economist

 [@anbrowndc](https://twitter.com/anbrowndc)

A presentation in two parts

A presentation in two parts

- Is science computationally reproducible – push button replicable?
- Do scientists provide replication files?

- What are replication research ethics?

Push button replication

If I run the same
code on the same
data, I should get
the published
results right?



Seems like
science 101.



freshspectrum.com

Push button replication

If I run the same code on the same data, I should reproduce the published results, right?



Presuming you can get the code and the data...



freshspectrum.com

Example 1: Wicherts, et al. (2006)

DOI: 10.1037/0003-066X.61.7.726

The Poor Availability of Psychological Research Data for Reanalysis

Jelte M. Wicherts, Denny Borsboom,
Judith Kats, and Dylan Molenaar
University of Amsterdam

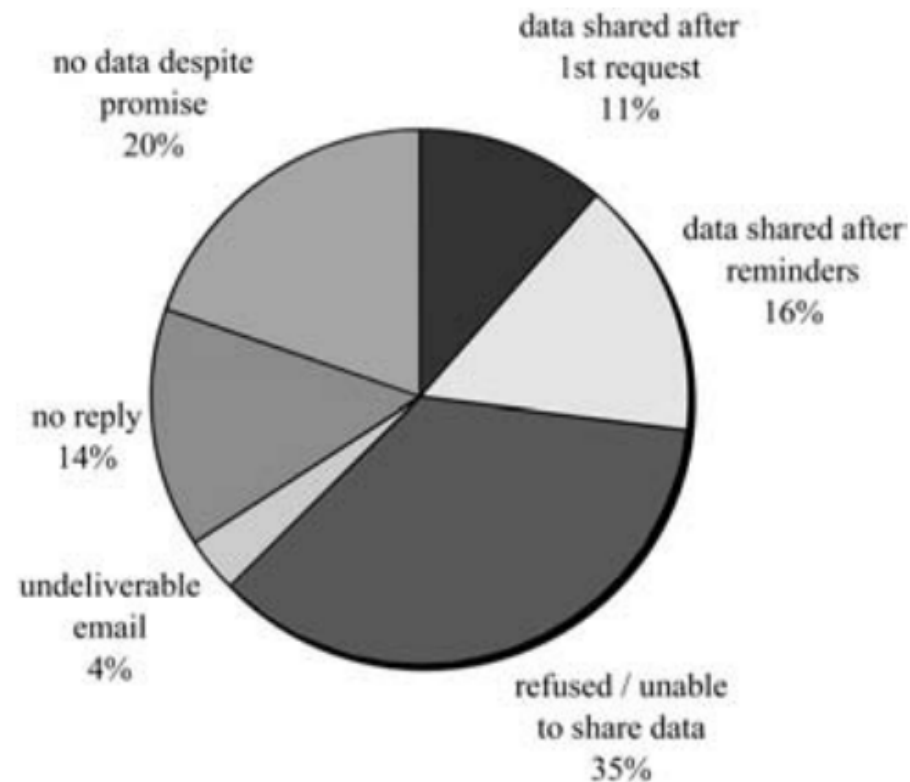
The origin of the present comment lies in a failed attempt to obtain, through e-mailed requests, data reported in 141 empirical articles recently published by the American Psychological Association (APA). Our original aim was to reanalyze these data sets to assess the robustness of the research findings to outliers. We never got that far.

- Original research objective was to assess robustness of research findings to outliers for psychological research
- Sample: 141 articles published by American Psychological Association
- Original authors had all signed Certification of Compliance with APA Ethical Principles, including principle on data sharing for reanalysis
- Emails sent to corresponding authors

Wicherts, JM, Borsboom, D, Kats, J and Molenaar, D. "The poor availability of psychological research data for reanalysis," *American Psychologist*, October 2006.

Wicherts, et al. (2006) findings

Figure 1.
Percentages of Empirical Articles' Corresponding Authors in Different Response Categories



Example 2: Stagge, et al. (2019)

Corrected: Author Correction

Assessing data availability and research reproducibility in hydrology and water resources

James H. Stagge^{1,2}, David E. Rosenberg¹, Adel M. Abdallah^{1,3}, Hadia Akbar¹, Nour A. Attallah¹ & Ryan James¹

There is broad interest to improve the reproducibility of published research. We developed a survey tool to assess the availability of digital research artifacts published alongside peer-reviewed journal articles (e.g. data, models, code, directions for use) and reproducibility of article results. We used the tool to assess 360 of the 1,989 articles published by six hydrology and water resources journals in 2017. Like studies from other fields, we reproduced results for only a small fraction of articles (1.6% of tested articles) using their available artifacts. We estimated, with 95% confidence, that results might be reproduced for only 0.6% to 6.8% of all 1,989 articles. Unlike prior studies, the survey tool identified key bottlenecks to making work more reproducible. Bottlenecks include: only some digital artifacts available (44% of articles), no directions (89%), or all artifacts available but results not reproducible (5%). The tool (or extensions) can help authors, journals, funders, and institutions to self-assess manuscripts, provide feedback to improve reproducibility, and recognize and reward reproducible articles as examples for others.

- Research objective: to develop a survey tool for assessing reproducibility and quantify the “current state of reproducible science in hydrology”
- Sample: 360 random-sampled articles from six hydrology and water resources journals
- Two journals required authors to state how files can be accessed; four journals only encouraged this
- Files accessed online; requirement to contact author or third party = unavailable

Stagge, JH, Rosenberg, DE, Abdallah, AM, Akbar, H, Attallah, NA, and James, R. “Assessing data availability and research reproducibility in hydrology and water resources,” *Nature Scientific Data*, 6:190030, February 2019.

Stagge, et al. (2019) findings

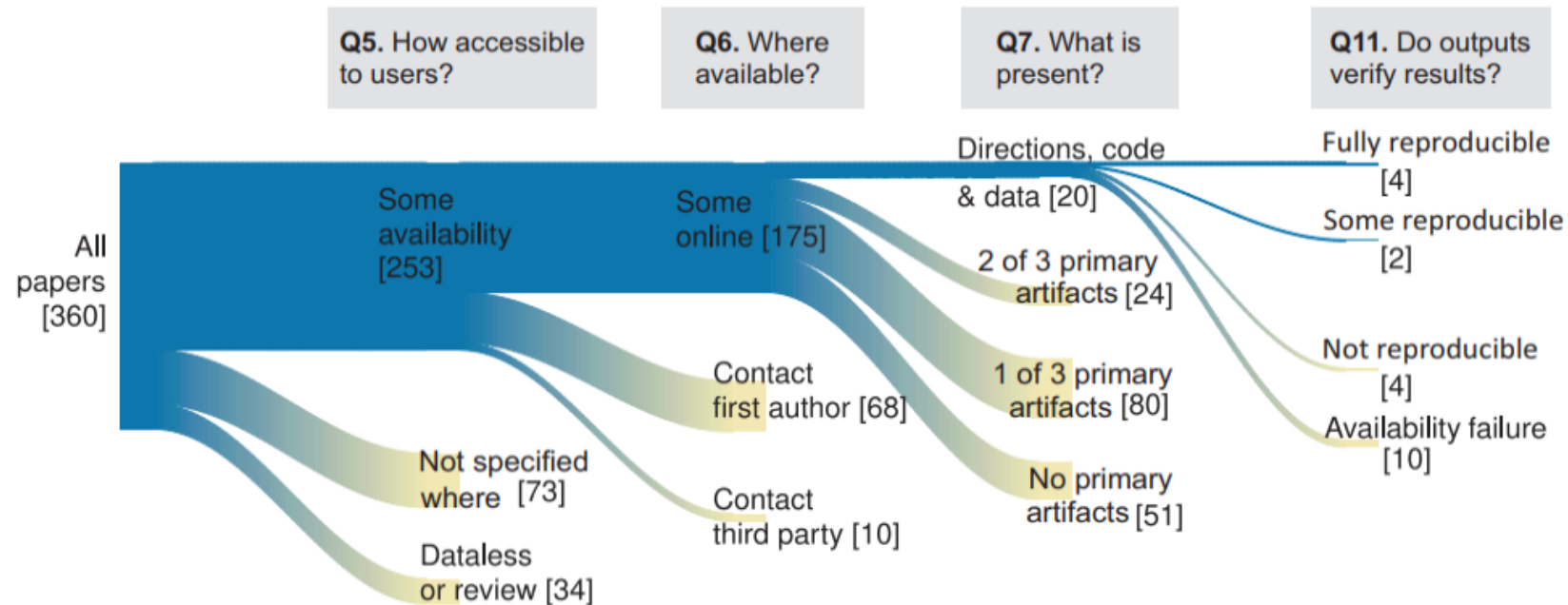


Figure 2. Number of papers progressing through the survey questions to determine availability and reproducibility requirements.

Example 3: Chang and Li (2017)

Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say “Often Not”

Andrew C. Chang* and Phillip Li†

September 26, 2017

Abstract

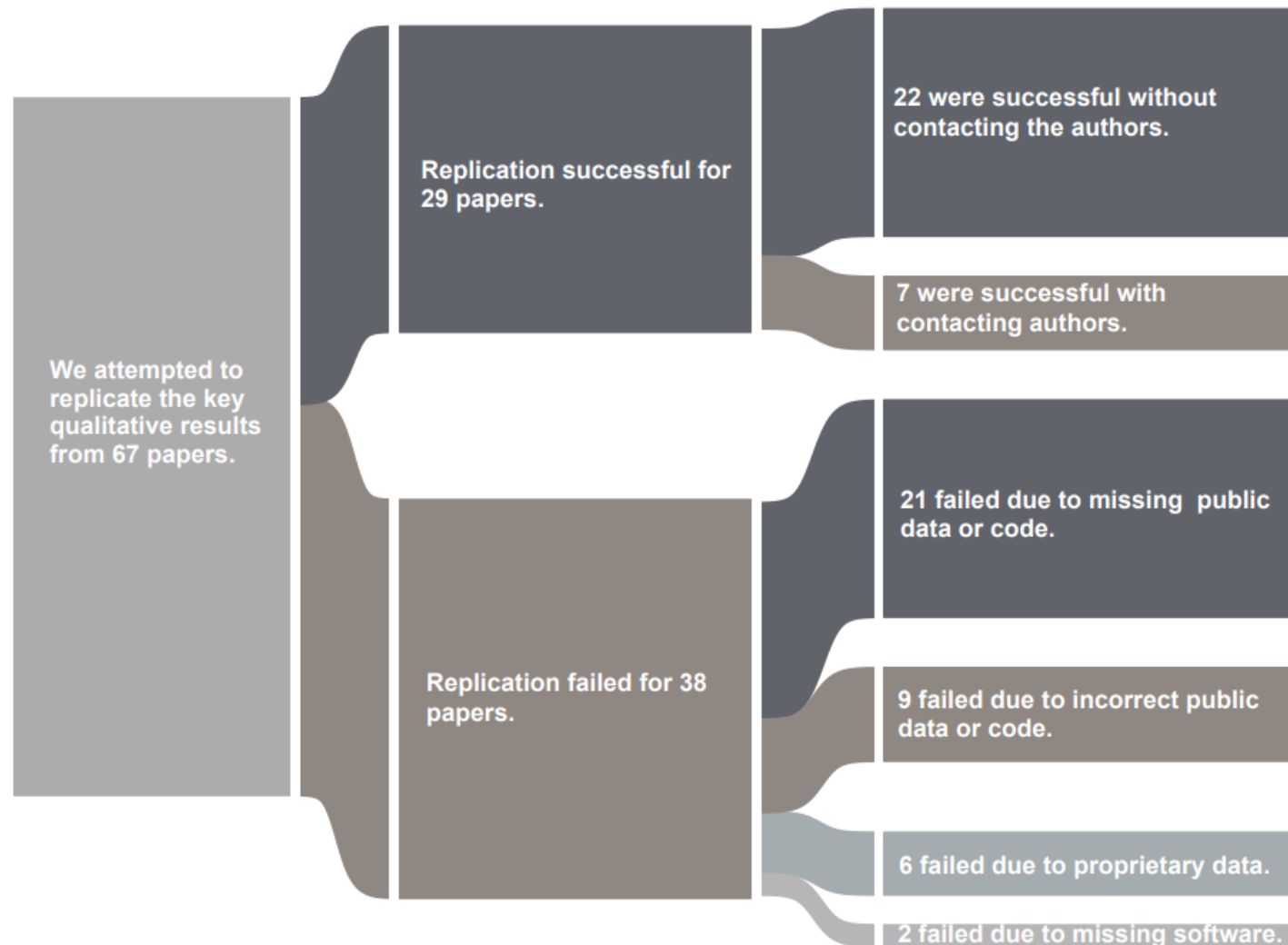
We attempted to replicate 67 macroeconomic papers published in 13 well-regarded economics journals using author-provided replication files that included both data and code by following a preanalysis plan. Aside from 6 papers that used confidential data, we obtained data and code replication files for 29 of 35 papers (83%) that were required to provide such files as a condition of publication, compared to 11 of 26 papers (42%) that were not required to provide data and code replication files. Defining replication success as our ability to use the author-provided data and code files to produce the key qualitative conclusions of the original paper, we successfully replicated 22 of 67 papers

- Research objective: analyze the state of replication in economics
- Sample: 67 empirical macroeconomics articles from 13 journals
- Some articles subject to data availability policy and some not
- Public files accessed first, then requests emailed to authors

Chang, AC and Li, P. “Is economics research replicable? Sixty published papers from 13 journals say ‘often not’” Pre-print accepted for publication at *Critical Finance Review*, November 2017.

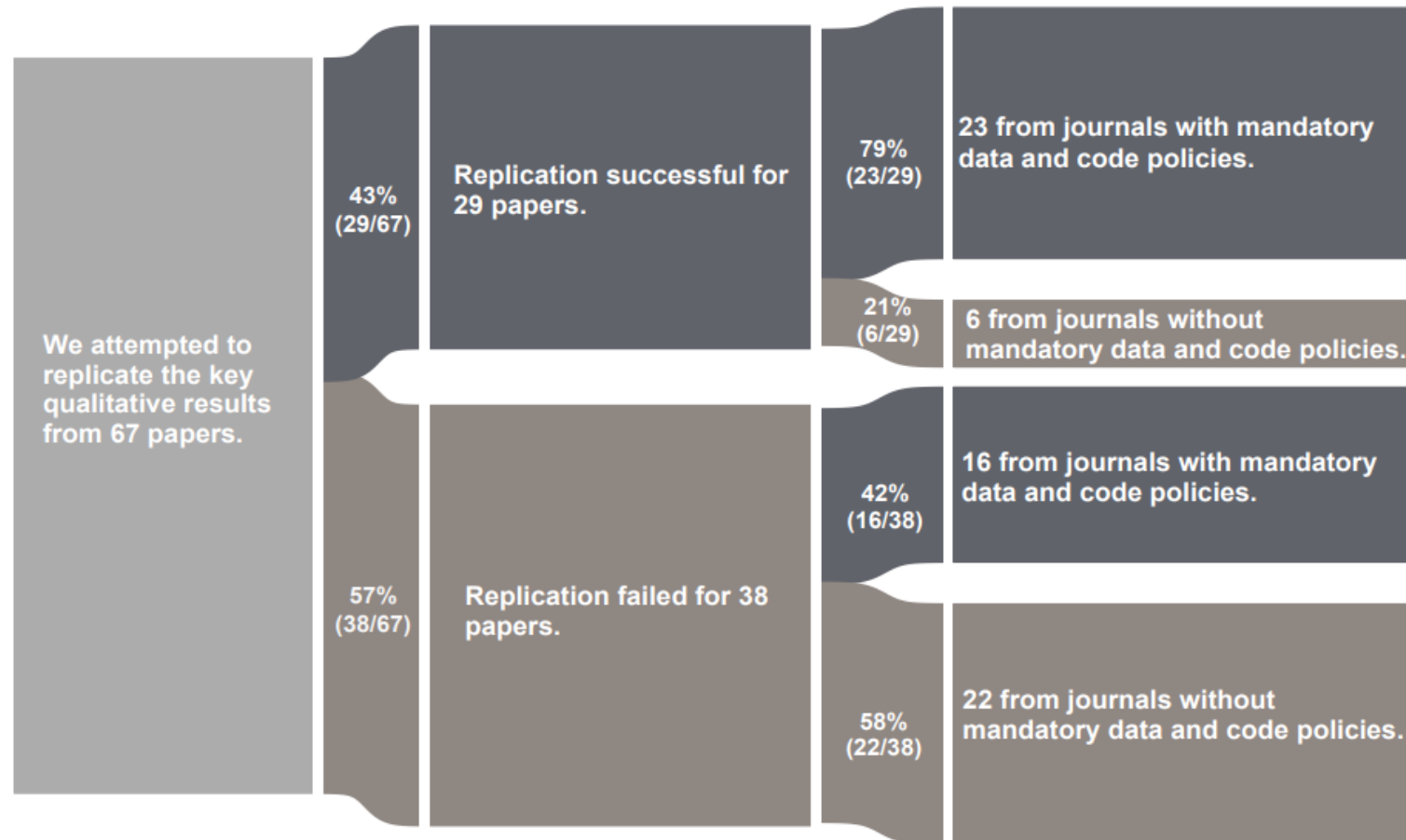
Chang and Li (2017) findings

Figure 1: Causes of Replication Success or Failure



Chang and Li (2017) findings continued

Figure 2: Replication Success or Failure by Journal Type



Our paper: Wood, Müller, and Brown (2018)

RESEARCH ARTICLE

Push button replication: Is impact evaluation evidence for international development verifiable?

Benjamin D. K. Wood^{1*}, Rui Müller², Annette N. Brown³

¹ International Initiative for Impact Evaluation (3ie), Washington, District of Columbia, United States of America, ² Department of Economics, University of Copenhagen, Copenhagen, Denmark, ³ Chief Science Office, FHI 360, Washington, District of Columbia, United States of America

✉ Current address: Monitoring, Evaluation, and Learning Department, Integra LLC, Washington, District of Columbia, United States of America

* bwood@integrallc.com

Wood, BDK, Müller, R and Brown, AN. “Push button replication: Is impact evaluation evidence for international development verifiable?” *PLoS ONE*, 13(12): e0209416.

- Research question: Is impact evaluation evidence for international development verifiable?
- Sample: 109 impact evaluation articles from 10 journals, including health science and social science journals
- One journal had public replication file requirement; two had replication file requirement; two encouraged replication files
- All authors notified, data requested when not public

Wood, et al. (2018) background

- Development impact evaluations defined as studies measuring the effect of an intervention or program in a low- or middle-income country using an experimental or quasi-experimental methodology.
- These studies can be highly policy influential.
- These studies span many academic disciplines and journals.
- Sample based on top ten journals from 2010-2012; sample drawn from 2014.

Wood, et al. (2018) methods

- Protocol
- Classifications
- Key results
- Transparency
 - OSF project site
 - Protocol public
 - All authors notified, even if data public
 - Key results public
 - PBR report accessible by original authors

Wood, et al. (2018) findings

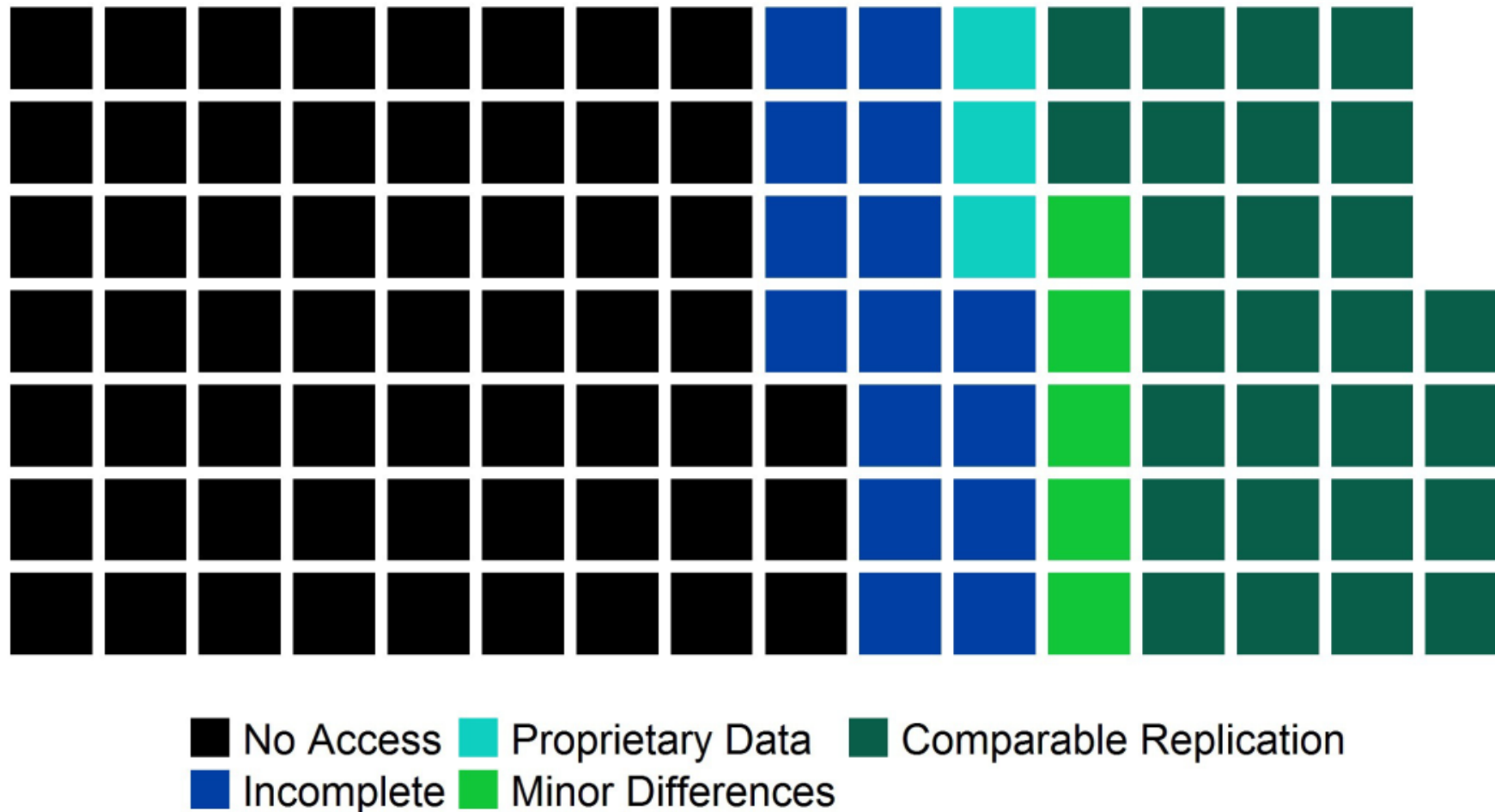


Fig 1. PBR classification results for entire sample. Each square represents one article, and the order of shading from left to right is no access, incomplete, proprietary data, (major differences), minor differences, and comparable.

Wood, et al. (2018) findings – studies with incomplete data

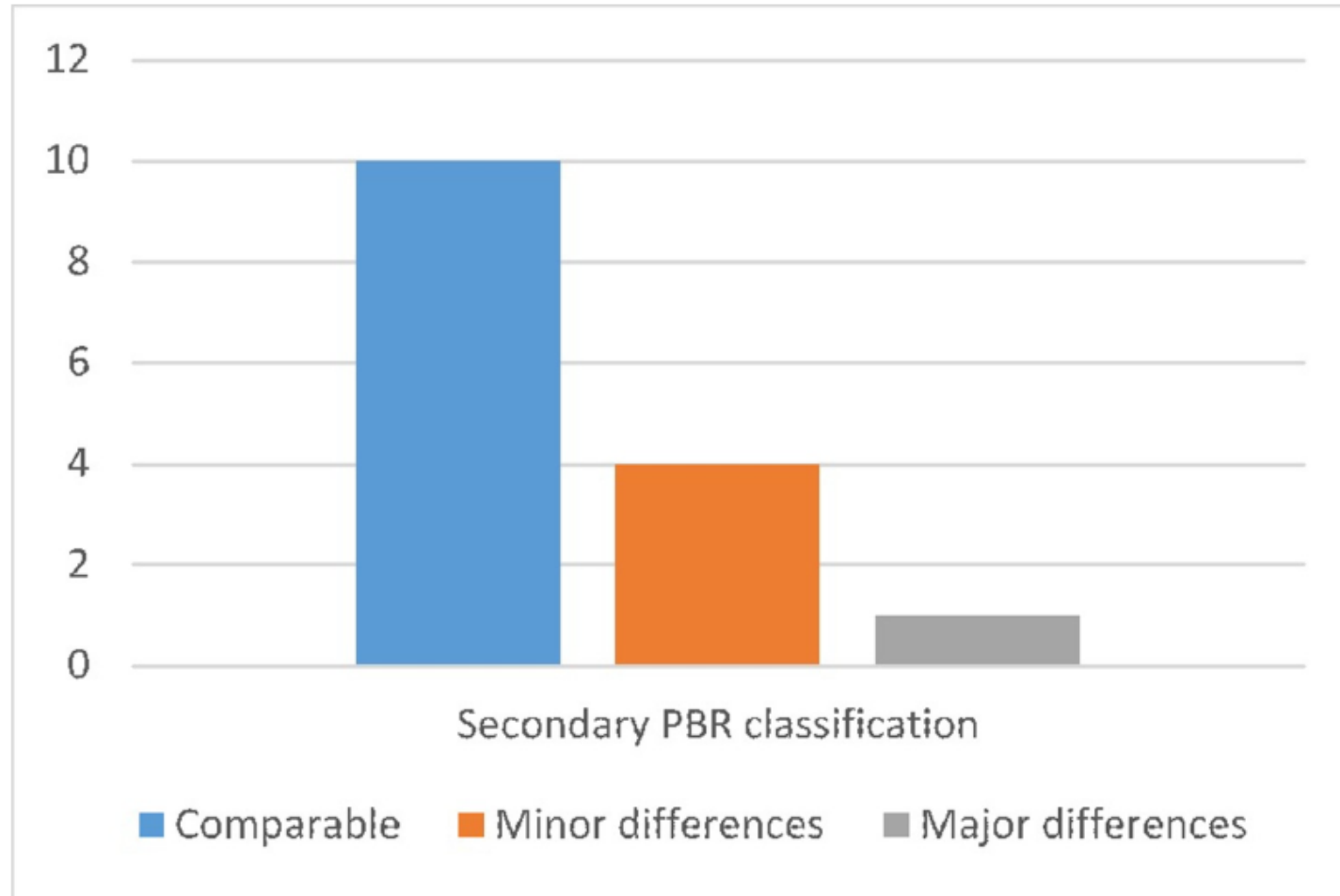


Fig 2. Secondary PBR classification of incomplete studies.

Wood, et al. (2019) findings – data access by journal

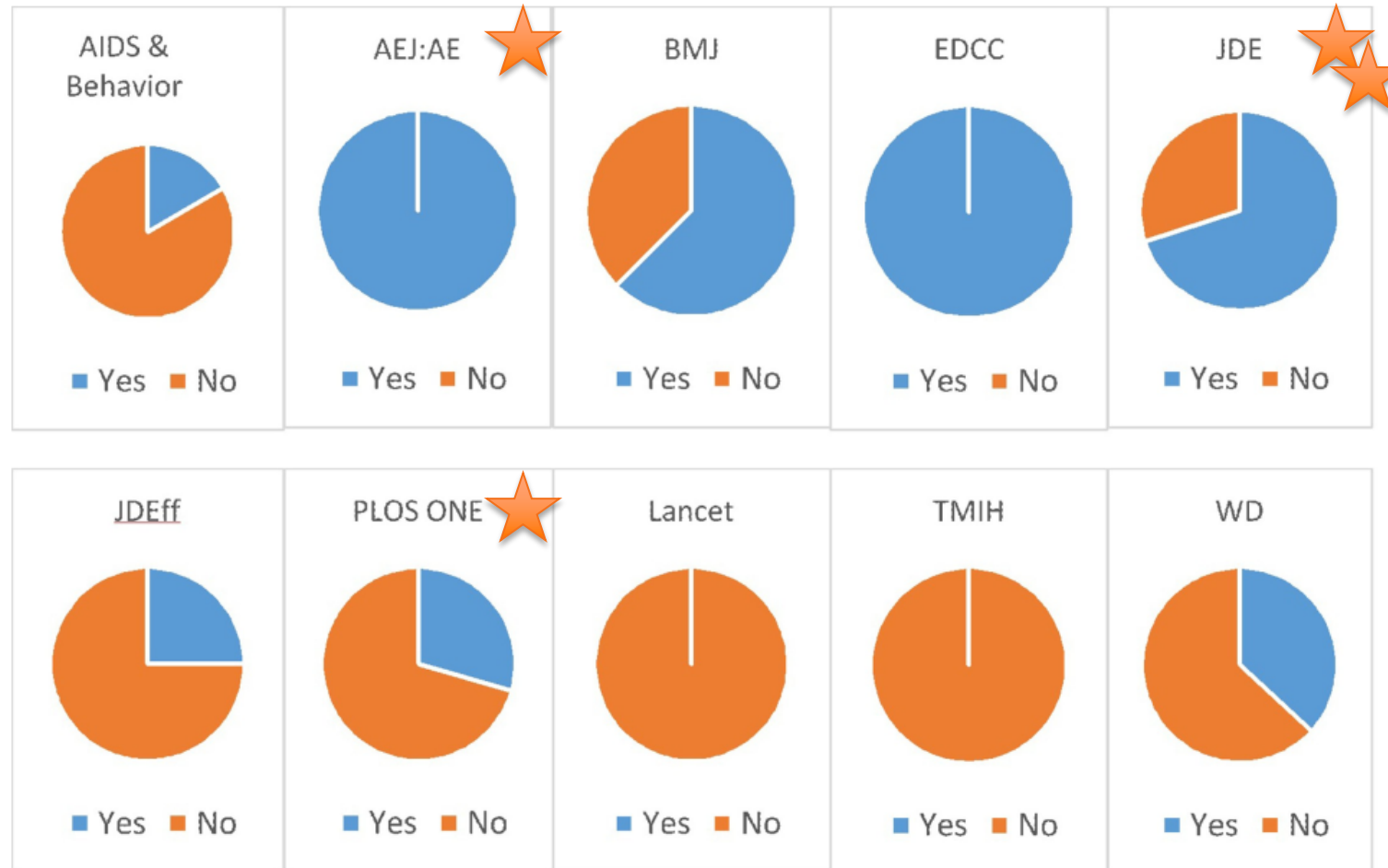


Fig 3. Accessibility of replication files through public access or by request, shares by journal. Those studies coded as 'no' include those with PBR classifications of no access and incomplete. Journal name abbreviations provided in [Table 2](#).

Wood, et al. (2019) findings – data access by funder

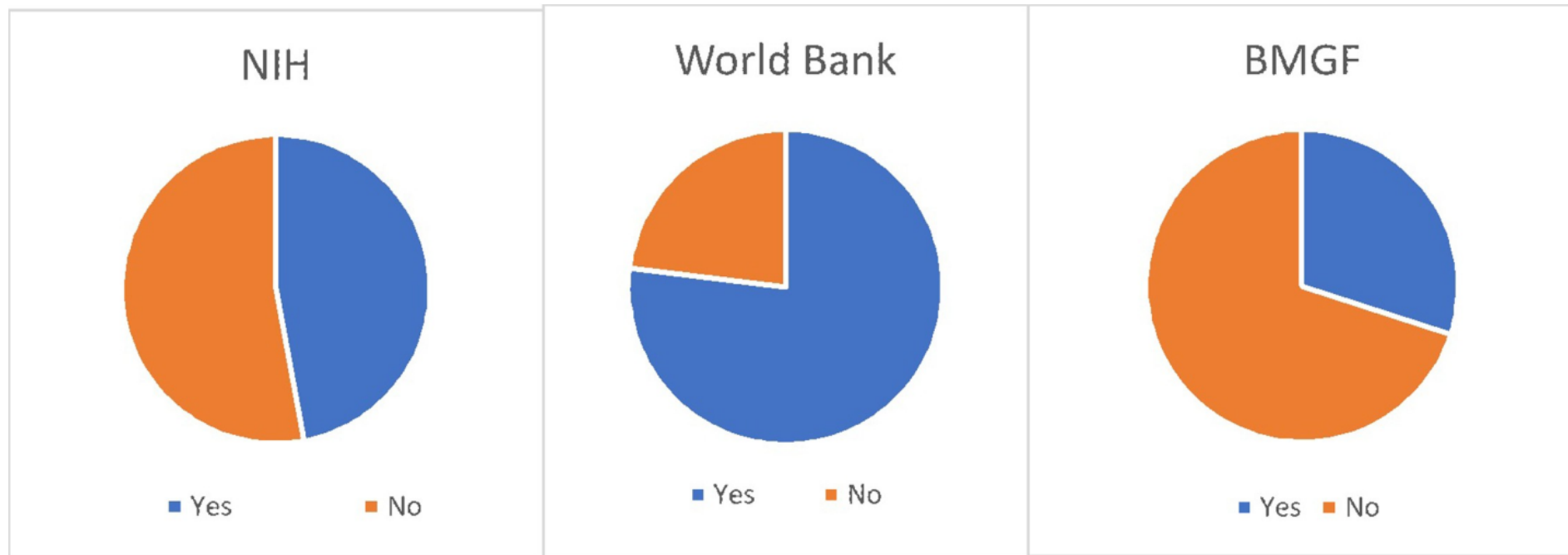


Fig 4. Provision of replication files, shares by three most prevalent research funders in the sample. Those studies coded as 'no' include those with PBR classifications of no access and incomplete.

Conclusions I

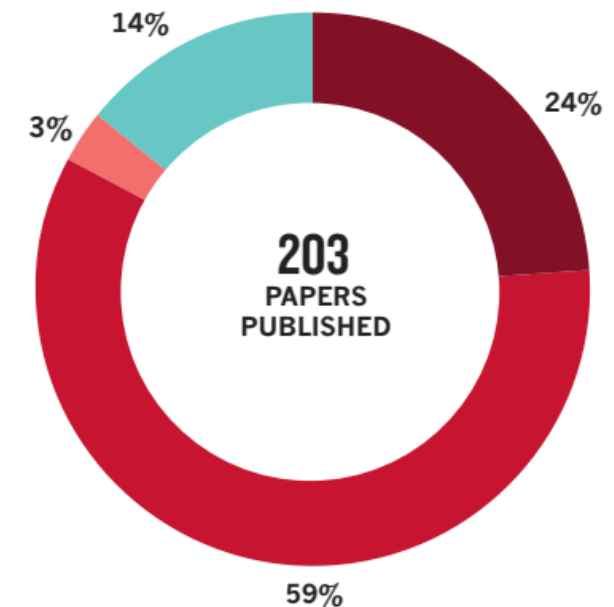
- We still have a long way to go in changing the culture of science concerning availability of replication files
 - Journals starting to make a difference, but some not enforcing
 - Available upon request is not a solution
- Hard to judge computational reproducibility (push button replicability) given limited access to data, however more recent results are encouraging
- Newer focus on raw data vs. estimation data

REPLICATION RARELY POSSIBLE

An analysis of 203 economics papers found that fewer than one in seven supplied the materials needed for replication.

ELEMENTS PROVIDED*:

■ None ■ One or more missing
■ All, code doesn't run ■ All, code runs



*The elements assessed were raw data, raw code, estimation data and estimation code.

Gertler, P, Galiani, S, and Romero, M. "How to make replication the norm?" *Nature* 554 February 2018

What is your ethics statement?

- *PLoS ONE* required an ethics statement.
 - Most similar studies make no mention of ethical approval.
 - Naudet, et al. (2018): “Ethical approval: Not required.”

Ethics statement

Ethical approval is not required. This investigation audits the availability of data and the computational accuracy of program code for published articles. All data provided to us were anonymized, and we only analyzed them by running the provided program code. See Naudet, et al. for a similar study not requiring ethical approval [[12](#)].

May we share identified data?

- *PLoS ONE* editor said no.
- Some other articles do and some do not.
 - Wicherts et al., 2006 – no
 - Stagge et al., 2019 – yes
 - Chang and Li, 2017 – yes (eventually)
 - Naudet et al., 2018 – yes for replications, appears not for no access



Is replication research the study of research or researchers?

- Meta-analysis, systematic review
 - The study of evidence, or research findings
- Metascience
 - The study of how we do research
- Replication studies
 - The study of both research findings and how we do research
 - Is replication research meaningful or useful if we cannot point to the original study?

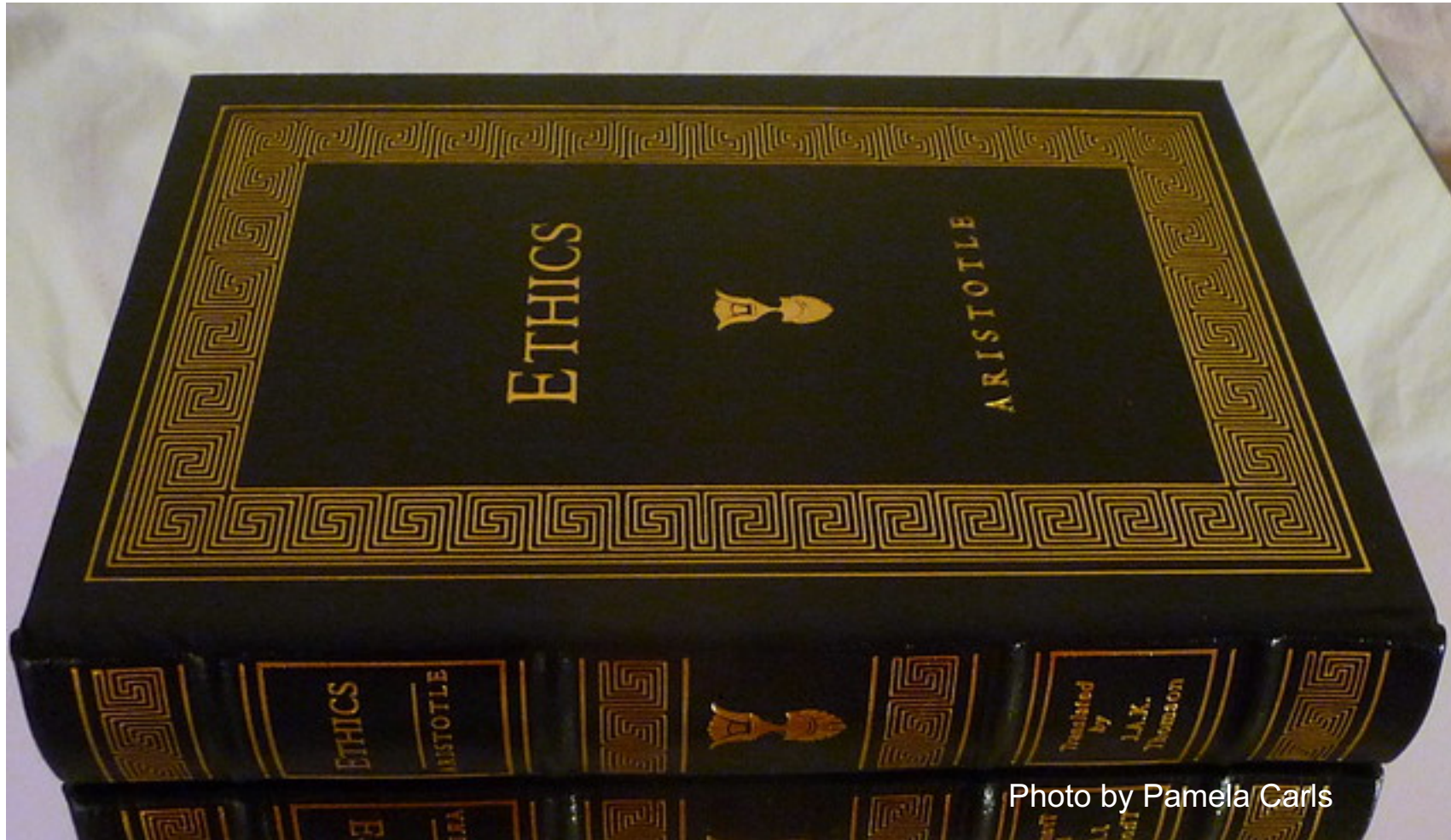


Photo by Pamela Carls

Systematizing, defending, and recommending concepts of right and wrong conduct

The 3ie replication program





What domains for replication ethics?

- Refusal
- Requirements
- Rhetoric
- Review
- Reply



Calum Davey @CalumDavey · Aug 20



Is it just me, or when researchers come up with a framework where all the components begin with the same letter does it make you think 'this is made-up rubbish, isn't it?'



Refusal (consent)

“We would prefer not to participate please.”

Requirements

- Data available upon *reasonable* request
 - Credentials
 - Purpose
 - Methods or design
- Information requirement or judgment call

Rhetoric

- Defining replication for the purpose of claiming success and failure
- Language within the replication study, e.g. mistake, error



Review

- Review to assist or to approve?
- Timing of review



Photo by Chris Waits

Reply



<http://thegreatwesternmovies.com/tag/spaghetti-westerns/>

Conclusions II

- Part of building the replication culture needs to be addressing the questions of replication ethics.

Thank you!