

REPLICATION, PREREGISTRATION & OPEN SCIENCE

why & how



EXPERIMENTAL SCIENCE IN CRISIS

Measuring reproducibility

270 contributors

massive replication of 100 studies
from 3 journals

Psych. Science

Journal of Personality & Social Psych.

J. of Exp. Psych.: Learning, Memory &
Cognition

all published in 2008

high statistical power

original material and/or feedback from org.authors

RESEARCH ARTICLE

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*



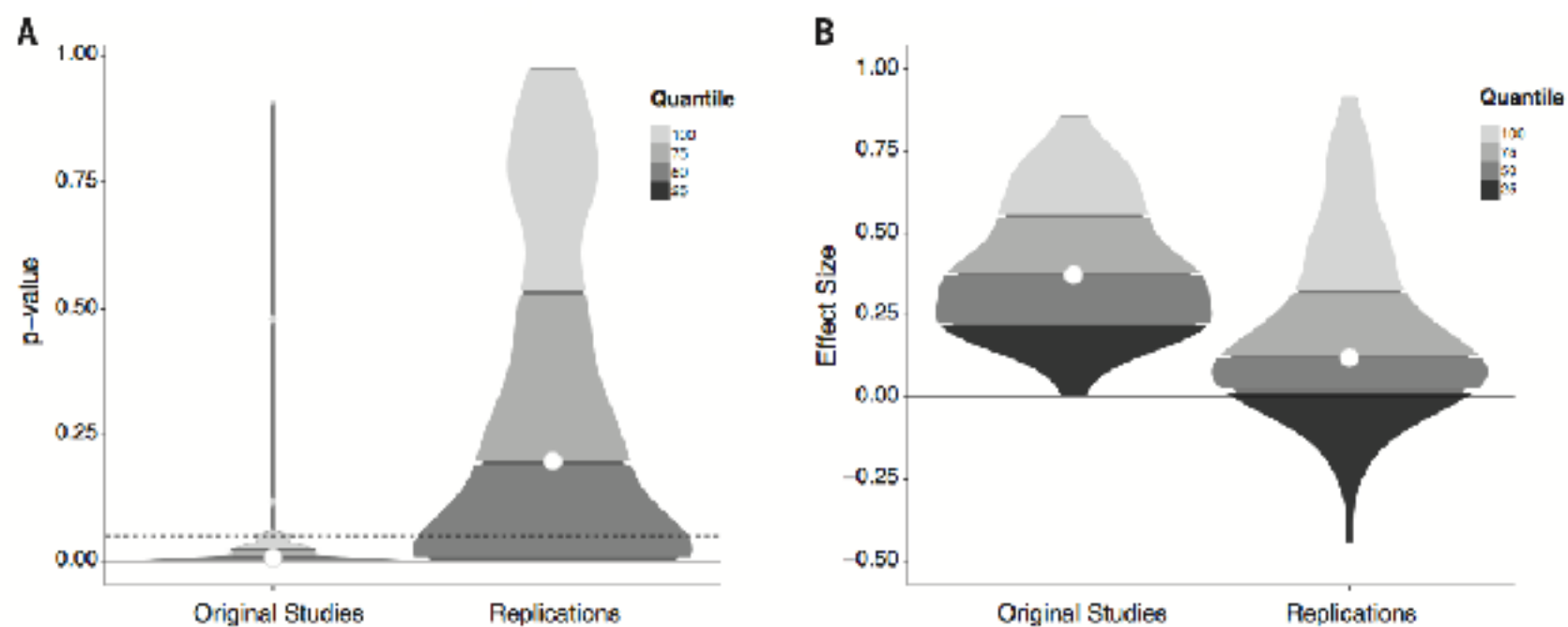


Fig. 1. Density plots of original and replication P values and effect sizes. (A) P values. (B) Effect sizes (correlation coefficients). Lowest quartiles for P values are not visible because they are clustered near zero.

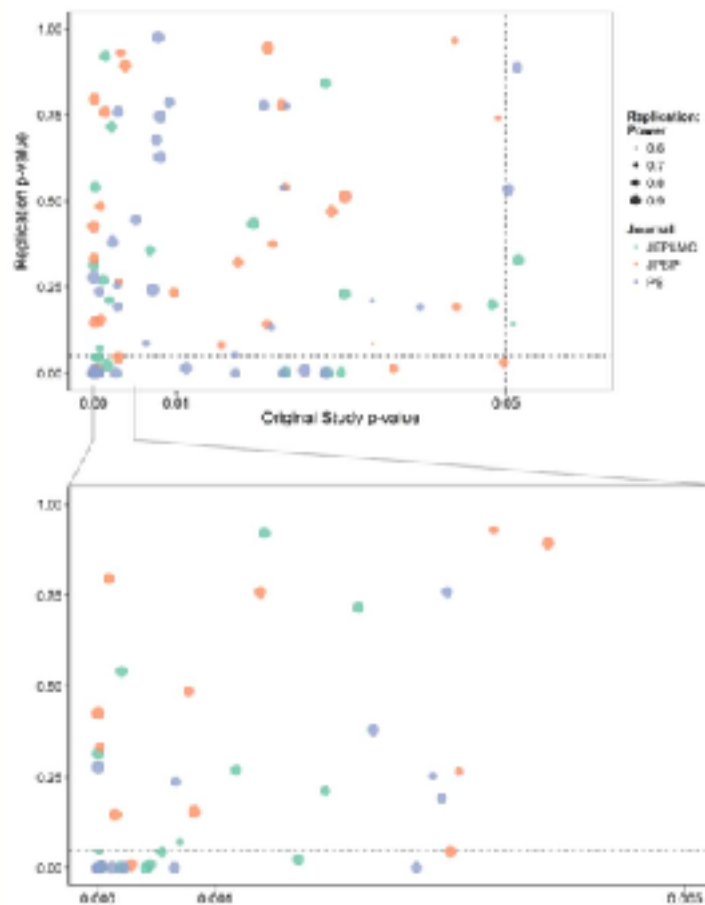


Fig. 2. Scatterplots of original study and replication P values for three psychology journals. Data points scaled by power of the replication based on original study effect size. Dotted red lines indicate $P = 0.05$ criterion. Subplot below shows P values from the range between the gray lines ($P = 0$ to 0.005) in the main plot above.

Measuring reproducibility

“large proportion of replications produced weaker evidence for the original findings”

“strength of initial evidence (such as original P value) was more predictive of replication success than [characteristics of the researchers]”

RESEARCH ARTICLE

PSYCHOLOGY

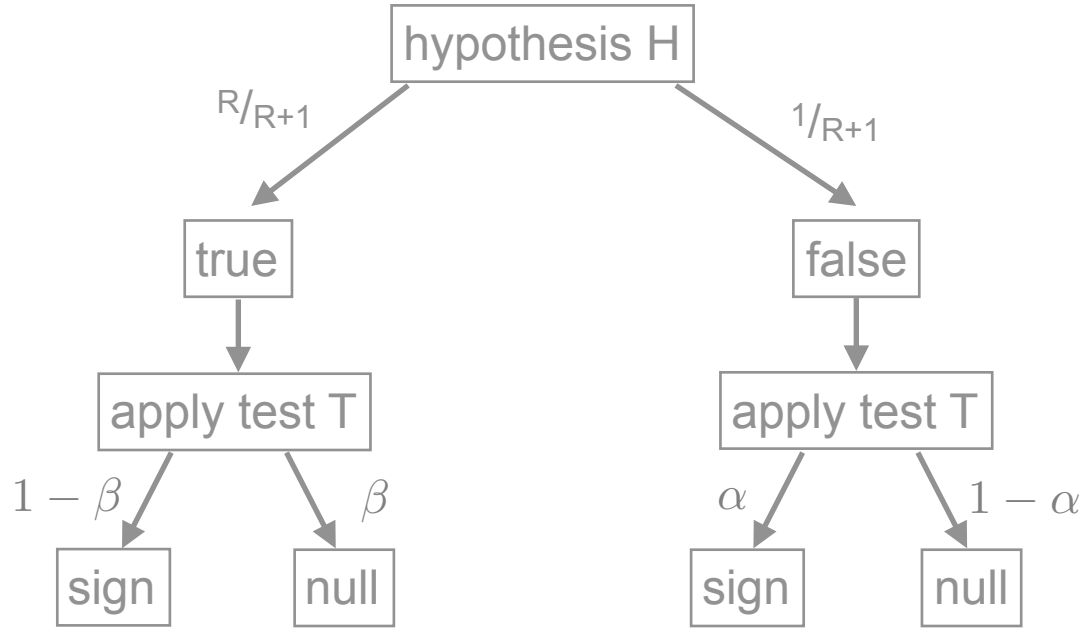
Estimating the reproducibility of psychological science

[Open Science Collaboration]

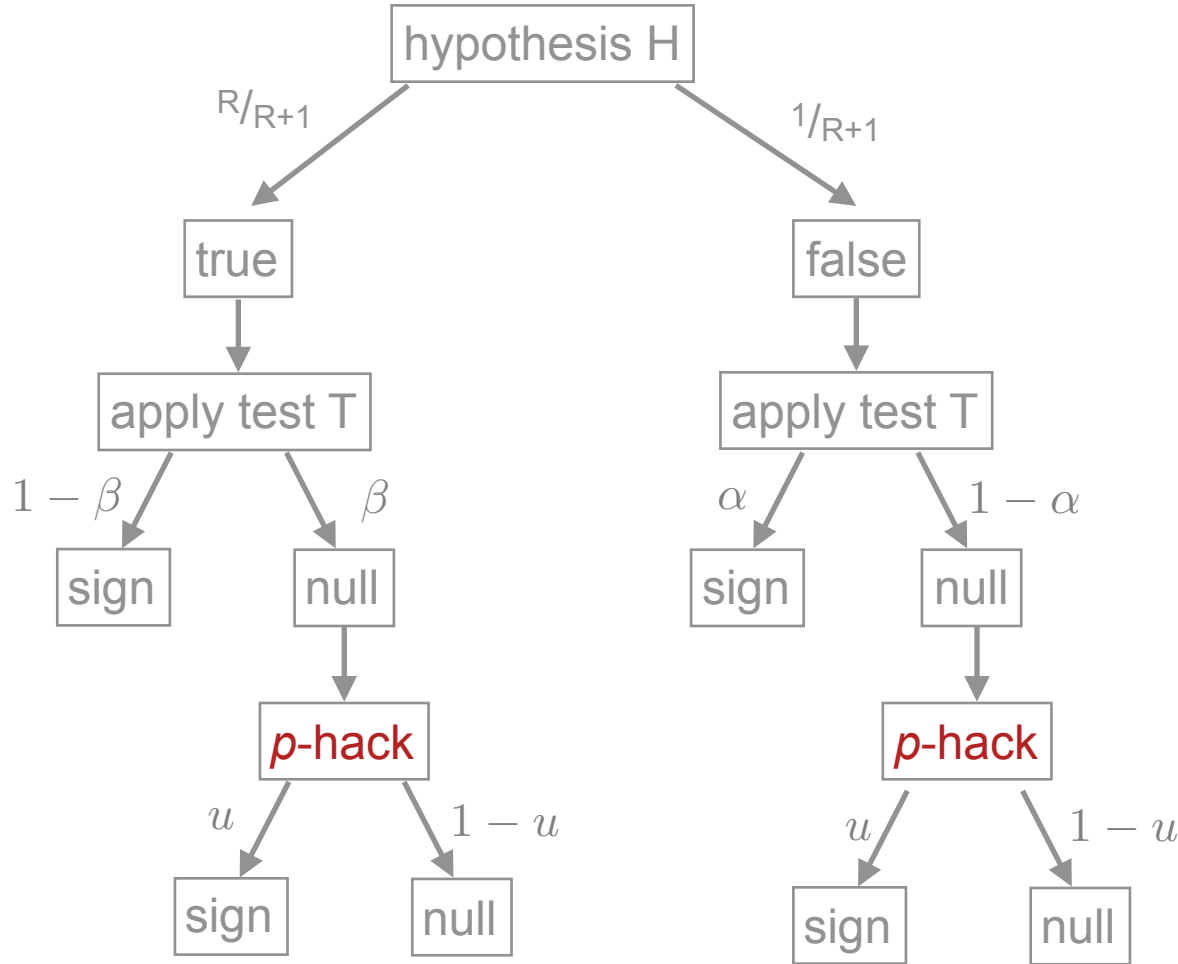
what's the problem?

what's the solution?

Example problem ::: p -hacking



Example problem ::: *p*-hacking



p-hacking
combination of design/
presentation/analysis
factors that favor a
significant test result
beyond the normal α
level

Potential (partial) remedies

(direct) **re**plication

preregistration

open **s**cience

Reproducibility

- conceptual replication** tests the same idea in a different experimental context
- direct replication** recreates the conditions believed necessary for effect X and tests whether X is observed
- reproducibility** extent to which direct replication is successful

Recommended reading

