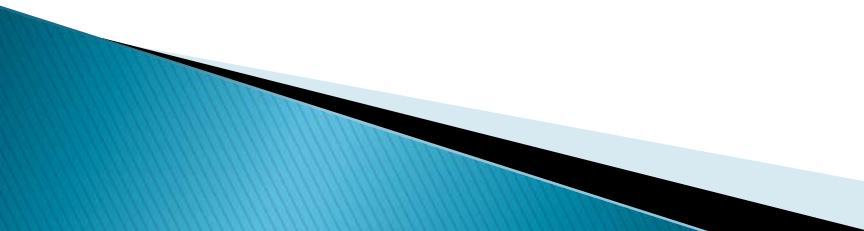


# *p*-values Had a Good Run: A Primer on the ‘New Statistics’

Rob Cribbie  
Quantitative Methods Program  
Department of Psychology  
York University

# Part 5: Replication

- ▶ Replication is a term referring to the repetition of a research study, generally in different situations and with different subjects, to determine if the basic findings of the original study can be applied to other participants and circumstances
  - ▶ Replication and internal validity, together, form a basis for inferring *external validity*
- 

# Why is Replication Important?

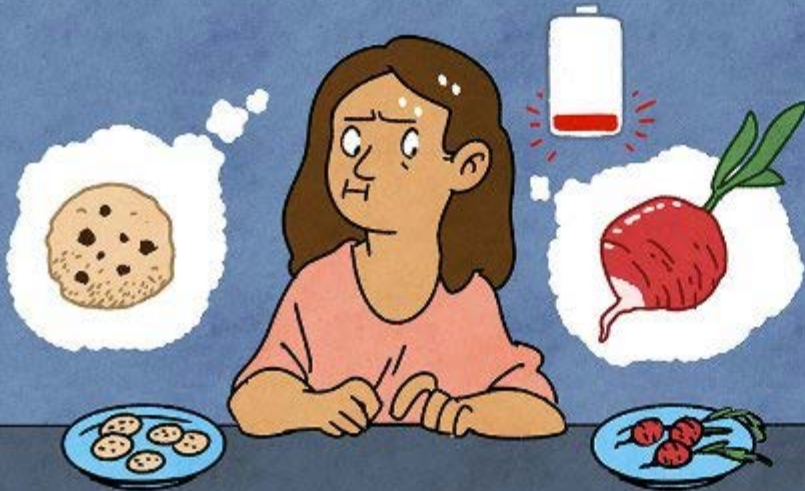
- ▶ Replication is EVERYTHING!
  - Or, external validity is everything!
    - But the current topic is ‘replication’, so that sounds better
- ▶ If a finding does not hold with different subjects, in different settings, with different researchers, etc. then what can we conclude regarding the finding? Not much

# Why is Replication Important?

- ▶ Without replication, the development of science is on thin ice
  - Many studies have a Type II error rate of at least .20, and a Type I error rate of at least 05
  - So ... given the thousands of research projects being conducted at any given time, MANY, MANY mistakes are reported in published studies
    - Replication serves the role of “getting rid of the junk”

# (Lack of) Replication in Psychology

A popular concept in psychology is that your self-control is a limited resource that can be worn down.



Called **ego-depletion**, the idea has been enjoying success in the literature for over a decade now.

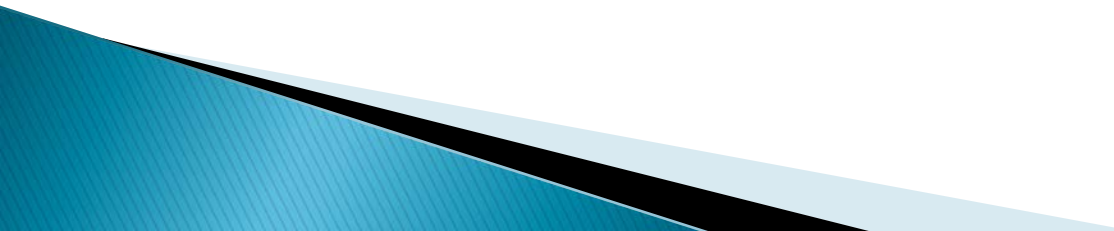
But when **Martin Hagger** and **Nikos Chatzisarantis** at Curtin University in Australia gathered over two thousand participants across 23 labs to repeat a seminal ego-depletion experiment, they found...



Nothing.  
Zip. Nada.

**PSYCHOLOGY'S REPRODUCIBILITY PROBLEM**

# (Lack of) Replication in Psychology

- ▶ Another popular example is the “bilingual advantage” effect in executive functions
  - ▶ Kenneth Paap, of San Francisco State University, concluded after 4 years of trying to replicate this effect, that bilingual advantages in executive functions “either do not exist or are restricted to very specific and undetermined circumstances”
- 

# Nosek et al. (2015)

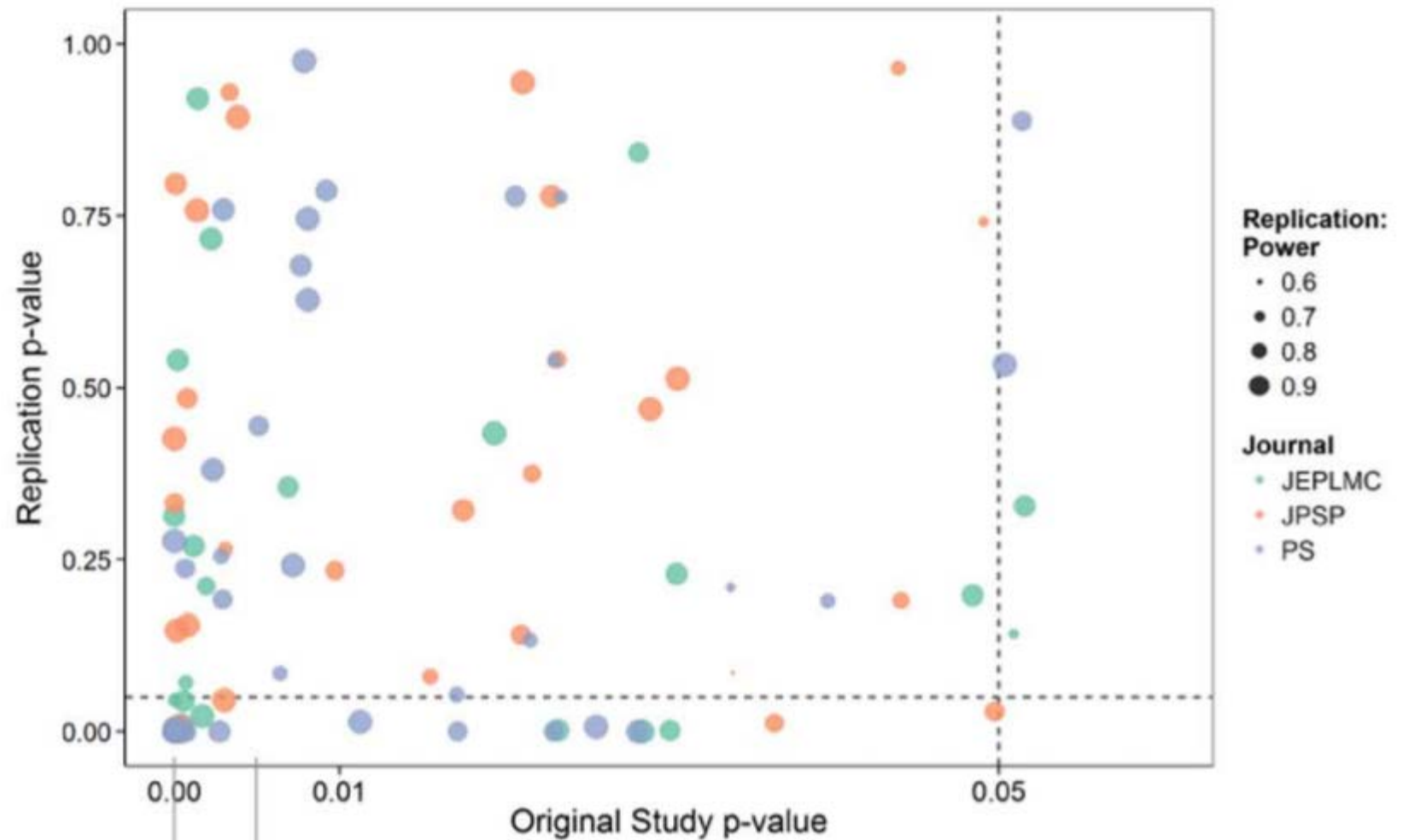
- ▶ Brian Nosek of the University of Virginia, and many colleagues from around the world, sought to replicate 100 different studies that were published in top-tier psychology journals in 2008
- ▶ The journals were:
  - *Psychological Science*
  - *Journal of Personality and Social Psychology*
  - *JEP: Learning, Memory, and Cognition*
- ▶ In their initial publications, 97 of these 100 studies claimed to have significant results

# Nosek et al. (2015)

- ▶ Tried to match the procedures, instruments, etc. of the original studies
  - Although some would argue that they didn't go far enough
- ▶ Only 35 of the 97 studies with statistically significant results in the original study replicated (36.1%) (i.e., both were statistically significant)
- ▶ Further, in only 45 of 95 (47%) studies did the CI for the replication include the sample effect
- ▶ If the studies did replicate, the replication effects were generally smaller than the initial effects



# Nosek et al. (2015)



# Nosek et al. (2015)

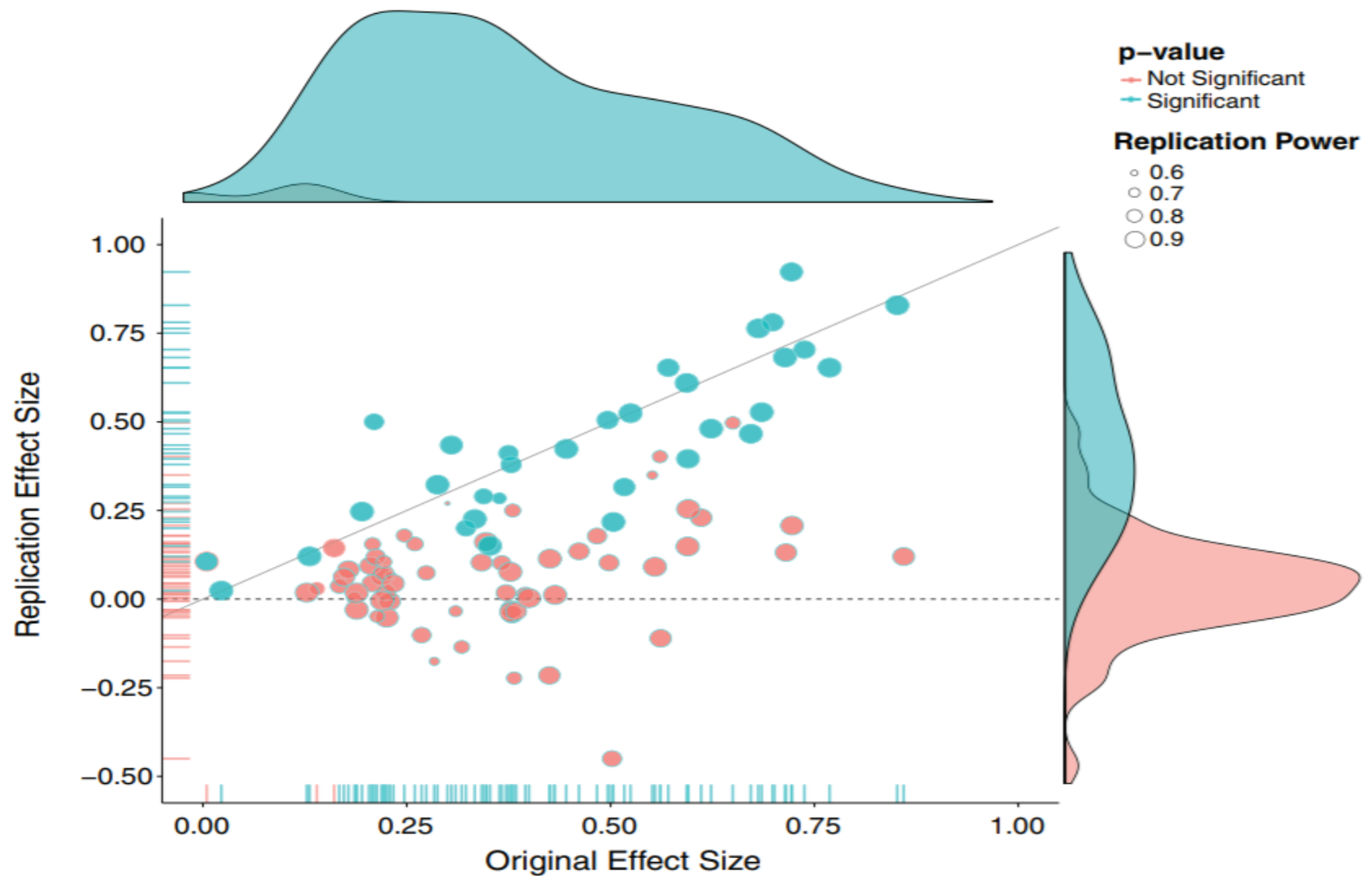


Fig. 3. Original study effect size versus replication effect size (correlation coefficients).

# Why do Studies Not Replicate?

- ▶ John Ioannidis, in a paper entitled “Why Most Published Research Findings Are False”, discussed three issues with current science that contribute to the ‘replication crisis’
  - Selective Publication
    - Studies with statistically significant effects are more likely to be published
      - File Drawer Problem
        - How many non-significant effects are sitting in file-drawers while this one statistically significant effect got published

# Why do Studies Not Replicate?

- Reaching  $p < .05$ 
  - The pressure to publish, and considering selective publication, means that researchers may resort to less than ideal practices in order to find statistically significant effects
- Lack of Replication Studies
  - There is little incentive in science for researchers to complete replications
    - They are hard to publish and receive little recognition (unless they find contradictory results)
    - Thus, there are a lot “single” studies that report effects that have never been replicated

# The “Troubling Trio”

- ▶ Steven Lindsay, the editor of Psychological Science, identified three characteristics of research papers/results that signify an unlikely probability of replication
  - He dubbed them the “troubling trio”
- ▶ 1) low statistical power
- ▶ 2) a “surprising” result
- ▶ 3) a  $p$  value only slightly less than .05

# Questionable Research Practices



# Questionable Research Practices

- ▶ Questionable research practices (QRPs) definitely contribute to a reduced probability of replication
- ▶ QRPs include:
  - Selective reporting of outcome variables
  - Flexible stopping rules
    - Collect data until you reach statistical significance
  - Not reporting the studies that ‘didn’t work’
  - HARKing (hypothesizing after the results are known)
- ▶  $p$ -Hacking is the generic term for QRPs for the purpose of achieving statistical significance
- ▶ Note that fraud (e.g., making up data) is not usually included as a QRP (but obviously wrong!)

## Estimates by Grad Students

In a paper, failing to report all of a study's dependent measures

Deciding whether to collect more data after looking to see whether the results were significant

In a paper, failing to report all of a study's conditions

Stopping collecting data earlier than planned because one found the result that one had been looking for

In a paper, 'rounding off' a p value (e.g., reporting that a p value of .054 is less than .05)

In a paper, selectively reporting studies that 'worked'

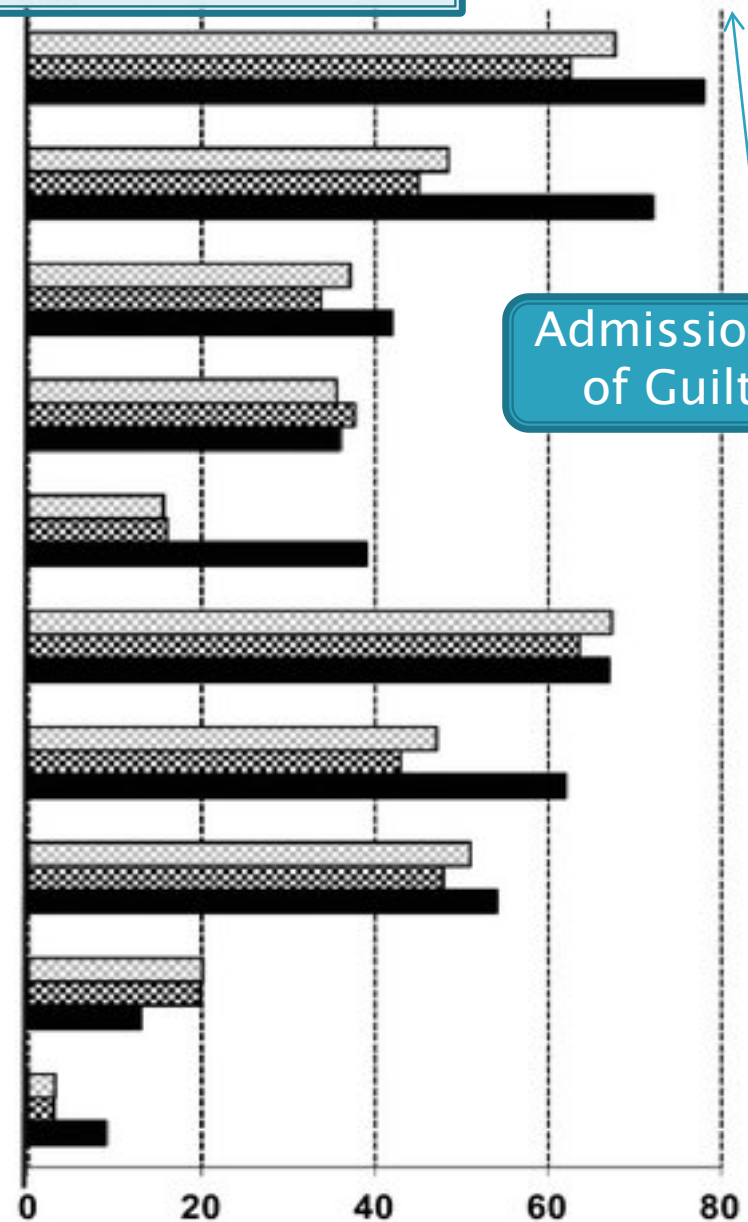
Deciding whether to exclude data after looking at the impact of doing so on the results

In a paper, reporting an unexpected finding as having been predicted from the start

In a paper, claiming the results are unaffected by demographic variables (e.g., gender) when one is actually unsure (or knows that they do)

Falsifying data

My field All scientists John et al. (2012)



Admissions of Guilt

Percentages



# *p*-Hacking

"People realise [p-hacking] is not quite right but they think it is a bit like jaywalking - they don't realise it is like committing burglary."

- PROFESSOR DOROTHY BISHOP

# Replication and Open Science

- ▶ The desire for replicability is the reason that scientific papers include a *Method* section, which describes exactly how the researchers performed the study
  - This includes what measures, procedures, subjects, etc. were used, as well as any discussing any issues that may have arisen during the study

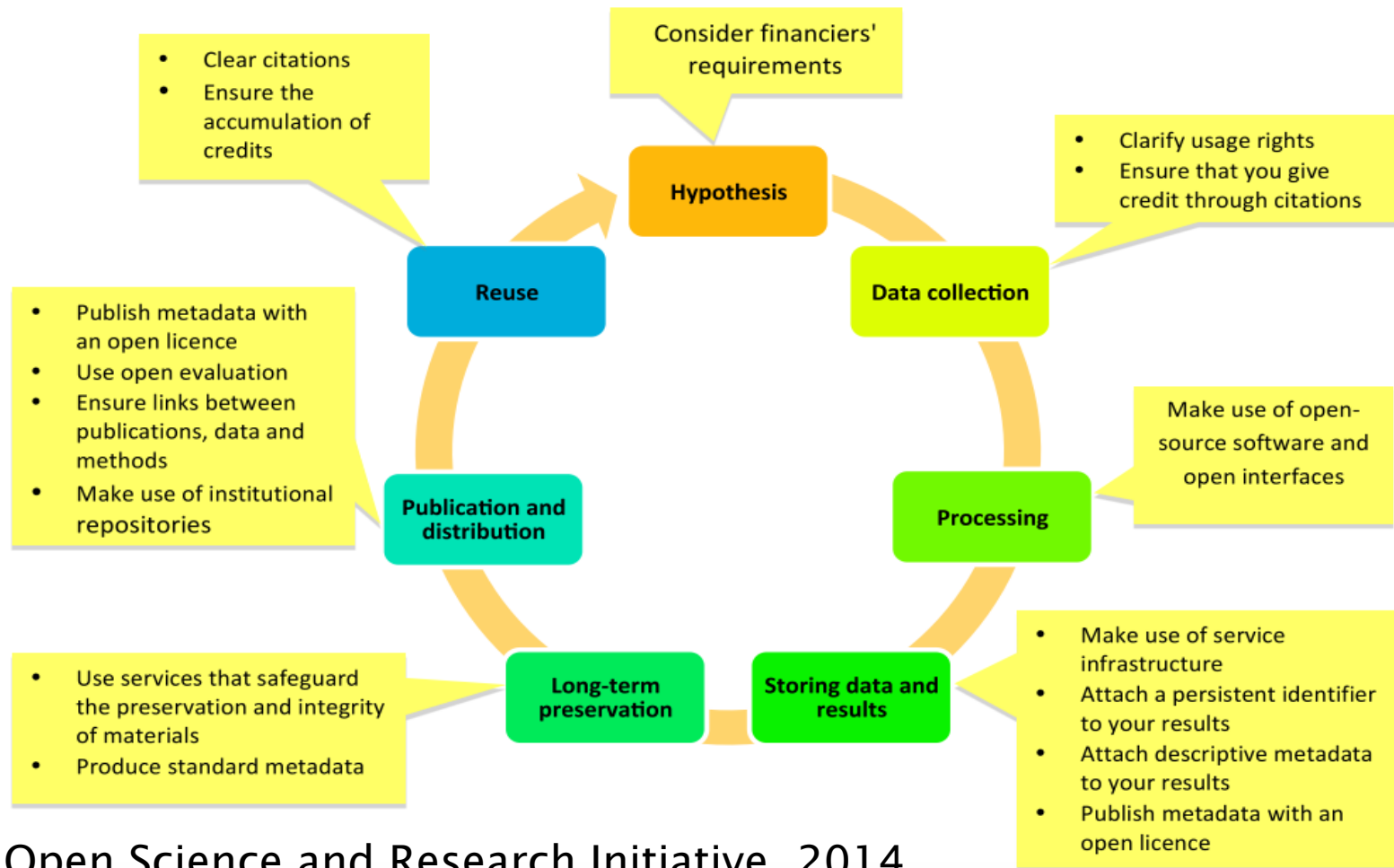


# Replication and Open Science

- ▶ These days *open science* implies a lot more about the nature of the research than it did just a few years ago, and much of it is in response to the current “replication crisis” facing the behavioral sciences
- ▶ In response to the reported low probability of replication in the behavioral sciences, one strategy for improving reproducibility is to make practices, materials, procedures, data, etc. more *open*

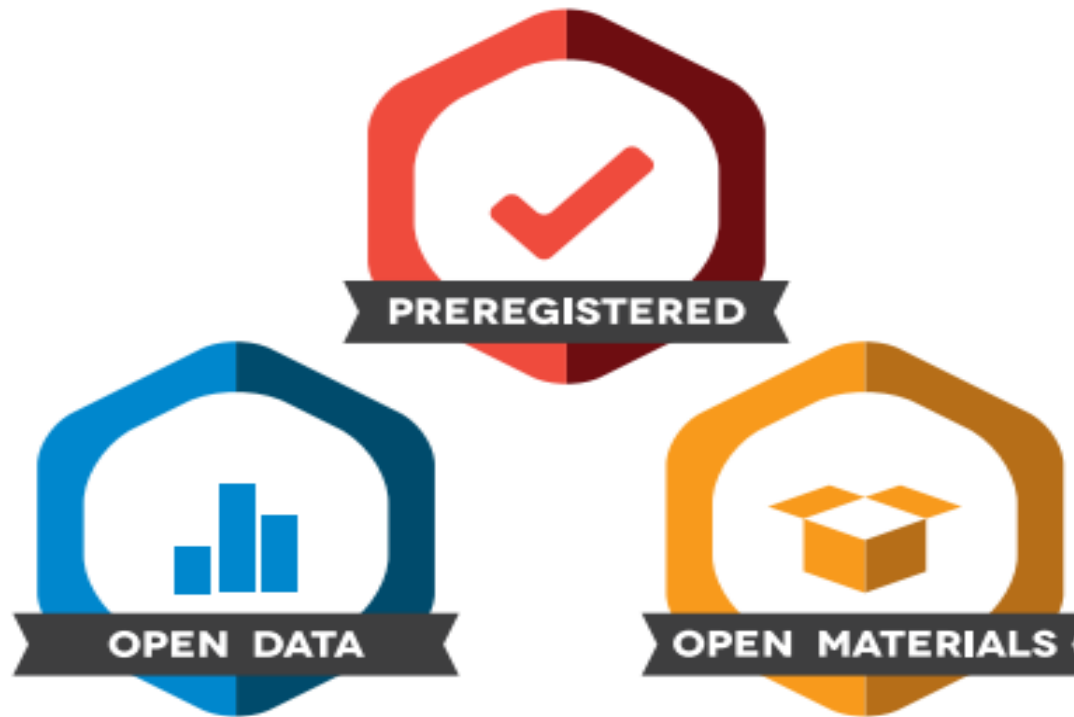
# Replication and Open Science

## ► What Is *Open Science*?



# Replication and Open Science

- ▶ Open Science *Badges* via the Centre for Open Science (COS)

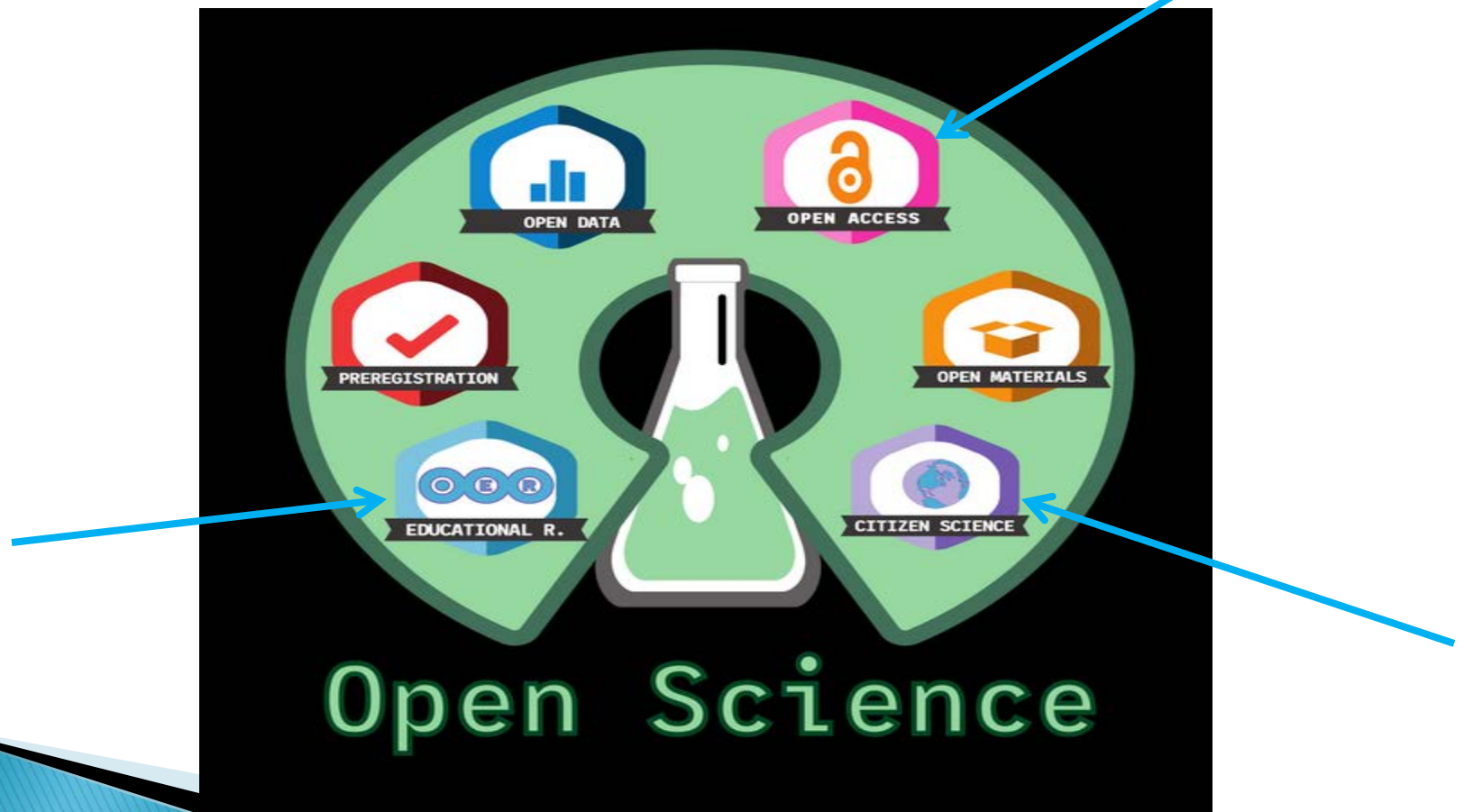


# Replication and Open Science

- ▶ Open Science *Badges*
  - Badges to acknowledge open science practices that are **incentives** for researchers to share data, materials, or to preregister.
- ▶ Badges signal to the reader that the content has been made available and certify its accessibility in a persistent location
- ▶ A recent systematic review identified this badging program as the only evidence-based incentive program that is associated with increased data sharing
  - Rowhani-Farid et al., 2017

# Replication and Open Science

- ▶ Is there more to *Open Science*?

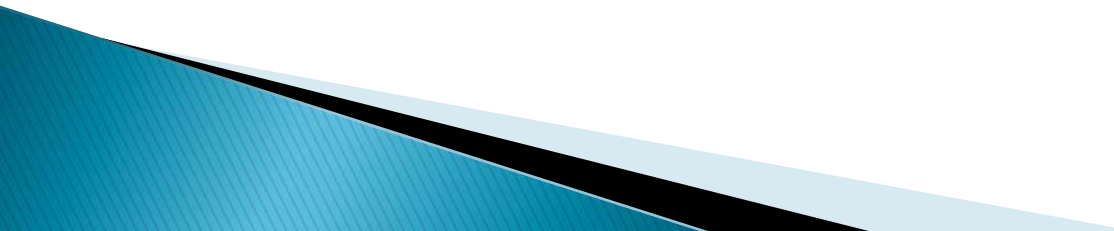


# Replication and Open Science

- ▶ Is there more to *Open Science*?
  - Open Access
    - Publish in journals that provide the research and supplementary materials free of charge
      - Which usually means publication costs for the author, and can also put the publisher in a conflict of interest
  - Open Educational Resources
    - Openly licensed materials that are useful to researchers
      - Includes full courses, course materials, modules, textbooks, streaming videos, tests, and other materials
  - Citizen Science
    - “crowd-sourced” science
    - Public participation in scientific research
      - E.g., smartphone apps for monitoring birds and other animals



# Why Open Science?

- ▶ One of the primary missions of the COS is to improve the reproducibility of scientific findings and encourage replication
    - The first project of the COS was Brian Nosek's *reproducibility project*, discussed earlier
  - ▶ A low probability of replication is definitely related to not preregistering a study, not sharing materials, not sharing data, etc.
    - *Strange* things can happen when transparency is not required ...
- 

# Summary

- ▶ Replicate, Replicate, Replicate!!!
  - *If something is worth doing, it's worth doing twice.*
  - Replications make great theses
  - Many journals now published “Registered Replication Reports” so it should not be hard to get replications published
  
- ▶ Don't  $p$ -hack or perform any other questionable research practices

# Summary

- ▶ Pre-Results Acceptance is Needed
  - A paper is ‘accepted’ based only upon the introduction and methods
    - Eliminates the chance of publication bias, because papers can’t be rejected for not finding statistical significance
- ▶ Get your badges!
  - Share your materials and data, pre-register your hypotheses, publish in open access journals, create open educational resources, etc.