

Publication Bias and Statistical Evidence in the Psychological Literature

Joachim Vandekerckhove
Alexander Etz
Maime Guan

Department of Cognitive Sciences
University of California, Irvine

Background

- Reproducibility Project: Psychology (Open Science Collaboration, 2015)
 - ▶ attempt to closely replicate 100 studies from leading journals
 - ▶ replications were pre-registered in order to avoid selection and publication bias
 - ▶ fewer than half successfully replicated (39% if by $p < .05$)

Background

- Reproducibility Project: Psychology (Open Science Collaboration, 2015)
 - ▶ attempt to closely replicate 100 studies from leading journals
 - ▶ replications were pre-registered in order to avoid selection and publication bias
 - ▶ fewer than half successfully replicated (39% if by $p < .05$)
- Reactions differ
 - ▶ “Scientists replicated 100 psychology studies, and fewer than half got the same results” (*Smithsonian Magazine*)
 - ▶ “More than half of psychology papers are not reproducible” (*Times Higher Education*)
 - ▶ “Study reveals that a lot of psychology research really is just ‘psycho-babble’ ” (*The Independent*)
 - ▶ “Psychology is not in crisis” (Lisa Feldman Barrett, *New York Times*)
 - ▶ “... hidden moderators...” (APA)

Publication bias

- What is publication bias?
 - ▶ A differential publishing rate between positive and negative results
 - ▶ “The small proportion of results chosen for publication are unrepresentative of scientists’ repeated samplings of the real world” (Young et al., 2008)

- What is publication bias?
 - ▶ A differential publishing rate between positive and negative results
 - ▶ “The small proportion of results chosen for publication are unrepresentative of scientists’ repeated samplings of the real world” (Young et al., 2008)
- A rising concern
 - ▶ Academic journals tend to prefer results that sound conclusive
 - ▶ Well-designed studies that do not find significant differences are less publishable
 - ▶ Competition for scarce funding and employment put researchers in a weak position to demand changes in journal policies

The consequences of publication bias were already foreseen in this abstract by an apparently horrified Theodore Sterling (1959):

There is some evidence that in fields where statistical tests of significance are commonly used, research which yields nonsignificant results is not published. Such research being unknown to other investigators may be repeated independently until eventually by chance a significant result occurs—an “error of the first kind”—and is published. Significant results published in these fields are seldom verified by independent replication. The possibility thus arises that the literature of such a field consists in substantial part of false conclusions resulting from errors of the first kind in statistical tests of significance.

Our view: Published data are **complex** and **different from lab data**

Our view: Published data are **complex** and **different from lab data**

- One might naively believe that published data are simply the result of a well-documented empirical process (e.g., an experiment)

Our view: Published data are **complex** and **different from lab data**

- One might naively believe that published data are simply the result of a well-documented empirical process (e.g., an experiment)
- Rather, they are the result of a combined process of empirical data collection and the sociopsychological process of scientific publication

Our view: Published data are **complex** and **different from lab data**

- One might naively believe that published data are simply the result of a well-documented empirical process (e.g., an experiment)
- Rather, they are the result of a combined process of empirical data collection and the sociopsychological process of scientific publication
- When we evaluate the evidence provided by published data, we implicitly use these naive assumptions that are **known to be false**

Our view: Published data are **complex** and **different from lab data**

- One might naively believe that published data are simply the result of a well-documented empirical process (e.g., an experiment)
- Rather, they are the result of a combined process of empirical data collection and the sociopsychological process of scientific publication
- When we evaluate the evidence provided by published data, we implicitly use these naive assumptions that are **known to be false**
- Retrospective analyses should instead quantify the sampling distribution of **the actual data at hand**

Our view: Published data are **complex** and **different from lab data**

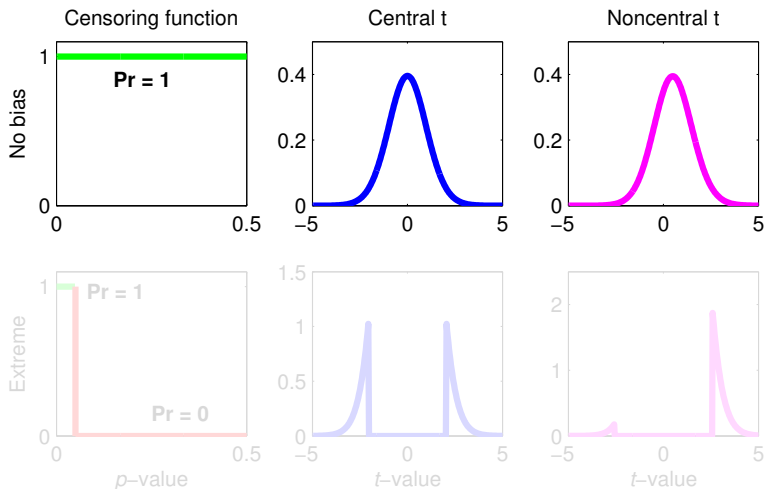
- One might naively believe that published data are simply the result of a well-documented empirical process (e.g., an experiment)
- Rather, they are the result of a combined process of empirical data collection and the sociopsychological process of scientific publication
- When we evaluate the evidence provided by published data, we implicitly use these naive assumptions that are **known to be false**
- Retrospective analyses should instead quantify the sampling distribution of **the actual data at hand**
- The appropriate sampling plan involves a joint process model of empirical data collection and incomplete (“censored”) publishing

Models of publication bias

We consider a series of four models for the publication process

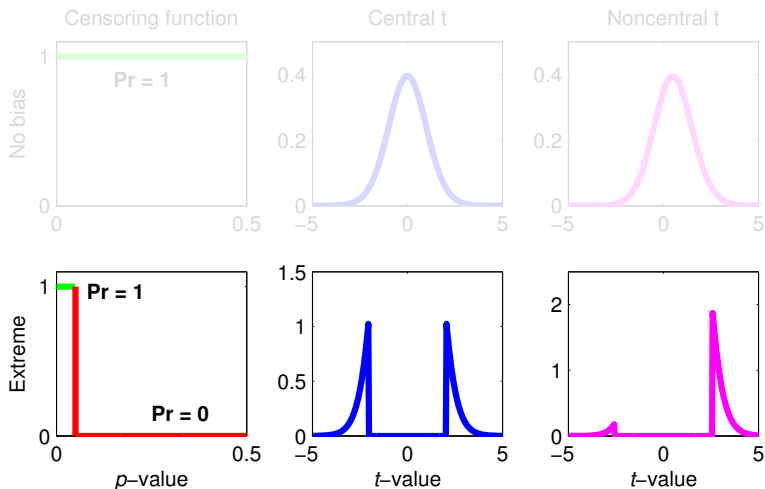
Models of publication bias

We consider a series of four models for the publication process:



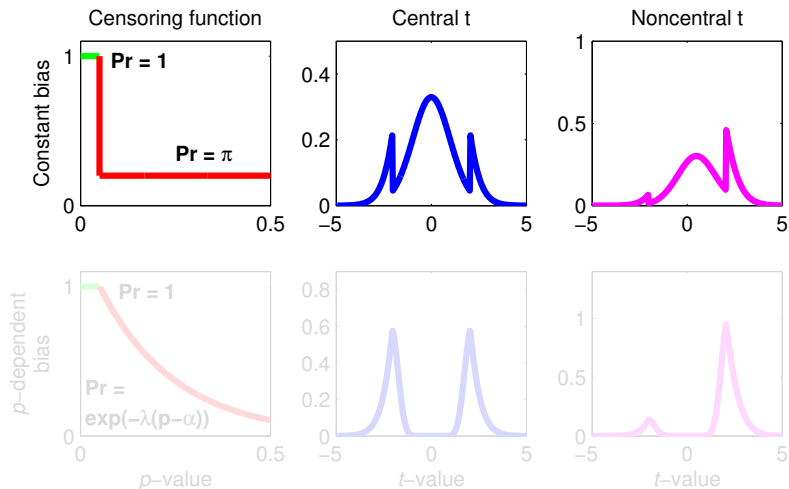
Models of publication bias

We consider a series of four models for the publication process:



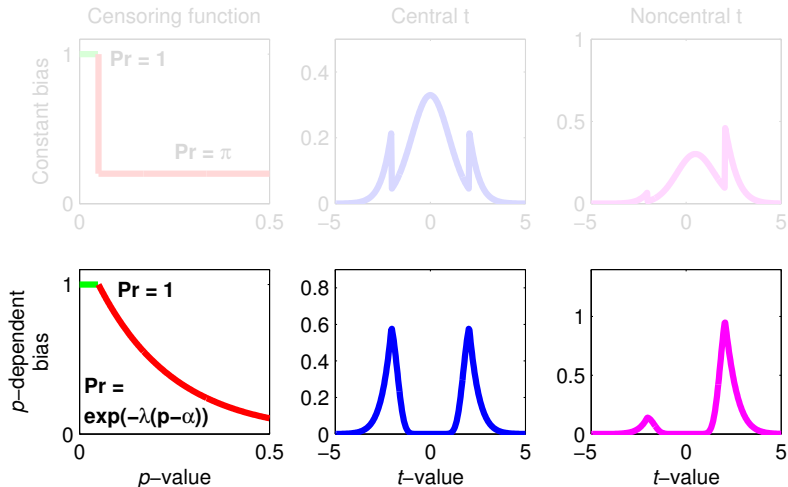
Models of publication bias

We consider a series of four models for the publication process:



Models of publication bias

We consider a series of four models for the publication process:



How does this help?

- We are rarely in a position to know which of these four sociopsychological censoring processes (if any) was active

How does this help?

- We are rarely in a position to know which of these four sociopsychological censoring processes (if any) was active
 - “Identity of censoring function is unknown”

How does this help?

- We are rarely in a position to know which of these four sociopsychological censoring processes (if any) was active
 - “Identity of censoring function is unknown”
- Exact censoring function is ultimately not of general interest

How does this help?

- We are rarely in a position to know which of these four sociopsychological censoring processes (if any) was active
 - “Identity of censoring function is unknown”
- Exact censoring function is ultimately not of general interest
- Issue of interest is \mathcal{H}_0 vs \mathcal{H}_A : which **empirical** process was true?

How does this help?

- We are rarely in a position to know which of these four sociopsychological censoring processes (if any) was active
 - “Identity of censoring function is unknown”
- Exact censoring function is ultimately not of general interest
- Issue of interest is \mathcal{H}_0 vs \mathcal{H}_A : which **empirical** process was true?
- Identity of the censoring function can be “integrated out”

How does this help?

- We are rarely in a position to know which of these four sociopsychological censoring processes (if any) was active
 - “Identity of censoring function is unknown”
- Exact censoring function is ultimately not of general interest
- Issue of interest is \mathcal{H}_0 vs \mathcal{H}_A : which **empirical** process was true?
- Identity of the censoring function can be “integrated out”
... through Bayesian model averaging

How does this help?

- We are rarely in a position to know which of these four sociopsychological censoring processes (if any) was active
 - “Identity of censoring function is unknown”
- Exact censoring function is ultimately not of general interest
- Issue of interest is \mathcal{H}_0 vs \mathcal{H}_A : which **empirical** process was true?
- Identity of the censoring function can be “integrated out”
... through Bayesian model averaging
- Compute a **mitigated measure of evidence** B^M (Guan & Vandekerckhove, 2016)

Evidence in the Reproducibility Project: Psychology

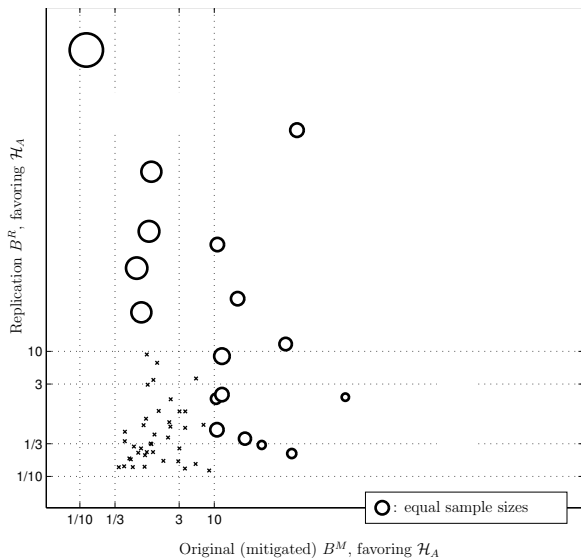
Evidence in the Reproducibility Project: Psychology

- Comparing 'target papers' to their replications, evidence in the original papers should be computed taking into account the possibility of publication bias

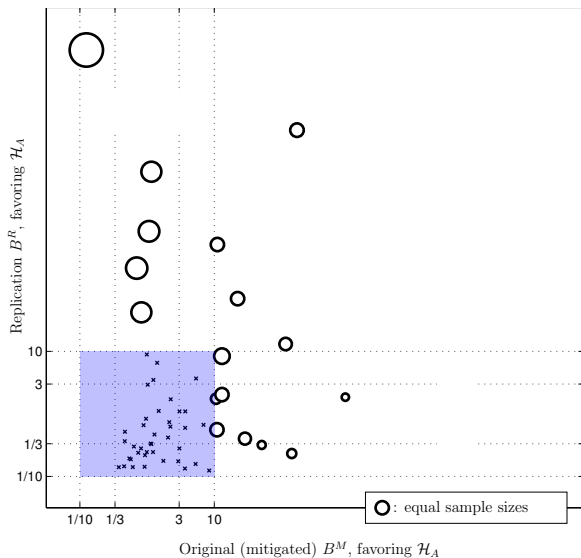
Evidence in the Reproducibility Project: Psychology

- Comparing 'target papers' to their replications, evidence in the original papers should be computed taking into account the possibility of publication bias
- Replications are pre-registered, so no mitigation is required

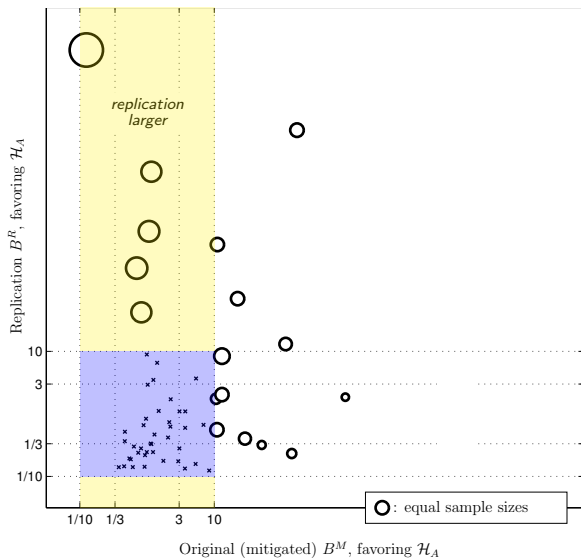
Evidence in the Reproducibility Project: Psychology



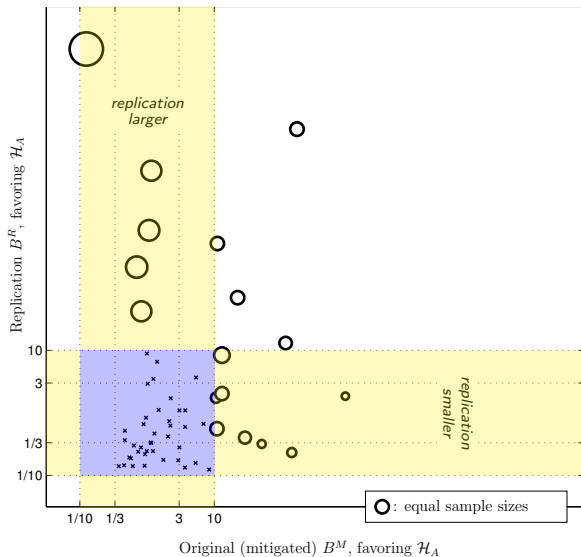
Evidence in the Reproducibility Project: Psychology



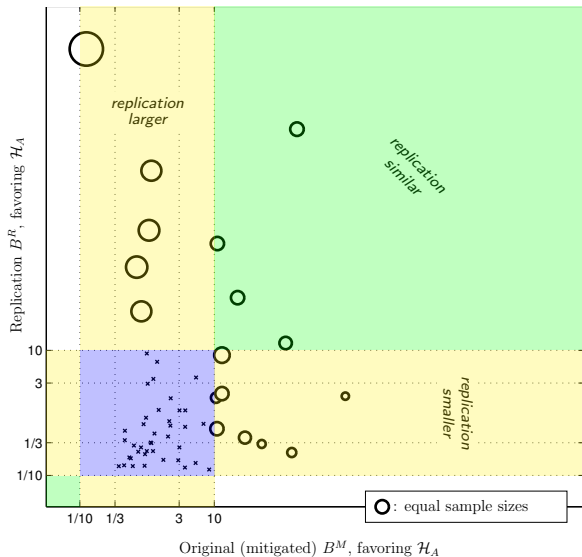
Evidence in the Reproducibility Project: Psychology



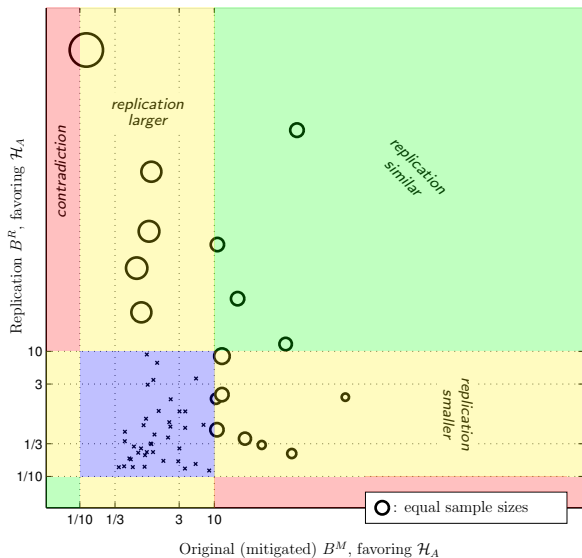
Evidence in the Reproducibility Project: Psychology



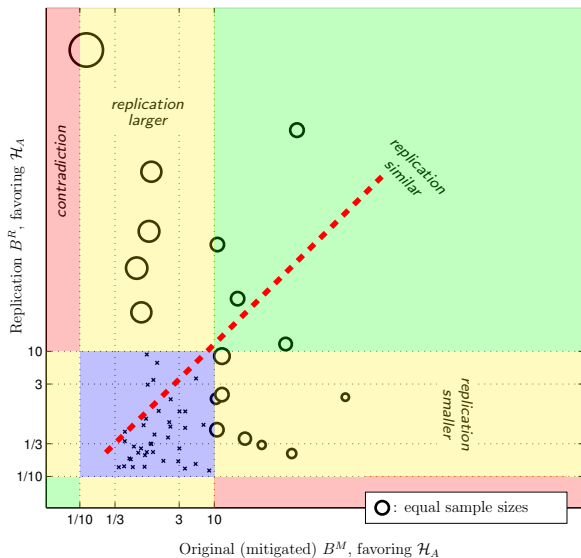
Evidence in the Reproducibility Project: Psychology



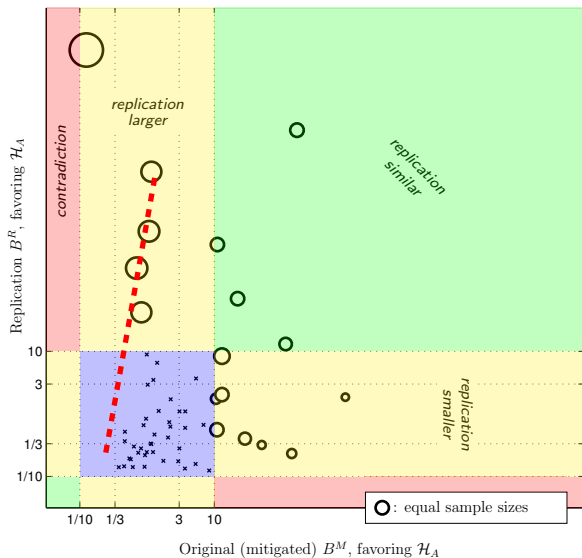
Evidence in the Reproducibility Project: Psychology



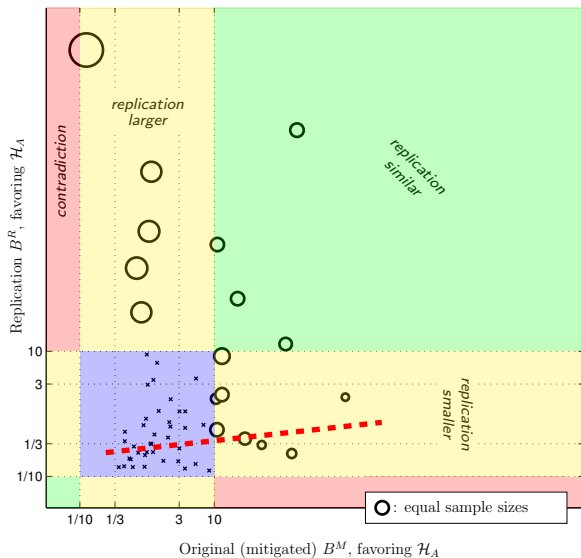
Evidence in the Reproducibility Project: Psychology



Evidence in the Reproducibility Project: Psychology

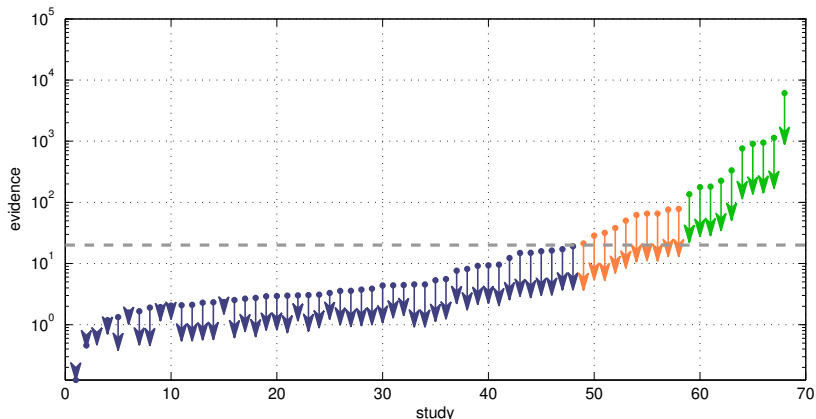


Evidence in the Reproducibility Project: Psychology



Evidence in the Reproducibility Project: Psychology

- Accounting for publication bias **raises the evidence bar for all studies** (Etz & Vandekerckhove, 2016)



Publication bias costs evidence

Reproducibility Project: Psychology emphasized that **evidence is precious**

Publication bias costs evidence

Reproducibility Project: Psychology emphasized that **evidence is precious**

- Evidence in the RP:P did not indicate any clear inconsistencies

Publication bias costs evidence

Reproducibility Project: Psychology emphasized that **evidence is precious**

- Evidence in the RP:P did not indicate any clear inconsistencies
 - ▶ Evidence generally pointed in one direction, never opposite directions

Publication bias costs evidence

Reproducibility Project: Psychology emphasized that **evidence is precious**

- Evidence in the RP:P did not indicate any clear inconsistencies
 - ▶ Evidence generally pointed in one direction, never opposite directions
 - ▶ There is **no need to assume hidden moderators** to explain RPP results

Publication bias costs evidence

Reproducibility Project: Psychology emphasized that **evidence is precious**

- Evidence in the RP:P did not indicate any clear inconsistencies
 - ▶ Evidence generally pointed in one direction, never opposite directions
 - ▶ There is **no need to assume hidden moderators** to explain RPP results
- Evidence in the Reproducibility Project depends greatly on sample size

Publication bias costs evidence

Reproducibility Project: Psychology emphasized that **evidence is precious**

- Evidence in the RP:P did not indicate any clear inconsistencies
 - ▶ Evidence generally pointed in one direction, never opposite directions
 - ▶ There is **no need to assume hidden moderators** to explain RPP results
- Evidence in the Reproducibility Project depends greatly on sample size
 - ▶ Obviously because each observation carries some evidence

Publication bias costs evidence

Reproducibility Project: Psychology emphasized that **evidence is precious**

- Evidence in the RP:P did not indicate any clear inconsistencies
 - ▶ Evidence generally pointed in one direction, never opposite directions
 - ▶ There is **no need to assume hidden moderators** to explain RPP results
- Evidence in the Reproducibility Project depends greatly on sample size
 - ▶ Obviously because each observation carries some evidence
 - ▶ ... but partly because of off-the-shelf analyses and designs

Publication bias costs evidence

Reproducibility Project: Psychology emphasized that **evidence is precious**

- Evidence in the RP:P did not indicate any clear inconsistencies
 - ▶ Evidence generally pointed in one direction, never opposite directions
 - ▶ There is **no need to assume hidden moderators** to explain RPP results
- Evidence in the Reproducibility Project depends greatly on sample size
 - ▶ Obviously because each observation carries some evidence
 - ▶ ... but partly because of off-the-shelf analyses and designs (ideally, sometimes we would apply better designs and analyses that squeeze more evidence out of each observation)

Publication bias costs evidence

Reproducibility Project: Psychology emphasized that **evidence is precious**

- Evidence in the RP:P did not indicate any clear inconsistencies
 - ▶ Evidence generally pointed in one direction, never opposite directions
 - ▶ There is **no need to assume hidden moderators** to explain RPP results
- Evidence in the Reproducibility Project depends greatly on sample size
 - ▶ Obviously because each observation carries some evidence
 - ▶ ... but partly because of off-the-shelf analyses and designs (ideally, sometimes we would apply better designs and analyses that squeeze more evidence out of each observation)
- Publication bias decreases evidence

Publication bias costs evidence

Reproducibility Project: Psychology emphasized that **evidence is precious**

- Evidence in the RP:P did not indicate any clear inconsistencies
 - ▶ Evidence generally pointed in one direction, never opposite directions
 - ▶ There is **no need to assume hidden moderators** to explain RPP results
- Evidence in the Reproducibility Project depends greatly on sample size
 - ▶ Obviously because each observation carries some evidence
 - ▶ ... but partly because of off-the-shelf analyses and designs (ideally, sometimes we would apply better designs and analyses that squeeze more evidence out of each observation)
- Publication bias decreases evidence
 - ▶ Obviously because some presumably good work goes unpublished

Publication bias costs evidence

Reproducibility Project: Psychology emphasized that **evidence is precious**

- Evidence in the RP:P did not indicate any clear inconsistencies
 - ▶ Evidence generally pointed in one direction, never opposite directions
 - ▶ There is **no need to assume hidden moderators** to explain RPP results
- Evidence in the Reproducibility Project depends greatly on sample size
 - ▶ Obviously because each observation carries some evidence
 - ▶ ... but partly because of off-the-shelf analyses and designs (ideally, sometimes we would apply better designs and analyses that squeeze more evidence out of each observation)
- Publication bias decreases evidence
 - ▶ Obviously because some presumably good work goes unpublished
 - ▶ ... but also in the published work itself!
- A parsimonious explanation for the failure to replicate these effects is that the **evidence for their existence was weak in the first place**

Some (current and recent) benefactors



JOHN TEMPLETON

FOUNDATION



These most excellent people

My students Maime Guan, Alexander Etz, and Beth Baribault.



References

- Etz, A., & Vandekerckhove, J. (2016). A Bayesian perspective on the Reproducibility Project: Psychology. *PLoS ONE*, *11*, e0149794.
- Guan, M., & Vandekerckhove, J. (2016). A Bayesian approach to mitigation of publication bias. *Psychonomic Bulletin & Review*, *23*(1), 74–86.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.
- Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association*, *54*, 30–34.
- Young, N. S., Ioannidis, J. P., & Al-Ubaydli, O. (2008). Why current publication practices may distort science. *PLoS Medicine*, *5*(10), e201.