

„Low replicability can support
robust and efficient science“

OSIP Journal Club

Robert Miller

16.05.2020

Worum geht es heute?

ARTICLE

<https://doi.org/10.1038/s41467-019-14203-0>

OPEN

Low replicability can support robust and efficient science

Stephan Lewandowsky ^{1,2*} & Klaus Oberauer ³



[...] Here we examine potential solutions by modeling a scientific community under various different replication regimes. In one regime, all findings are replicated before publication to guard against subsequent replication failures. In an alternative regime, individual studies are published and are replicated after publication, but only if they attract the community's interest.

We find that the publication of potentially non-replicable studies minimizes cost and maximizes efficiency of knowledge gain for the scientific community under a variety of assumptions. Provided it is properly managed, our findings suggest that low replicability can support robust and efficient science.

Worum geht es heute?

ARTICLE

<https://doi.org/10.1038/s41467-019-14203-0>

OPEN

Low replicability can support robust and efficient science

Stephan Lewandowsky ^{1,2*} & Klaus Oberauer ³

Dreber, A. et al. Using prediction markets to estimate the reproducibility of scientific research. *Proc. Natl Acad. Sci. USA* 112, 15343–15347 (2015)

Oberauer, K. & Lewandowsky, S. Addressing the theory crisis in psychology. *Psychon. Bull. Rev.* 26, 1596–1618 (2019).



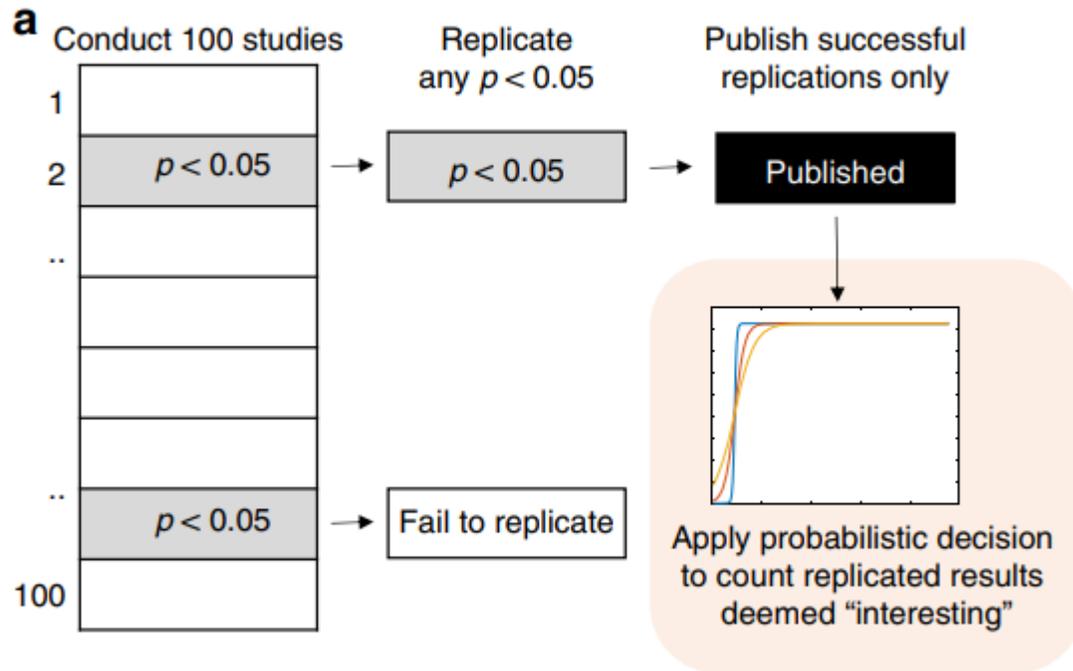
Basisrate wahrer
psychologischer Hypothesen:
 $\Pr(H1) = 9\%$

Exploration (discovery):
Nur “signifikante” Ergebnisse sind
informativ

Konfirmation (theory-testing):
Zuverlässig “nicht-signifikante”
Ergebnisse sind auch informativ

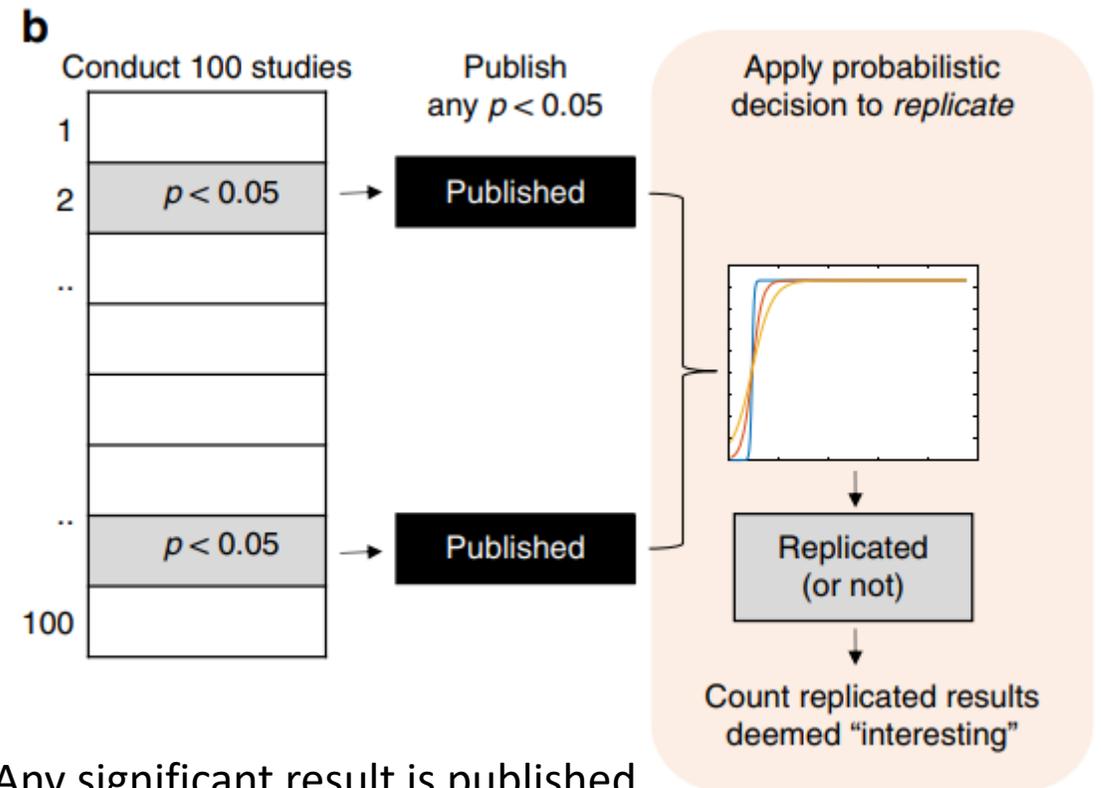
Wie die Replikationskrise begegnen?

Private replication regime



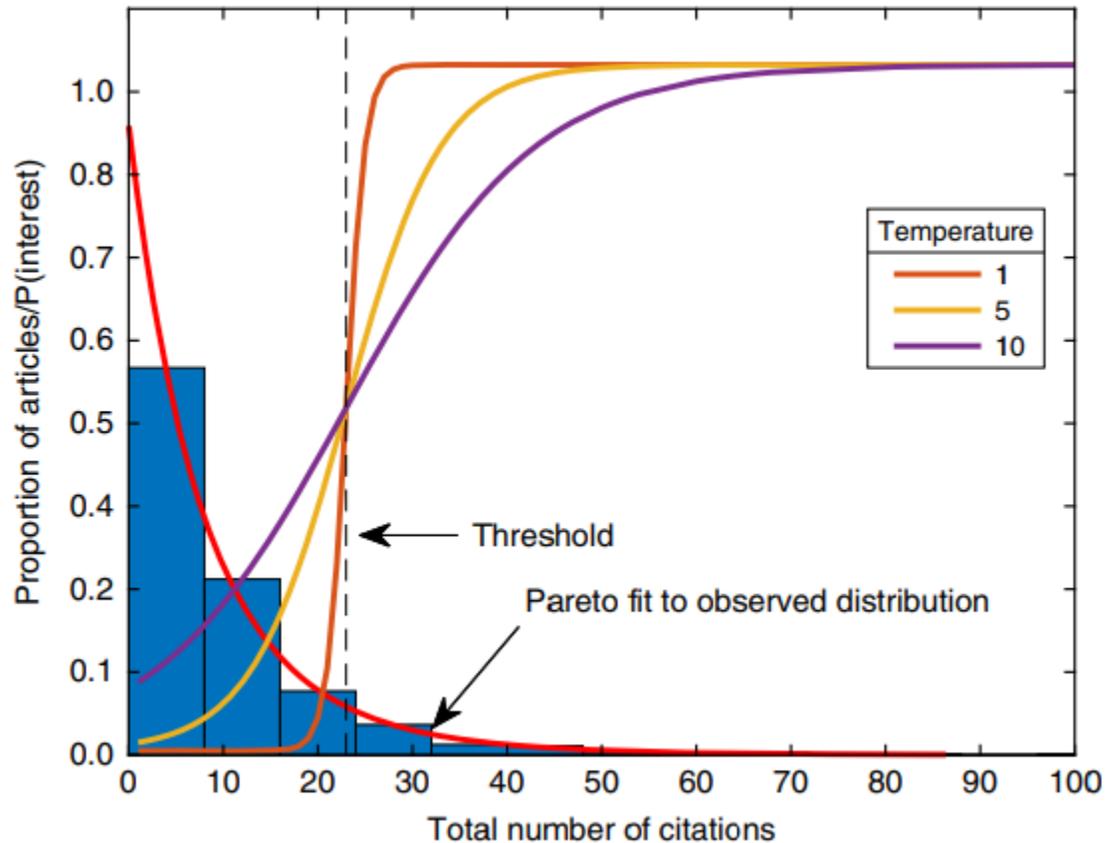
- Investigator replicates any significant result
- Publication If replication successful
- Peers decide which publication is interesting

Public replication regime



- Any significant result is published
- Peers decide which publication is interesting and hence worthy of replication

Wie entscheidet die wissenschaftliche Gemeinschaft was interessant ist?



All psychology articles published in 2014 (Scopus)

Logistische Response Funktion

$$P(I_k) = \frac{1}{1 + e^{-(n_k - q)/t}}$$

- $P(I_k)$: Wahrscheinlichkeit für Interesse an Publikation k
- n_k : Zitationshäufigkeit der Publikation k
- t : "Temperatur" Parameter
- q : "Schwellen" Parameter ($P = 0.5$)

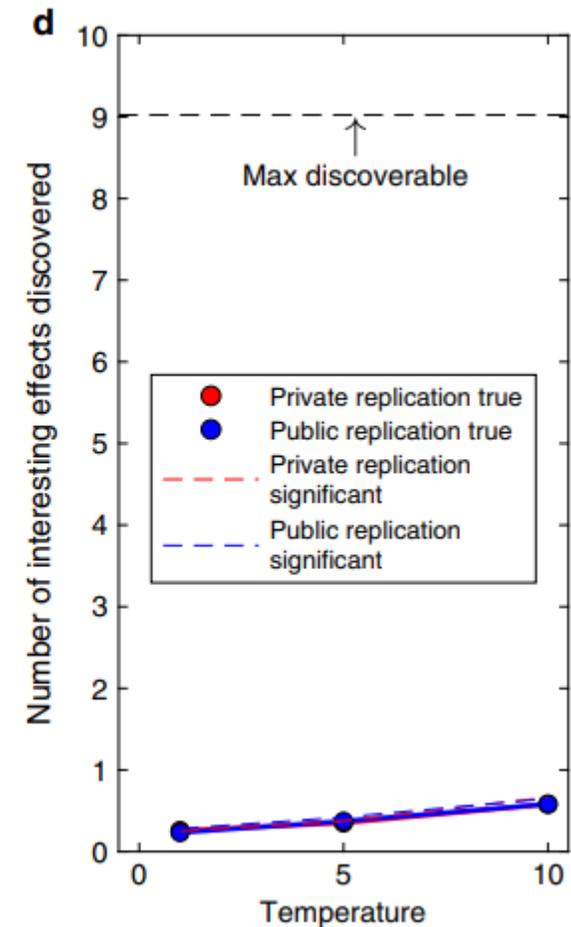
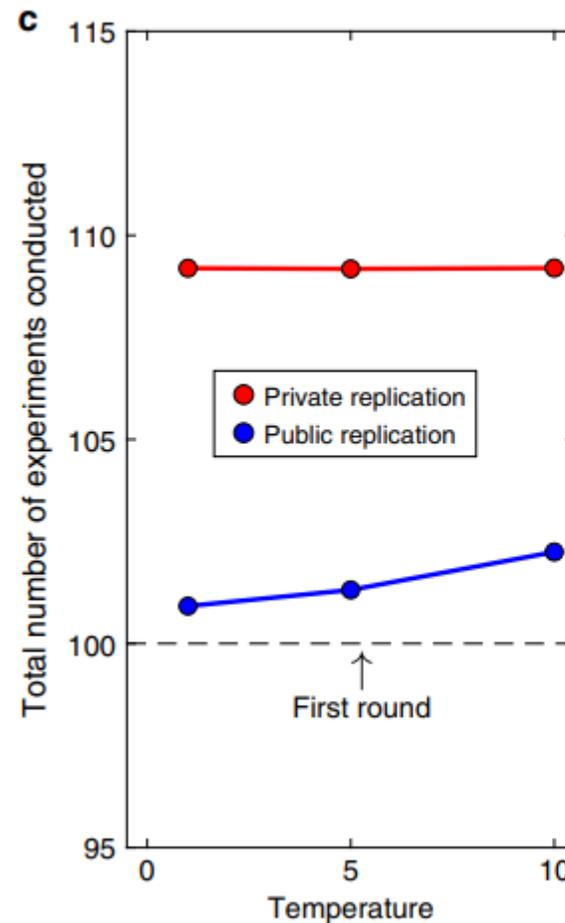
Simulation Explorative Forschung

Simulationsparameter:

- 1000 Iterationen
- 100 Primärstudien pro Iteration
- Basisrate H1 = 9%
- Power $P(H1 | H1)$: 80% vs. 50%
- p-hacking $P(H1 | H0) = 5%$ vs. 20%
- **Temperatur der Interesse-Funktion**
- Schwelle der Interesse-Funktion:
Zitationsquantil 10% - 90% (= 22.98 Zitationen)

Endpunkte pro 100 Studien:

- Studienkosten (N Experimente)
- Effektivität (Anzahl H1 repliziert)



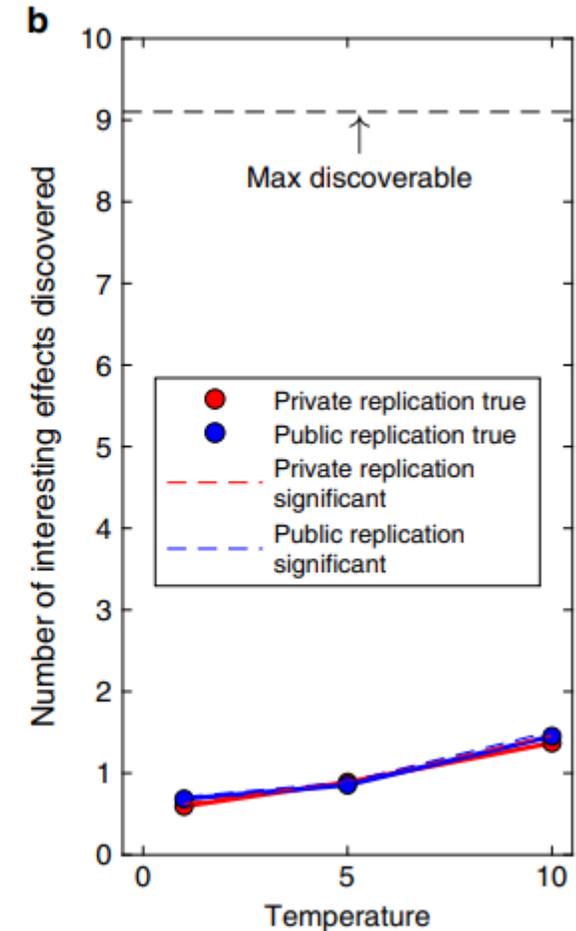
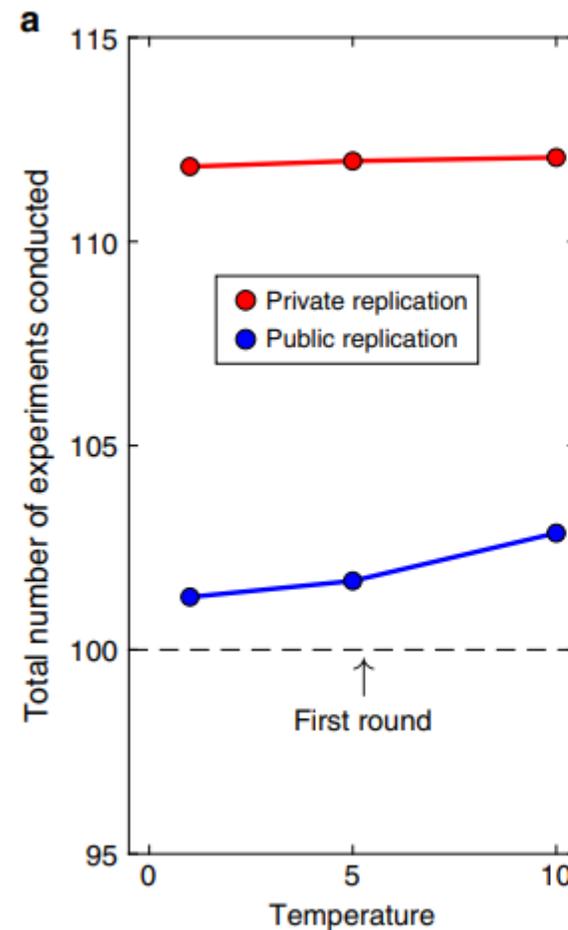
Simulation Explorative Forschung

Simulationsparameter:

- 1000 Iterationen
- 100 Primärstudien pro Iteration
- Basisrate H1 = 9%
- Power $P(H1 | H1)$: 80% vs. 50%
- p-hacking $P(H1 | H0)$ = 5% vs. 20%
- **Temperatur der Interesse-Funktion**
- Schwelle der Interesse-Funktion:
Zitationsquantil 10% - 90% (= 22.98 Zitationen)

Endpunkte pro 100 Studien:

- Studienkosten (N Experimente)
- Effektivität (Anzahl H1 repliziert)



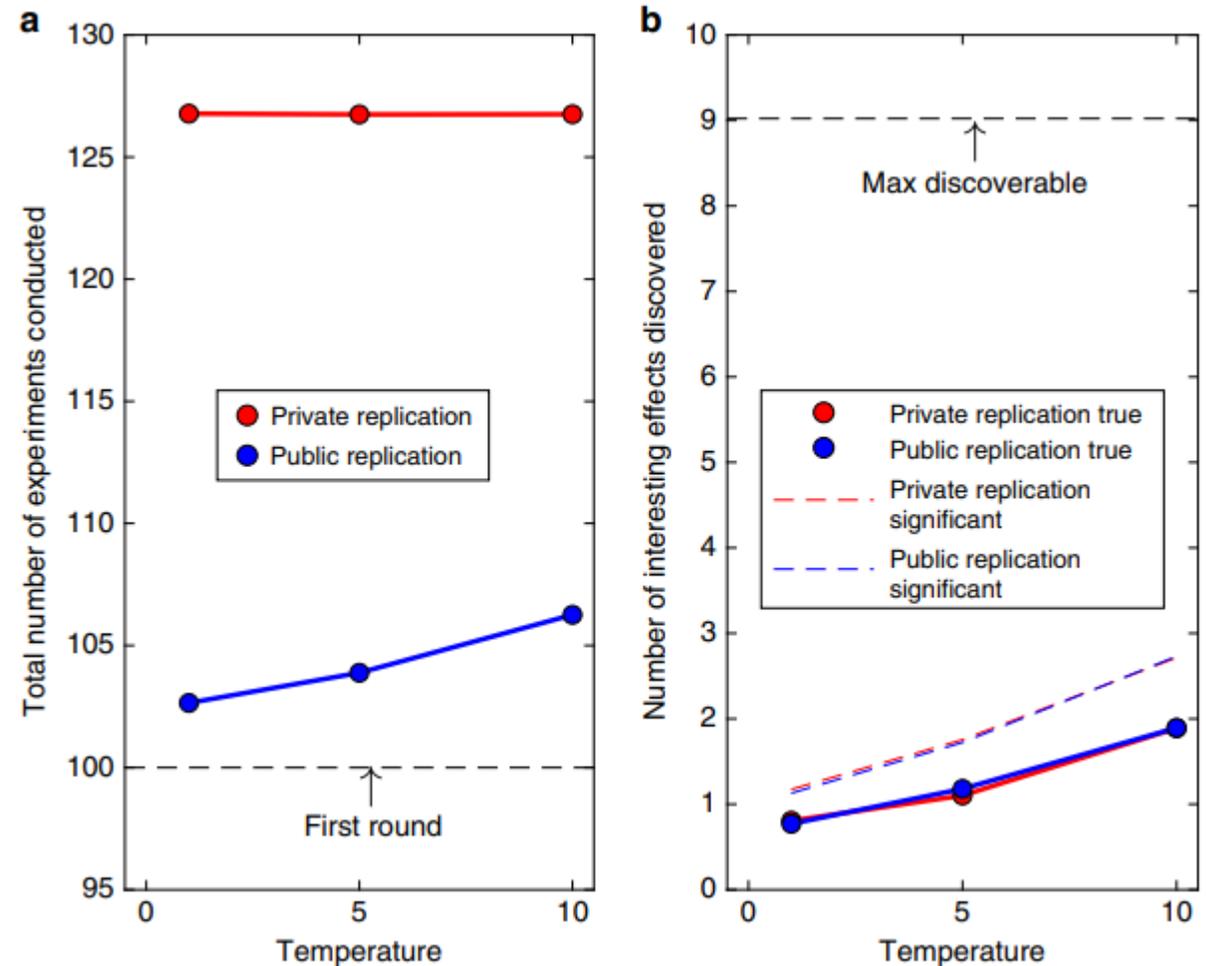
Simulation Explorative Forschung

Simulationsparameter:

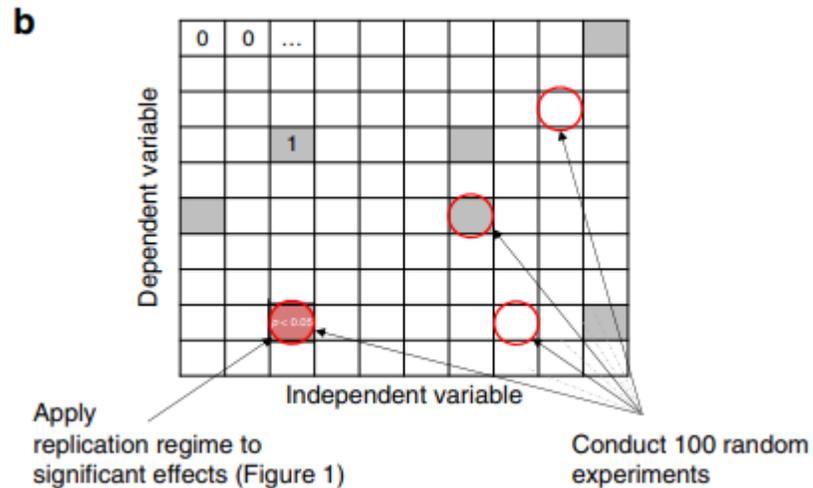
- 1000 Iterationen
- 100 Primärstudien pro Iteration
- Basisrate H1 = 9%
- Power $P(H1 | H1)$: 80% vs. 50%
- p-hacking $P(H1 | H0)$ = 5% vs. 20%
- Temperatur der Interesse-Funktion
- Schwelle der Interesse-Funktion:
Zitationsquantil 10% - 90% (= 22.98 Zitationen)

Endpunkte pro 100 Studien:

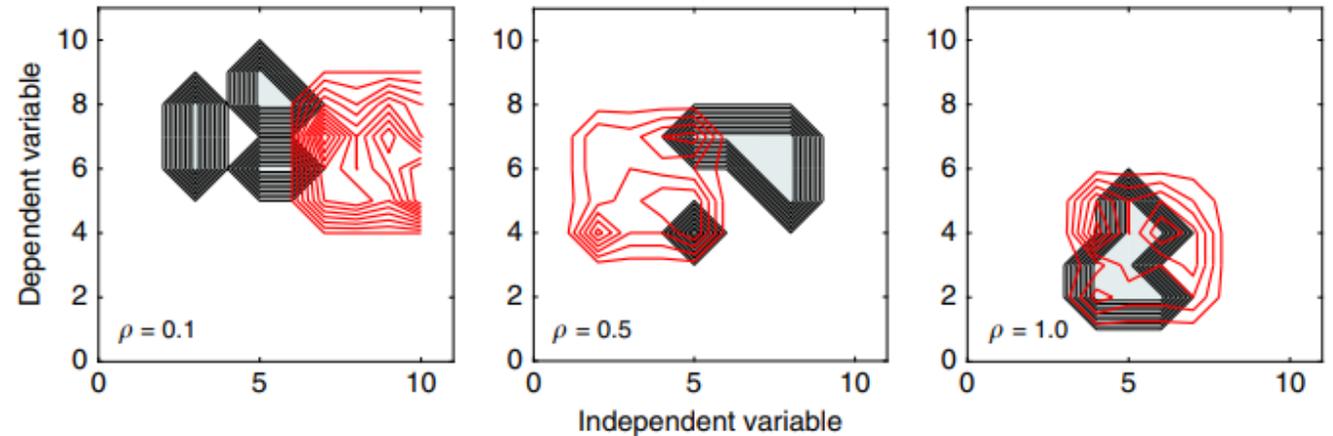
- Studienkosten (N Experimente)
- Effektivität (Anzahl H1 repliziert)



Explorative vs. Konfirmatorische Forschung



Keine Theorie



Miese Theorie

Brauchbare Theorie

Perfekte Theorie

Die Qualität einer Theorie wird

- bestimmt durch Parameter ρ
- visualisiert durch die Überlappung zwischen wahrer Beschaffenheit der Welt und Experimenten

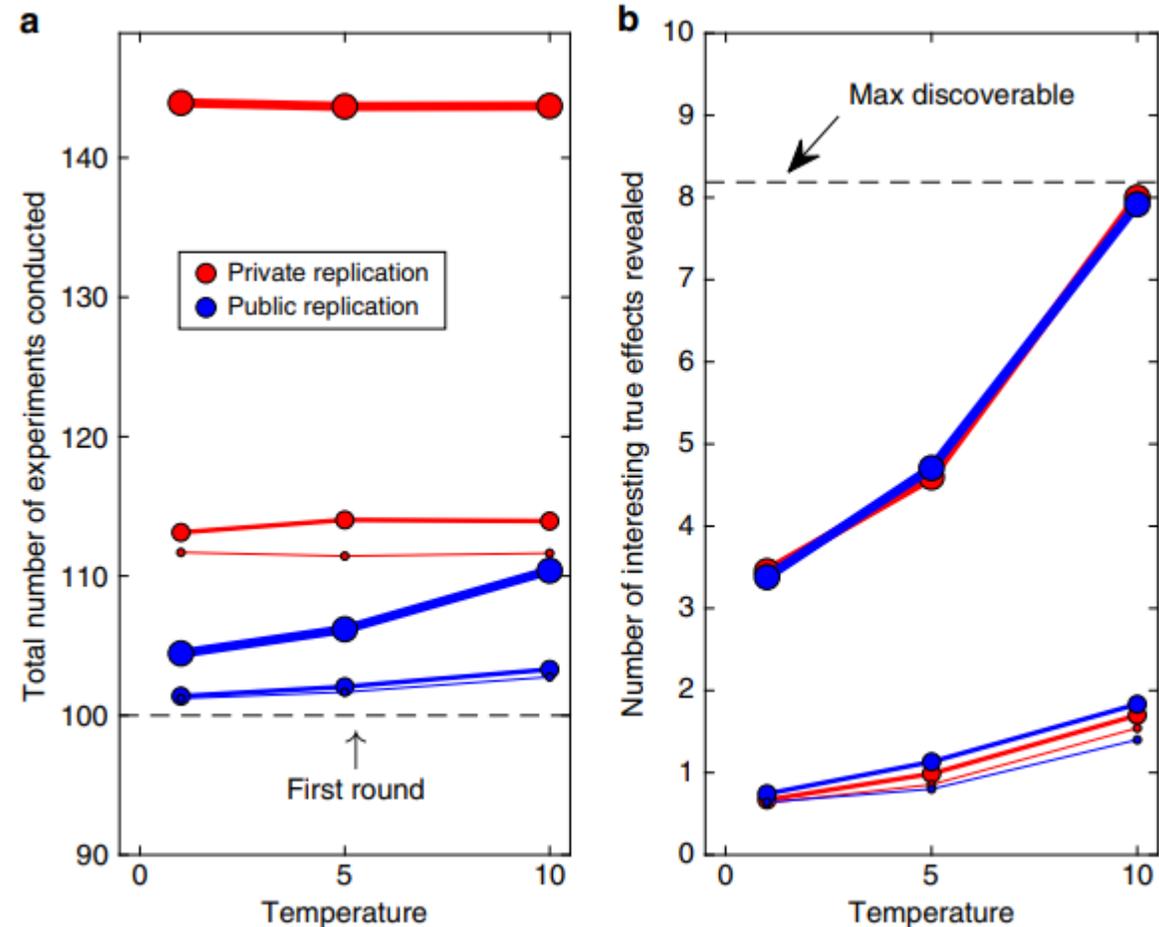
Simulation Konfirmatorische Forschung

Simulationsparameter:

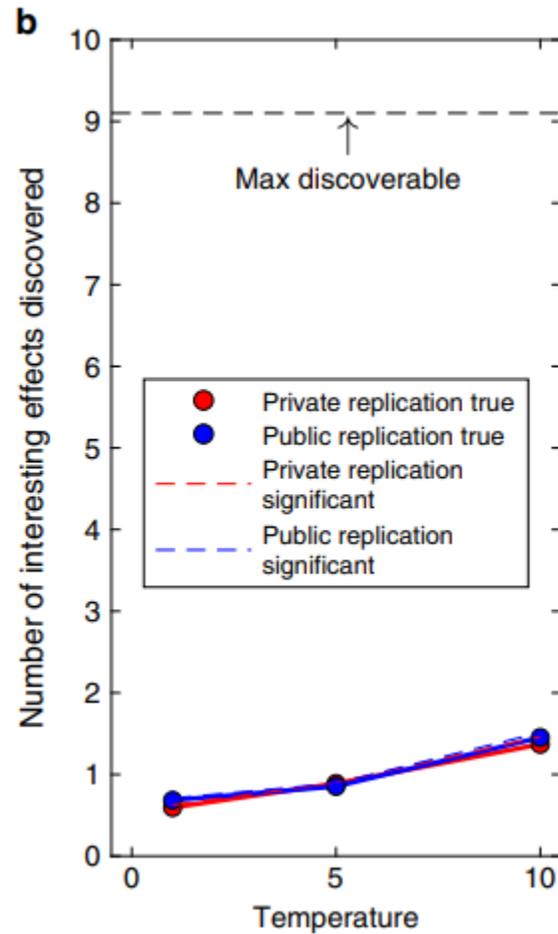
- 1000 Iterationen
- 100 Primärstudien pro Iteration
- Basisrate H1 = 9%
- Power $P(H1 | H1)$: 80% vs. 50%
- p-hacking $P(H1 | H0) = 5\%$ vs. 20%
- **Temperatur der Interesse-Funktion**
- Schwelle der Interesse-Funktion:
Zitationsquantil 10% - 90% (= 22.98 Zitationen)
- **Qualität der Theorie ρ : Dickere Linie = bessere Theorie**

Endpunkte pro 100 Studien:

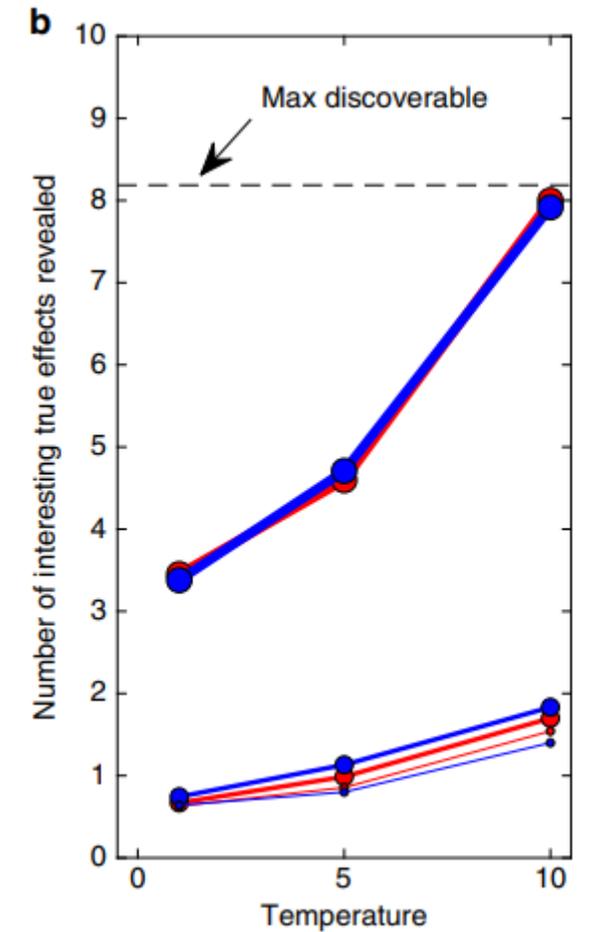
- Studienkosten (N Experimente)
- Effektivität (Anzahl H1 repliziert)



Explorative vs. Konfirmatorische Forschung

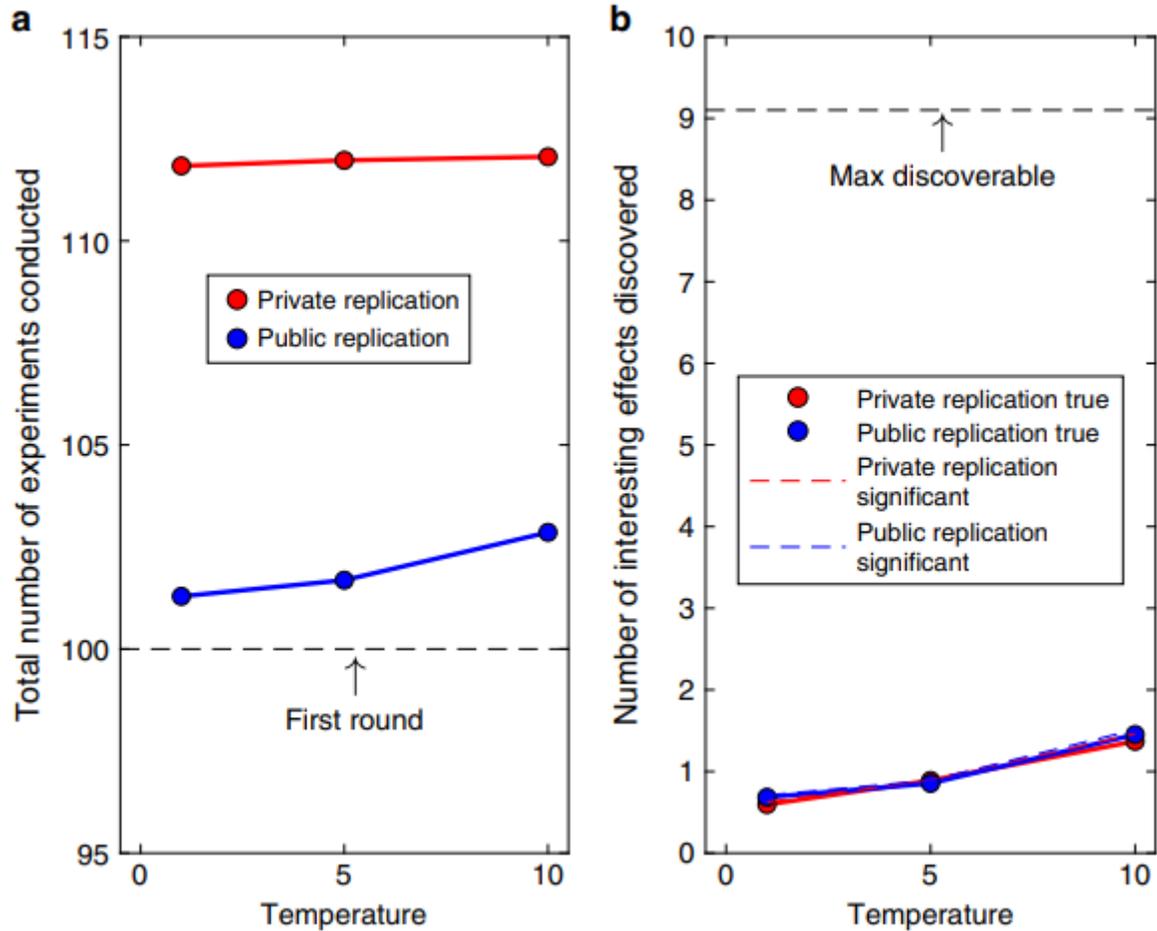


Exploration

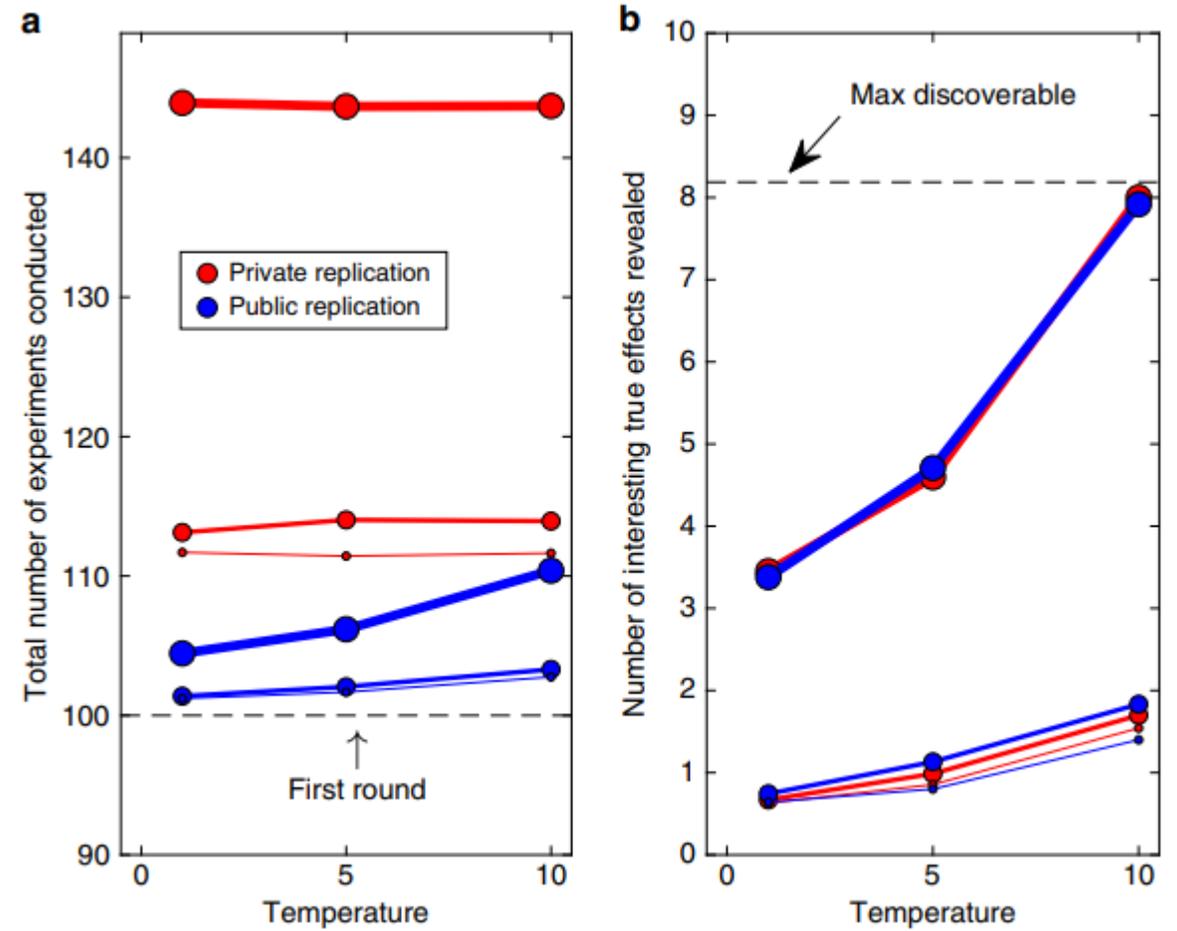


Konfirmation

Explorative vs. Konfirmatorische Forschung

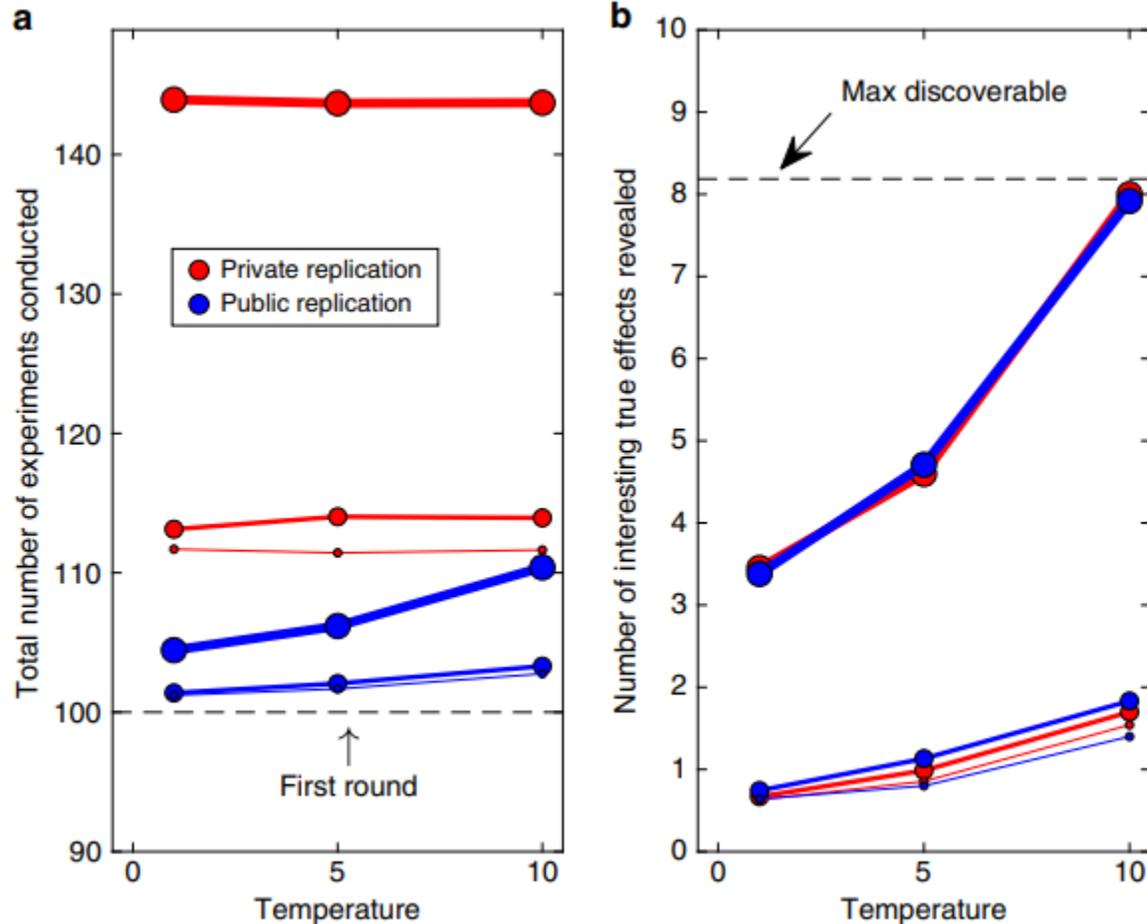


Exploration



Konfirmation

Schlussfolgerung der Autoren



Public > Private Replikation, wenn

1. Embargozeit gegenüber Medienberichten: 1 Jahr
2. Deklaration von Befunden als „provisorisch“ bis Ablauf Embargo oder Replikationsversuch
3. Bei Replikation: Replikateure werden Koautoren auf Original Publication
4. Bei Nicht-Replikation: Withdrawal + Citable Replication Acknowledgement
5. Bei Ablauf Embargo: Archivierung als „kein Replikationsinteresse“

Ein Gedankenpiel..

Szenario: Sie arbeiten in einem Forschungsfeld mit hervorragender theoretischer Fundierung, welches Sie aufgrund der Breite der Thematik jedoch nur schwer überblicken können.

- Sie konzipieren gerade ein neues Experiment, stellen jedoch fest dass Ihnen für das adäquate Design Ihrer Stimuli bestimmte Informationen fehlen.
- Ein Google-Scholar Alert zeigt Ihnen einen neuen, aber niedrig publizierten Artikel, der Ihnen genau diese fehlenden Informationen zu liefern scheint.
- Wie wahrscheinlich ist es, dass Sie ihre Stimuli auf Basis dieser neuen Informationen designen?

(Score: 0: „nur mit eigener Pilotstudie “ bis 5: „direkt übernommen“)

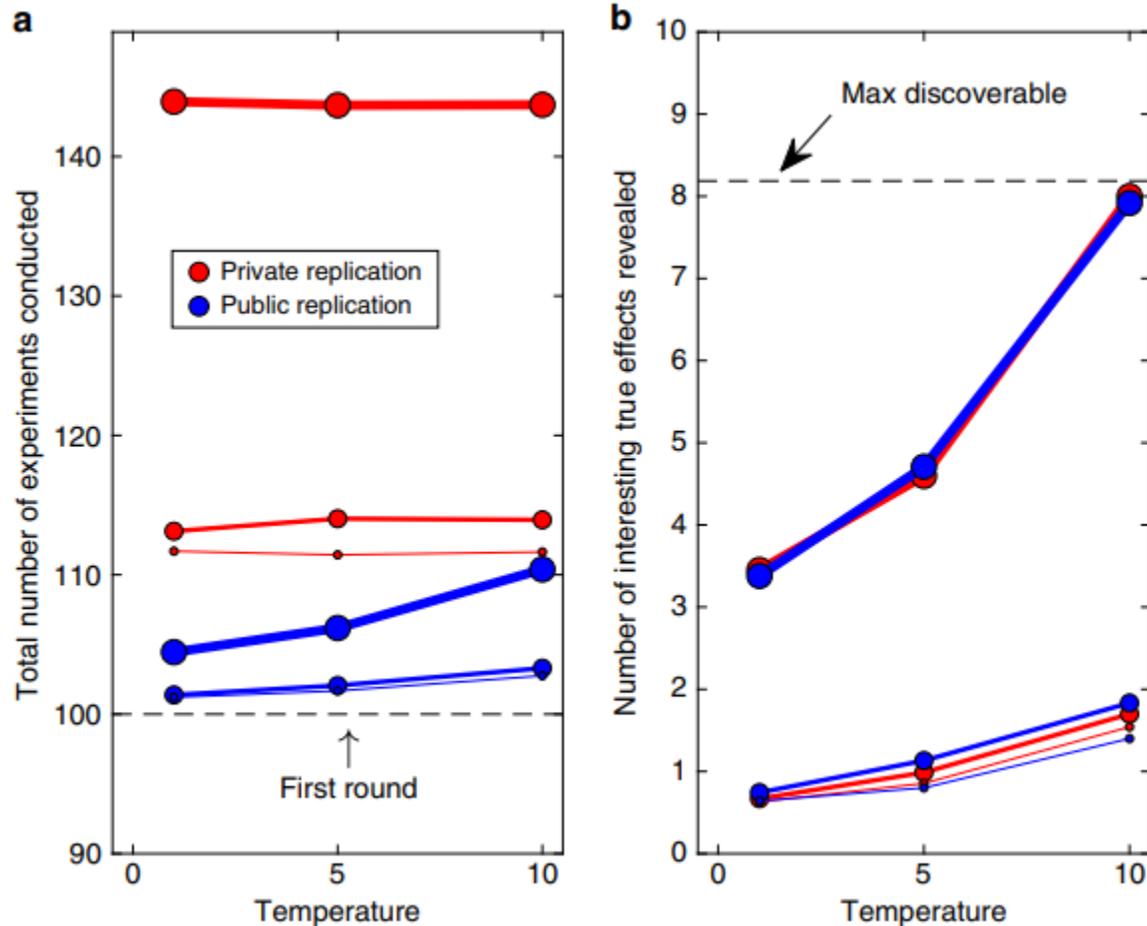
Ein weiteres Gedankenexperiment..

Szenario: Sie arbeiten in einem Forschungsfeld mit hervorragender theoretischer Fundierung, welches Sie aufgrund der Enge der Thematik relativ gut überblicken können.

- Sie konzipieren gerade ein neues Experiment, stellen jedoch fest dass Ihnen für das adäquate Design Ihrer Stimuli bestimmte Informationen fehlen.
- Ein Google-Scholar Alert zeigt Ihnen einen neuen, aber niedrig publizierten Artikel, der Ihnen genau diese fehlenden Informationen zu liefern scheint.
- Wie wahrscheinlich ist es, dass Sie ihre Stimuli auf Basis dieser neuen Informationen designen?

(Score: 0: „nur mit eigener Pilotstudie “ bis 5: „direkt übernommen“)

Unberücksichtigte Endpunkte + Kritik



- Anzahl von Publikationen im System:
 - Private R.: < 10 pro 100 Primärstudien
 - Public R.: 100 pro 100 Primärstudien
- Assoziation zwischen
 - Anzahl Publikationen im System
 - „Temperatur“ / „Schwellen“ Parameter
- Kosten der Qualitätssicherung (Peer-Review) + Kosten der Veröffentlichung (Lizenzverträge) vs. Studienkosten

Diskussionszeit!

