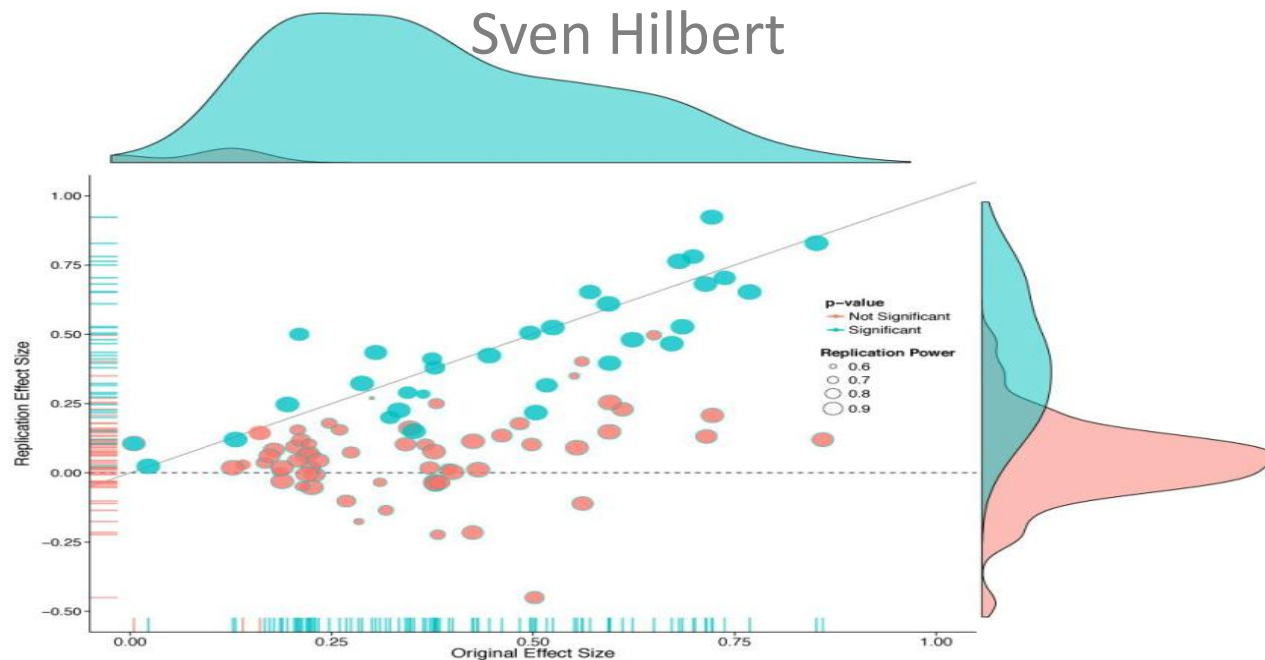


Open Science vor dem Hintergrund der Replikationskrise





Universität Regensburg

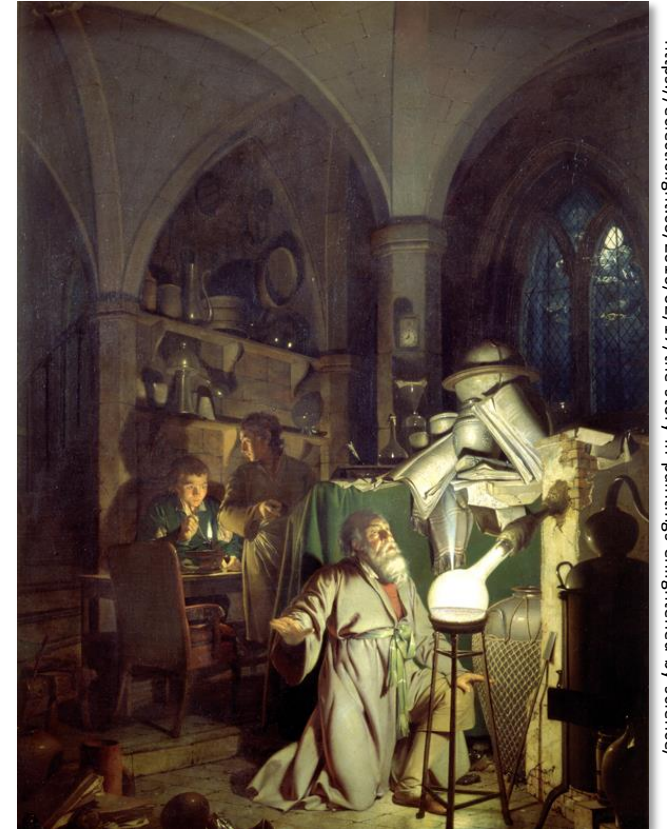
Wie steht es um die Wissenschaft?

Ob die Wissenschaft noch Wissen schafft?

Ziele der Wissenschaft

- Wissen schaffen
 - Im Idealfall zur Verbesserung der Welt
- Verlässliche Regeln aufstellen

- Empirische Wissenschaften stehen vor einer besonderen Herausforderung
 - Es kann nicht nur logisch gearbeitet werden
 - Ein Gegenbeispiel invalidiert keine Theorie
 - Statistische Annahmen werden getroffen



<https://electrictlight.co/2016/02/17/the-story-in-paintings-enlightened-by-science/>

Wissenschaft als Gemeinschaft

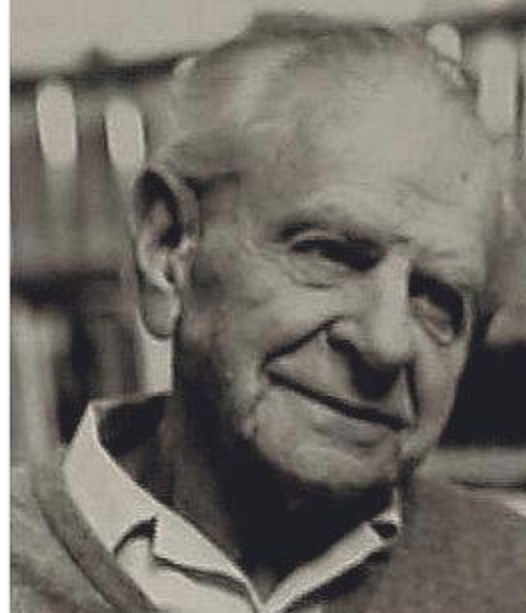
- **Gemeinsames Schaffen von Wissen**
 - Finanziert von der Gesellschaft
- **Eine Gemeinschaft muss transparent sein**
 - Glaubwürdigkeit ist Teil von geschaffenem Wissen
- **Eine Gemeinschaft muss Fehler machen können**
 - Wissen kann nur durch einen Lernprozess geschaffen werden
- **Fehler müssen gemeinsam korrigiert werden können**
 - Einer der wichtigsten Aspekte: Lernen aus den Fehlern



Wichtig für die Qualitätskontrolle: Transparenz

Fundamentale Aspekte:

- Methodische Transparenz
- Reproduzierbarkeit der Analyse und der Ergebnisse
- Replizierbarkeit der Effekte



"In so far as a scientific statement speaks about reality, it must be falsifiable; and in so far as it is not falsifiable, it does not speak about reality."

Karl Popper

<https://warosu.org/lit/thread/5921170>

➤ **Minimalstandards** brauchbarer Forschung

- Ohne Transparenz, Reproduzierbarkeit und Replizierbarkeit ist keine Falsifizierbarkeit möglich!

Forschung in den Medien

- Wissenschaftliche Magazine werden zitiert von populärwissenschaftlichen Magazinen

This Simple 'Power Pose' Can Change Your Life And Career

BI Henry Blodget, Business Insider
© 3.05.2013, 16:33

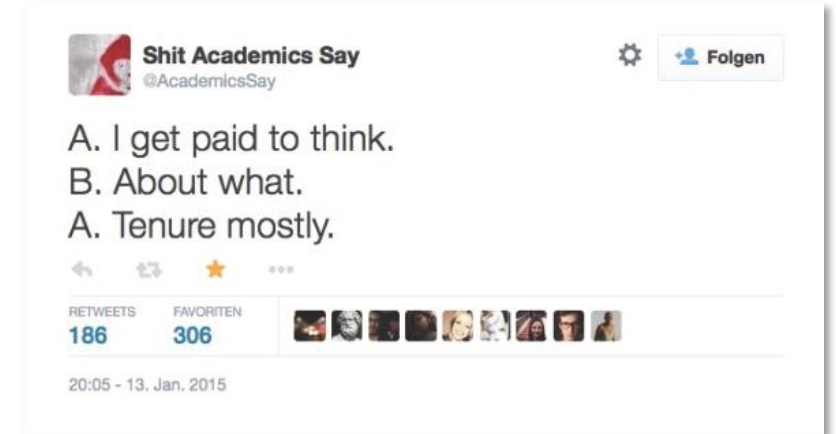
- Besonders aufregende Befunde werden in den allgemeinen Nachrichten übernommen



Anreizsystem der Wissenschaft

- Belohnungssystem der Wissenschaft

- Publish or Perish
- Einwerben von Geldern
- Spannende Ergebnisse produzieren



- Ein großer Teil der wissenschaftlichen Qualifikation wird durch Veröffentlichungen repräsentiert

- Wichtig hierbei: vor allem Anzahl der Arbeiten und Renommee der Journals

- Qualitätskontrolle organisiert wissenschaftliche Community

- Peer-Review Verfahren

Anreizsystem der Wissenschaft

- Top fünf Faktoren bei der Besetzung einer Professur

(Abele-Brehm & Bühner, 2016)

1. **Anzahl der Peer-Review Publikationen**
2. **Passung Forschungsprofil zum berufenden Institut**
3. **Qualität des Forschungsvortrags**
4. **Anzahl der Publikationen**
5. **Volumen bisher eingeworbener Drittmittel**

...

17. **Qualitätsbeurteilung der besten drei Publikationen**

- $n = 729$
- 66 % der Befragten war bereits Mitglied in einer Berufungskommission



<http://www.simpsoncrazy.com/dictionary>



Universität Regensburg

Was sind die Konsequenzen?

Genauer hinsehen schmerzt

Publication Bias

Hypothetische Situation zur Illustration des File-Drawer-Problems:

- **20 unabhängige Studien**, die untersuchen ob Schokolade die Intelligenz mindert
- **19 Negativbefunde** ($p > .05$)
 - Befunde nicht veröffentlicht, da nicht spannend
- **1 Positivbefund** ($p \leq .05$)
 - Befund wird veröffentlicht, da sehr spannend

ONE MORE NOSTALGIC
PART OF CHILDHOOD
GOES THBPPTH.

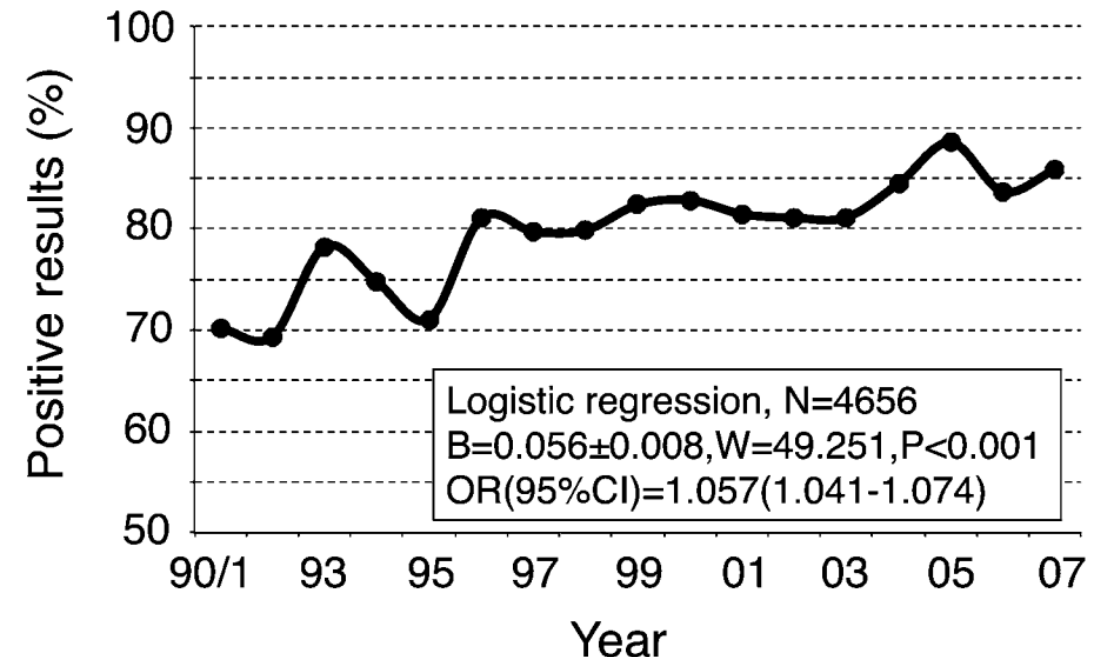


<https://www.gocomics.com/calvinandhobbes/1993/01/26>

➤ Öffentliche Wahrnehmung: **Schokolade macht dumm!**

Publication Bias in der Wissenschaft

- Wie stark ist der Bias wirklich?
 - Ioannidis und Trikalinos (2007)
 - Sechs von acht untersuchten Metaanalysen zu Neuroleptika
 - Egger, Smith, Schneider und Minder (1997)
 - 38% der untersuchten Metaanalysen im Bereich der Medizin
 - Francis, Tanzman und Matthews (2014)
 - 15 von 18 der untersuchten Artikel mit vier oder mehr Studien im Bereich der experimentellen Psychologie



Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891-904.

Die gegenwärtige Situation (die Krise)

- Replizierbarkeit sozialwissenschaftlicher Studien unter 50%
- Starker Einfluss des Publication Bias
 - Metaanalysen nutzen nur publizierte Studien
- Mangelhafte statistische Fähigkeiten der Wissenschaftler
- Image(-Management) macht einen großen Teil des Erfolges aus → Rockstars mit $n = 42$
- Viele Probleme entstanden und entstehen durch **Questionable Research Practices (QRPs)**

This is part of our series, "50 Women Who Are Changing The World." [redacted] ranked #37.

Everyone talks about the importance of "body language," but few people understand how much of an impact it actually has – not just in the way others perceive us, but in terms of how we actually perform.



Sven Hilbert

27. August 2015, 21:30 Uhr Psychologie

Viele Psychologie-Experimente lassen sich nicht wiederholen

- Ein internationales Team aus 270 Forschern hat 100 psychologische Experimente wiederholt.
- Bei über der Hälfte der Studien konnte das Ergebnis nicht repliziert werden.
- Um die Qualität von Studien zu erhöhen, müsste es andere Anreize geben.

<http://www.sueddeutsche.de/wissen/psychologie-stimmt-das-ueberhaupt-1.2623513>



<https://www.youtube.com/watch?v=0Rnq1NpHdmw>

The New York Times

TheUpshot

THE NEW HEALTH CARE

Science Needs a Solution for the Temptation of Positive Results

<https://www.nytimes.com/2017/05/29/upshot/science-needs-a-solution-for-the-temptation-of-positive-results.html>

#13

Leider ist nicht nur der Publication Bias schuld

Die Ergebnisse vieler Studien basieren auf Questionable Research Practices (QRPs)

- **Data dredging** oder ***p*-Hacking** kann leicht eingesetzt werden, um „statistisch signifikante“ Ergebnisse zu erzielen
- Jahr 2011 erschien (u. a.) der wegweisende Artikel „False-Positive Psychology“ von Simmons, Nelson und Simonsohn und prangerte die Situation an
 - Besonders die Konsequenzen für die gesamte Wissenschaft wurden deutlich:

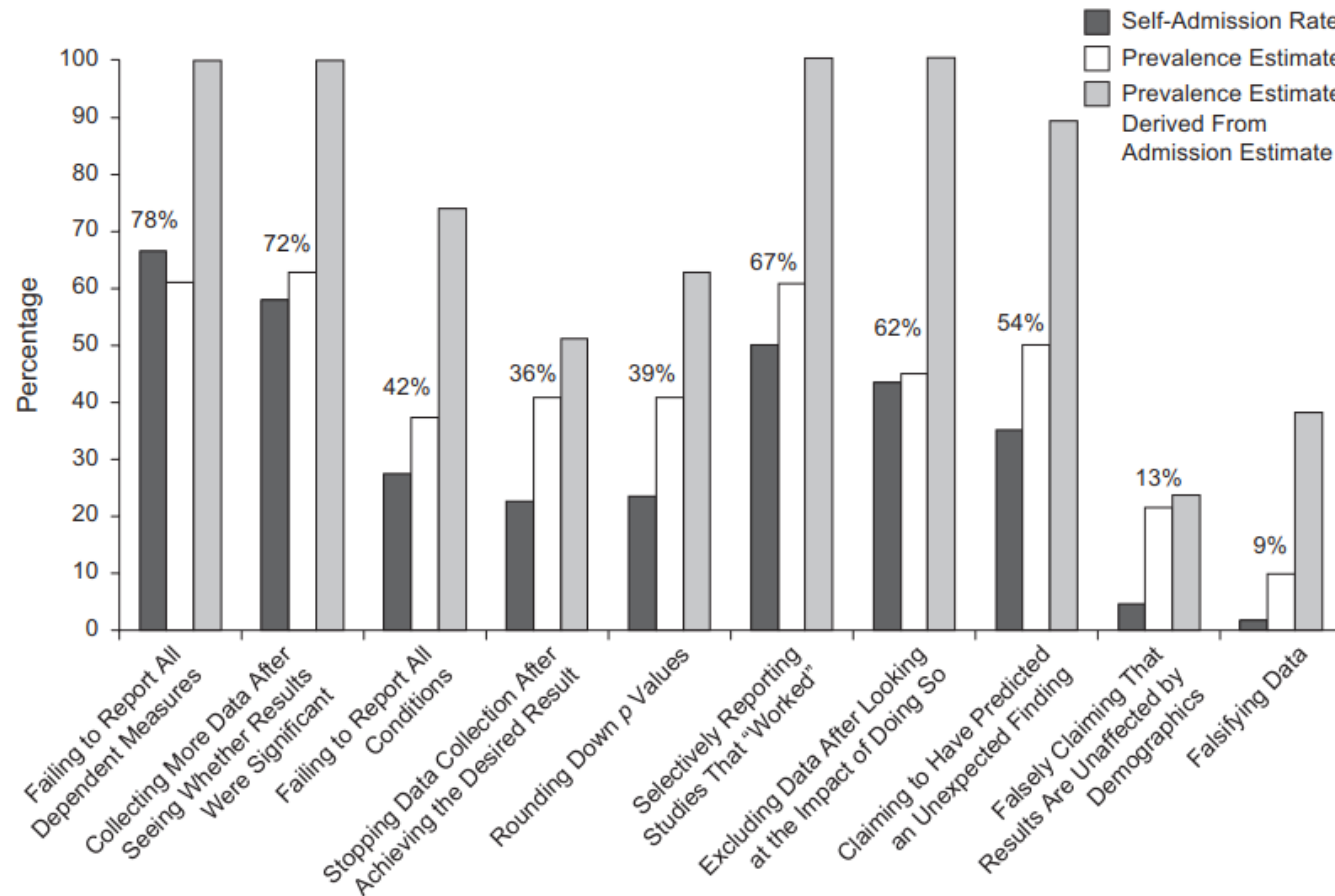
“Everyone knew it was wrong, but they thought it was wrong the way it’s wrong to jaywalk. We decided to write ‘False-Positive Psychology’ when simulations revealed it was wrong the way it’s wrong to rob a bank.” (Simmons, Leif & Simonsohn, 2017)

p -Hacking

Beispiele verbreiteter Praktiken (teilweise auch bezeichnet als *Cherry Picking*, *Fishing Expedition* oder *Data Snooping*)

- Mitmessen vieler Variablen
- Selektive Auswahl von Versuchsbedingungen
- Einschluss von Kovariablen
- Optional Stopping
- Selektiver Ausschluss von Fällen
- Beispiele methodischer Aufklärungsmaßnahmen
 - p -Curve
 - Funnel Plot
 - Test for Excess Significance (TES)

For Vohs et al. (2006), “the authors conducted two additional money priming studies that showed no effects, the details of which were shared with us.” and “**reported nine dependent measures that were statistically affected by the manipulation** in the predicted direction (one in each experiment) but **did not report 19 additional measures** that were statistically unchanged”.

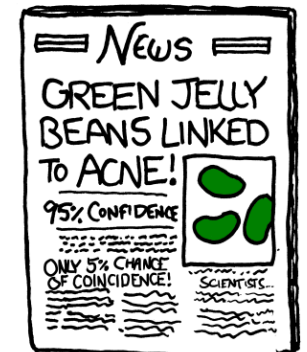
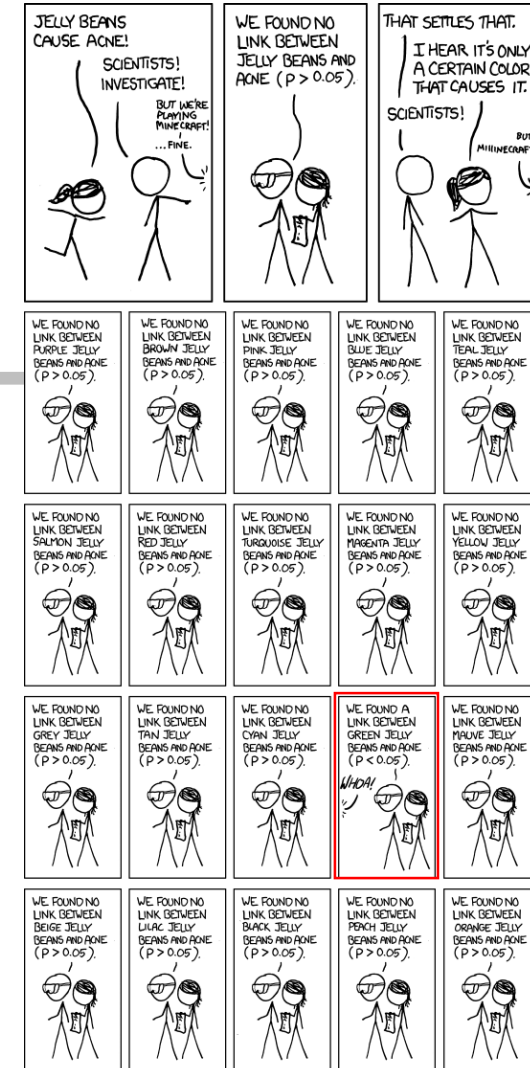


Studie von John, Loewenstein und Prelec (2012) zur Selbsteinschätzung von 2000 Psychologen, welche Praktiken sie schon angewandt haben.

➤ Alle diese sind QRPS

Cherry Picking

- Wenn nur signifikante Ergebnisse von Untersuchungen veröffentlicht werden, entspricht die Sicherheitswahrscheinlichkeit des Ergebnisses nicht den Angaben
- Gilt auch wenn Bedingungen weggelassen werden
- Fällt im Besonderen ins Gewicht, wenn der Fehler 1. Art nicht korrigiert wird!
- Remember:



Folglich halten viele Befunde keinen Replikationen stand - Replikationskrise

- Beispiele
 - Money Priming
 - Powerpose
 - Romantic Priming



<https://www.moviepilot.de/news/dragon-ball-super-promo-teaser-enthullt-neue-super-sayajin-stufe-176968>

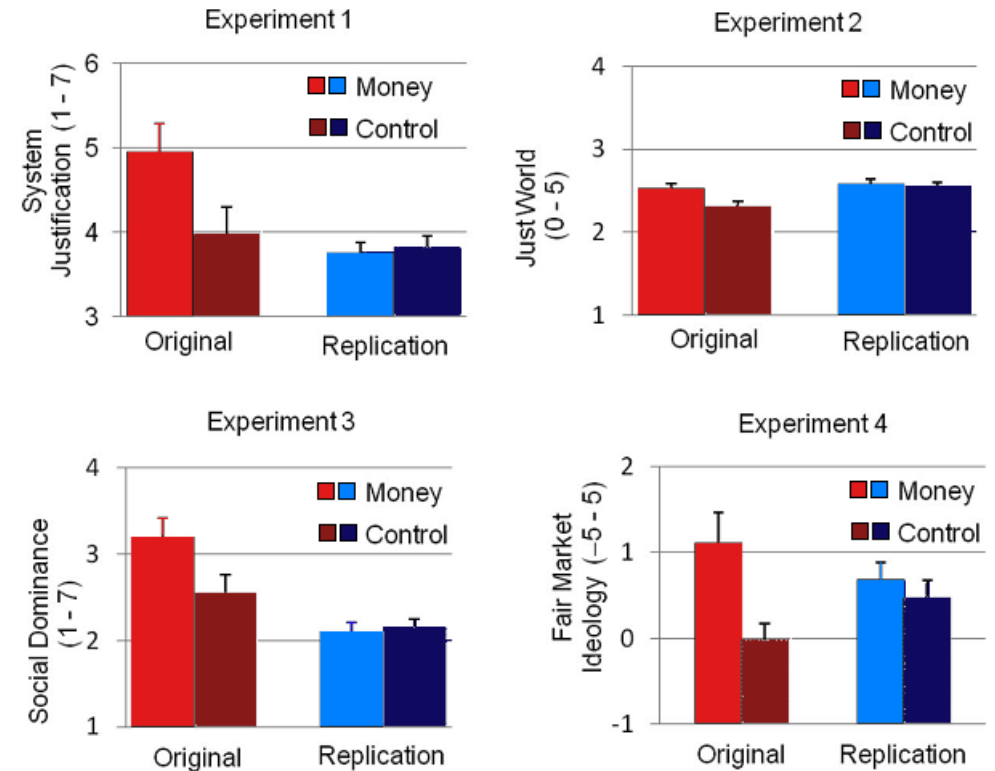


Figure 3. Results of the original studies (Caruso et al., 2013) and the replication attempts presented here. Priming effects were observed in each of the original studies but in none of the replication attempts. Error bars represent one standard error.

Rohrer et al. (2015) *JEPG*

Was macht die wissenschaftliche Community dagegen?

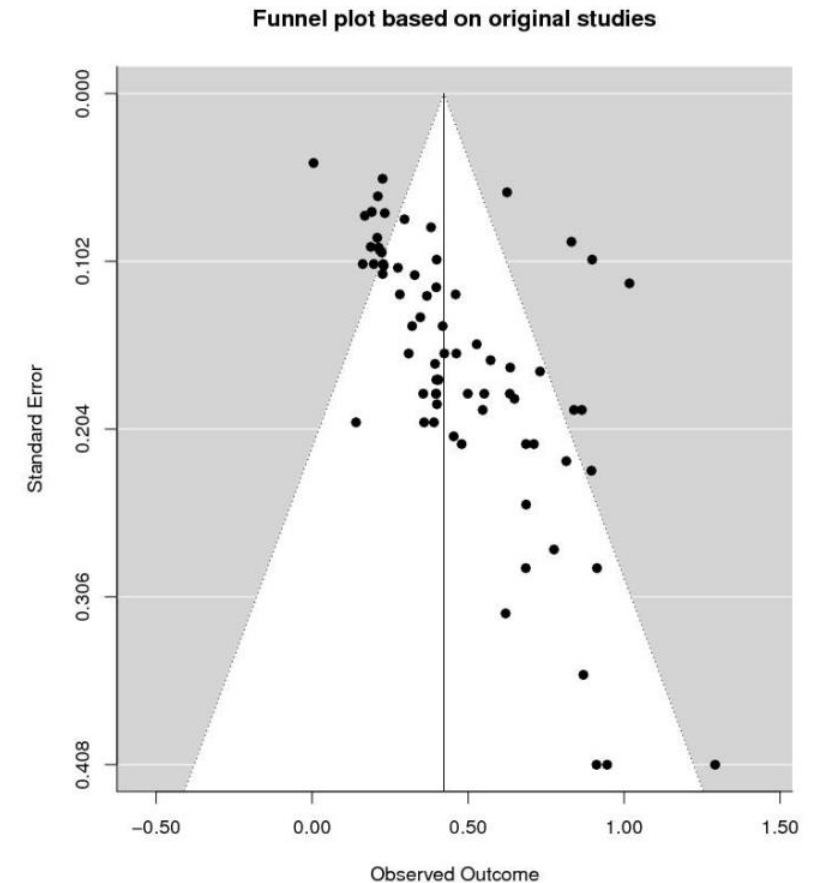
Die Methodiker schlagen zurück

Wege zur Identifikation schwacher Befundlagen

- Mittlerweile existiert eine Vielzahl statistischer Verfahren, welche auf einen „Bias“ in den veröffentlichten Studien eines Bereichs hinweisen können
 - Sie basieren Grundsätzlich auf Wissen über die Verteilungen statistischer Kennwerte unter Annahmen bzgl. Effektstärken und Teststärke (Power)
- Drei Beispiele dieser Verfahren
 - Funnel Plot
 - Test for Excess Significance
 - p -Curve

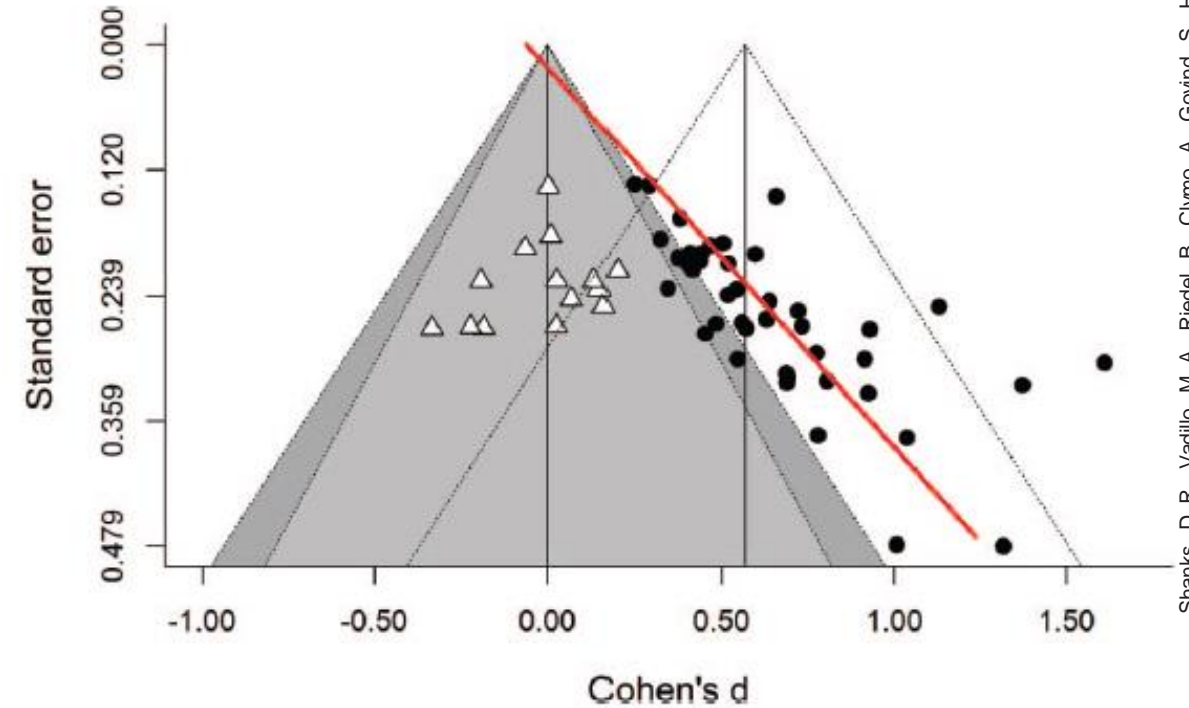
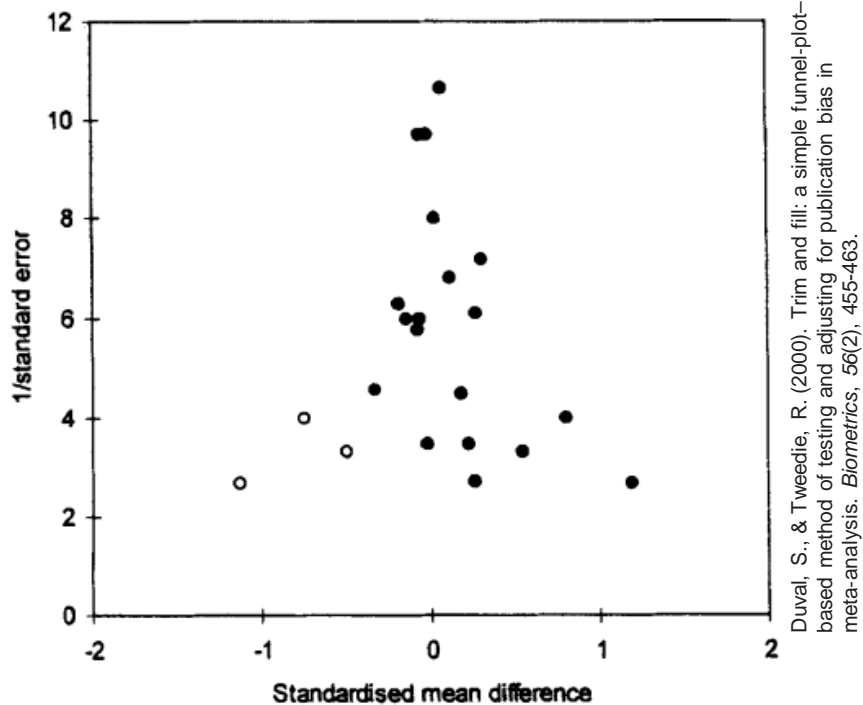
Funnel Plot

- In der Inferenzstatistik besteht ein direkter Zusammenhang zwischen folgenden drei Größen:
 1. Standardfehler
 2. Effektstärke
 3. p -Wert
- Einfach gesagt: Je kleiner der Standardfehler, desto kleiner zufällige Abweichungen der Stichprobeneffekte vom Populationseffekt
 - Der **Funnel Plot** illustriert diesen Zusammenhang



Funnel Plot

- Funnel Plots können also auf Unstimmigkeiten im Zusammenhang zwischen Effektstärke und Standardfehler hinweisen



Shanks, D. R., Vadillo, M. A., Riedel, B., Clymo, A., Govind, S., Hickin, N., ... & Puhmann, L. (2015). Romance, risk, and replication: Can consumer choices and risk-taking be primed by mating motives?. *Journal of Experimental Psychology: General*, 144(6), e142.

Quantitative Verfahren

- Funnel Plot ist qualitativ
- Quantitative Verfahren
 - Trim and Fill Technique (Duval & Tweedie, 2000): Nonparametrische Methode zur Schätzung fehlender Studien
 - Regressionsbasierter Ansatz (Egger, Smith, Schneider & Minder, 1997)
 - **Test for Excessive Significance (Ioannidis & Trikalinos, 2007)**
 - **p -Curve (Simonsohn, Nelson & Simmons, 2014)**
- Zu beachten:
 - Asymmetrie kann auch durch Heterogenität zwischen Studien entstehen
 - Wenn die Anzahl an Studien gering ist, ist die Power für Signifikanztests auf Funnel-Plot-Asymmetrie auch nicht groß

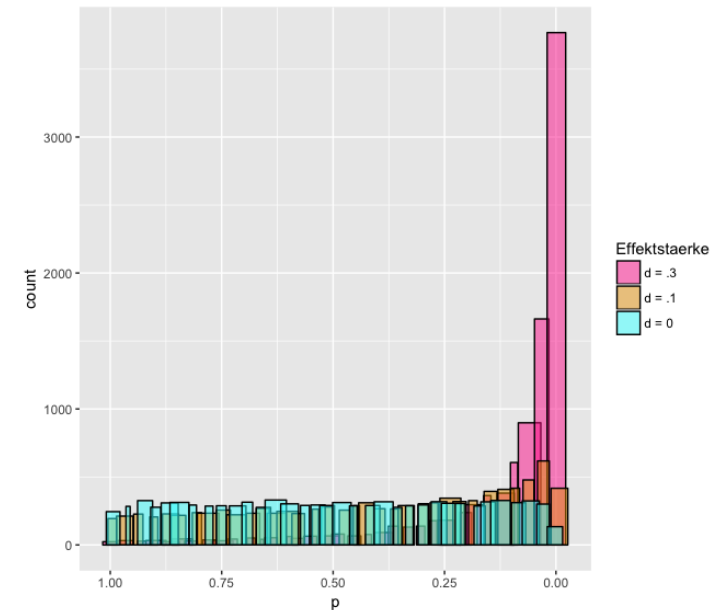
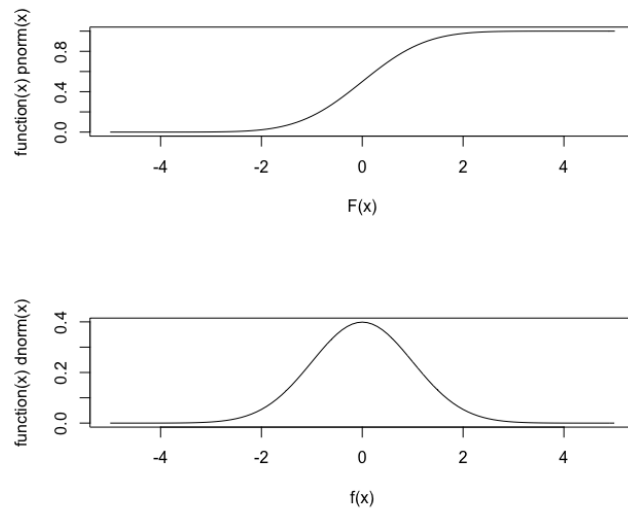
Test for Excess Significance (TES)

- Test zur Erkundung möglicher Bias' in der Gesamtmenge von Veröffentlichungen zu einem Thema
- Auf Basis der Effektstärke der beobachteten Befunde kann die Power der Studien geschätzt werden
- Die geschätzte Power führt dann zu der Anzahl zu erwartender signifikanter Ergebnisse
- Vergleich von erwarteter und beobachteter Anzahl signifikanter Ergebnisse auf Basis eines χ^2 -Tests:

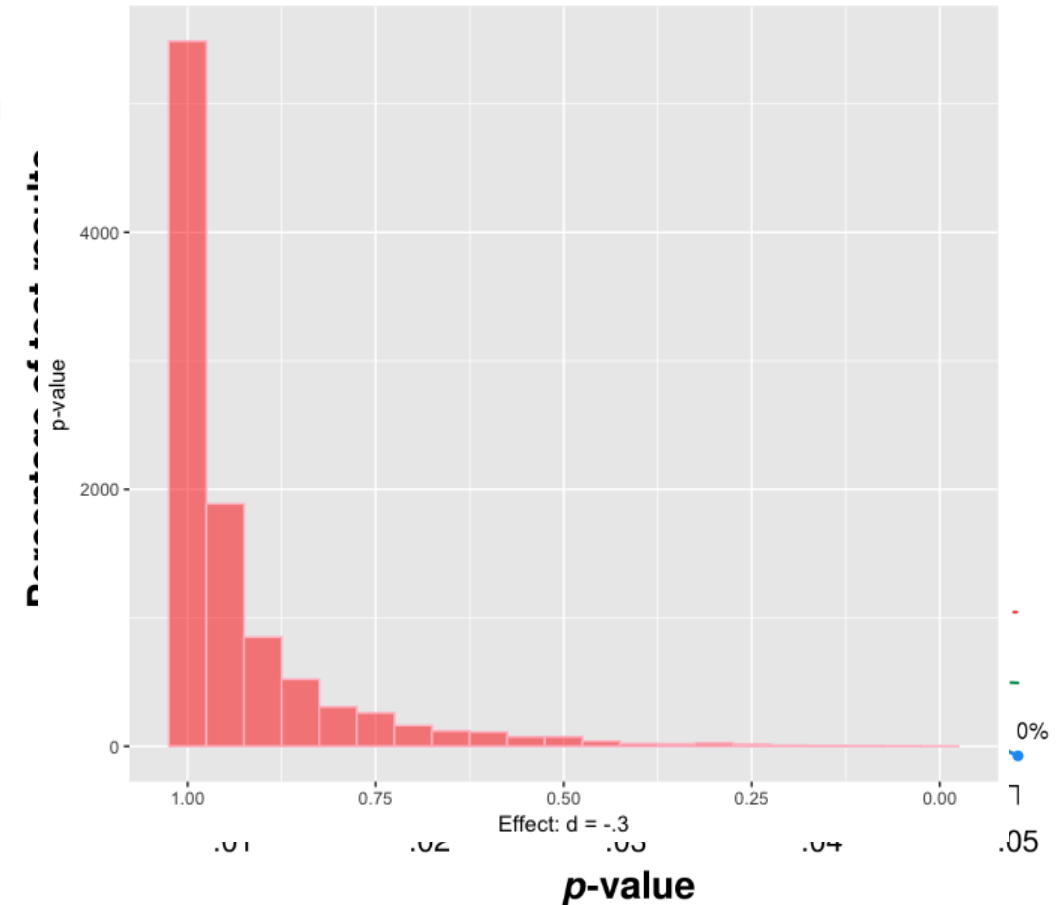
$$A = \left(\frac{(O - E)^2}{E} + \frac{(O + E)^2}{n - E} \right) \sim \chi_1^2$$

p -Curve

- Ein weiterer Zusammenhang ist der zwischen Verteilung der p -Werte und der Effektstärke:
 - Ist kein Effekt vorhanden, sind p -Werte unter der H_0 gleichverteilt
 - Ist ein Effekt vorhanden, sind p -Werte unter der H_0 rechtssteil verteilt



- Ein Vergleich mit der erwarteten Kurve des Effekts mit 33% Power simuliert den Verlauf von Studien mit überschätzten Effekten
- So wie ein Effekt existierender Effekt eine rechtssteile Kurve ergibt, ergibt ein negativer (überschätzter) Effekt eine linkssteile Kurve
- Ein Effekt mit Power < 50% ist immer überschätzt!



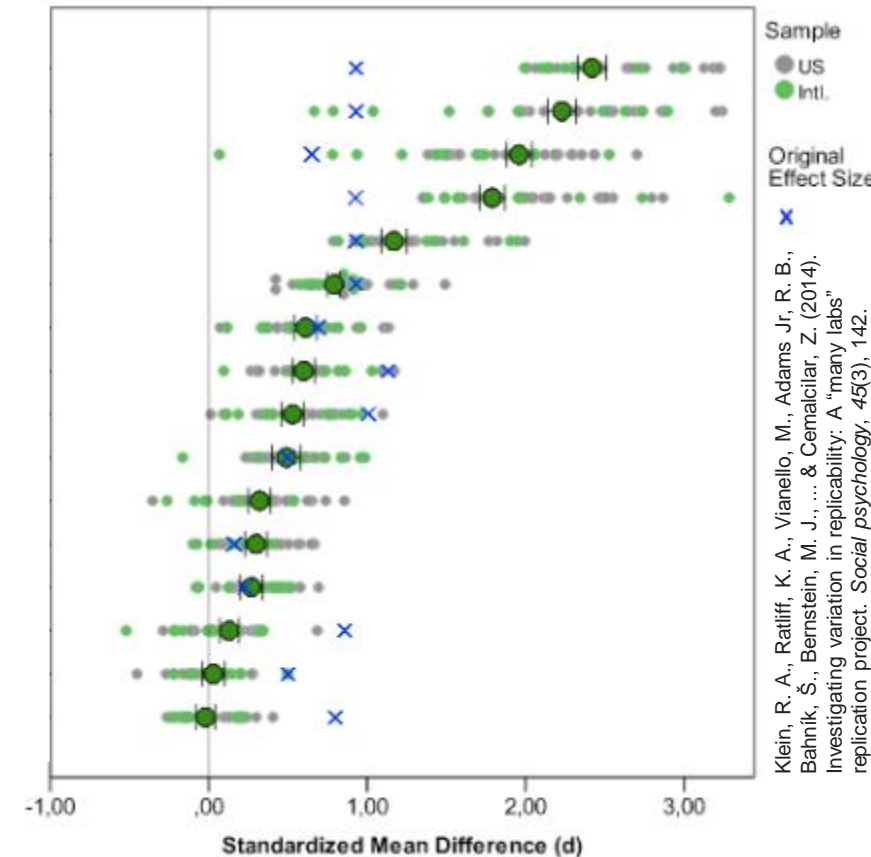
Das Alpha Level: Warum bedeutet $p \leq 0,05$ Signifikanz?

- Probleme des Cutoffs
 - P-Hacking
 - Unterschiedliche Interpretationen
- Mögliche Lösungen
 - Strengerer Cutoff (z. B. $p \leq 0,005$)
 - Beseitigt allerdings das p-Hacking Problem nicht
 - Bayesianische Ansätze (Bayes-Faktoren)
 - Kontinuierliche Maße statt strikter Cutoffs
 - Lakens et al. (2018) Vorschlag: „Bring your own Alpha“
 - **Alle Daten und Analysen mitveröffentlichen**
 - **Serien aus (Open Data) Studien analysieren**



Ansätze zu Replikationsserien

- Reproducibility Project (OSC, 2015)
 - Replikation 100 wichtiger Studien der Psychologie
 - Gegründet im Rahmen der Open Science Collaboration
 - 36% der Replikationen erfolgreich
- Many Labs Project (Klein et al., 2014)
 - Untersucht Variation bei multiple Replikationen
 - Multiple Replikation 13 psychologischer Studien
 - Replikation abhängiger von Effekt selbst als von Labor
- The Replication Network (replicationnetwork.com)
 - Listet, verbindet und kategorisiert Replikationen

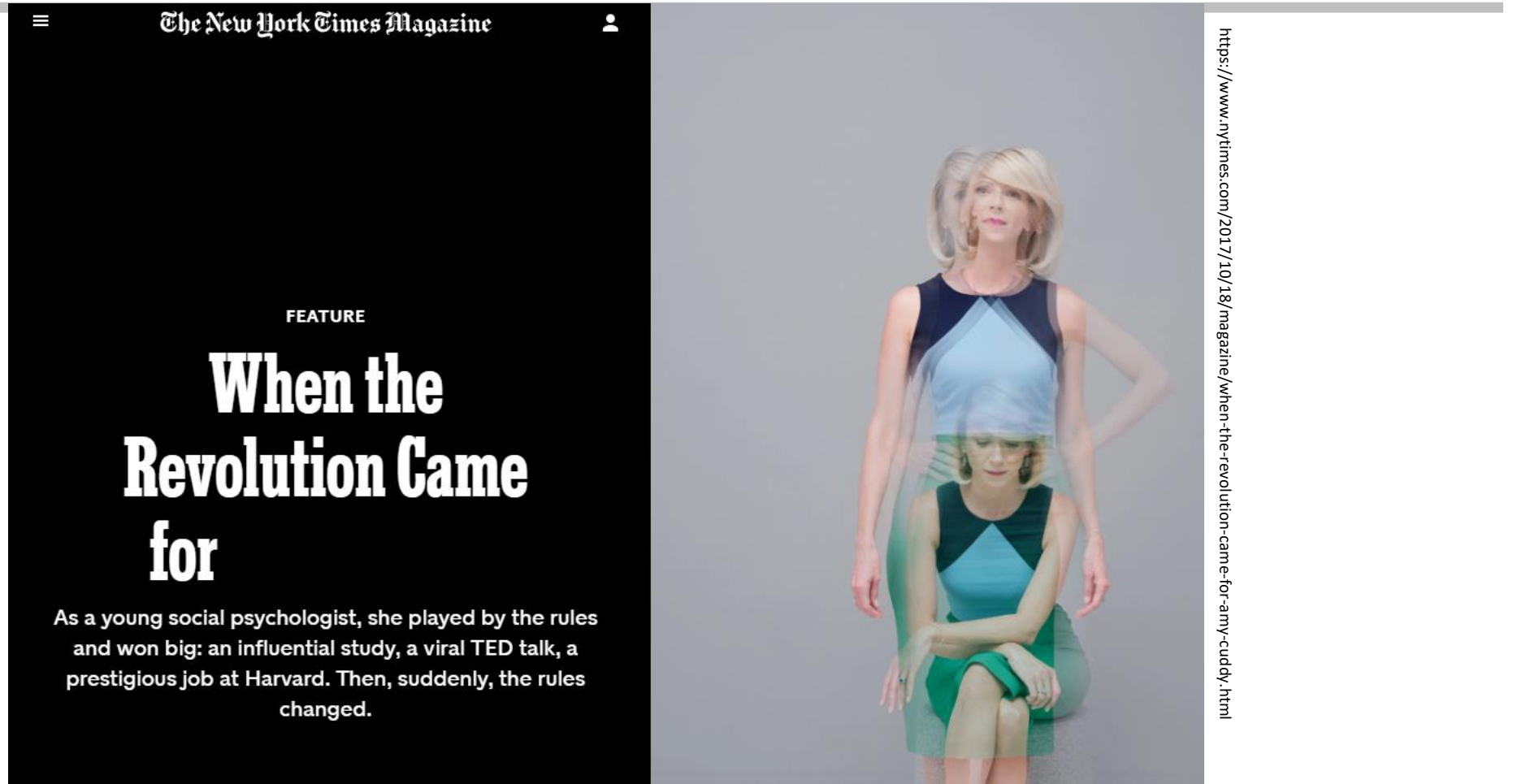


Die Folgen der Bewegung sind spürbar

Auswahl zurückgezogener Befunde:

- Diederik Stapel:
 - Menschen äußern sich rassistischer, wenn sie in vermüllten Umgebungen sind, und werden aggressiver, wenn sie Fleisch sehen
- Dirk Smeesters:
 - Medienberichte mit Inhalten über Tod bringen Menschen dazu, lokale Produkte zu bevorzugen
- Jens Förster:
 - Wenn Menschen als Ganzes schmecken, sind sie kreativer, wenn sie detailliert schmecken, analytischer

Auch die Powerpose hat nicht standgehalten



Was kann ich persönlich zu einer Verbesserung beitragen?

Quellen, Möglichkeiten und Anlaufstellen für Open Science

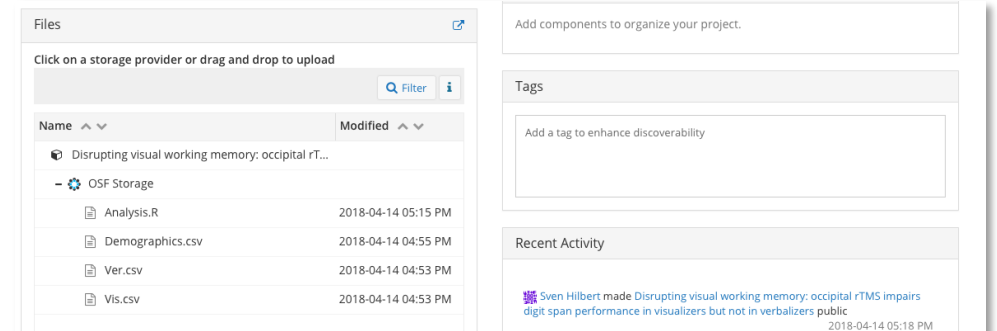
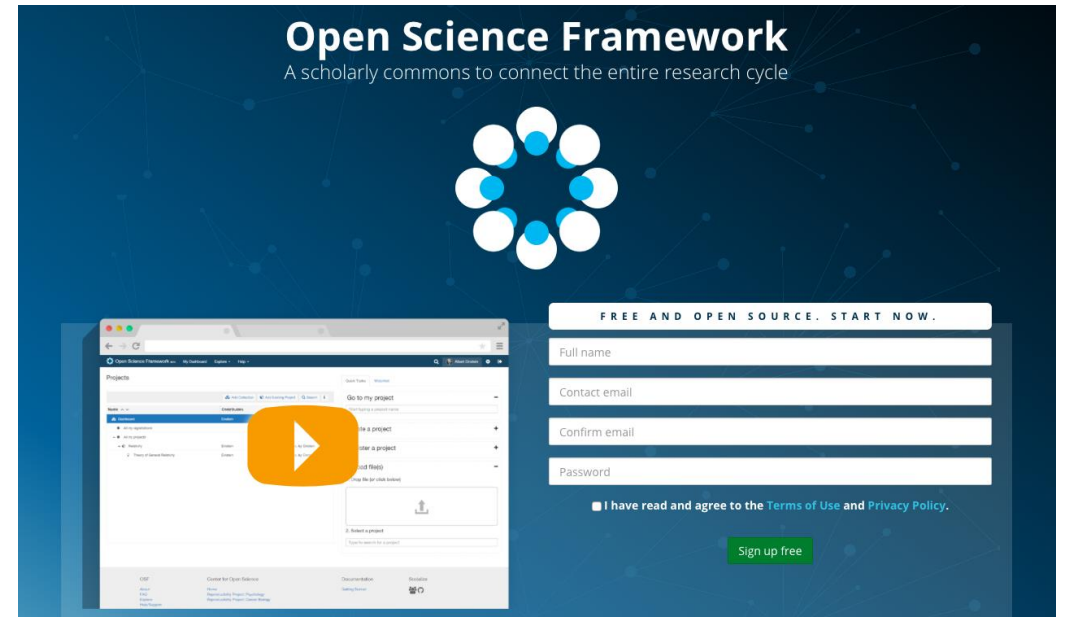
Was kann ich als persönlich tun?

- Präregistrierung
 - Studienplanung mit Fallzahl, Bedingungen und Analyse vor der Datenerhebung festlegen und öffentlich machen
- Open Data und Open Materials
 - Daten und Analyseskripte in Repository veröffentlichen
 - In Artikeln angeben, wo Daten und Analyse zu finden sind
- Gewissenhaft analysieren
 - Nach Wissenserwerb, nicht nach Signifikanz
- Alle Ergebnisse berichten
 - Alle Bedingungen und Analysen berichten, nicht nur die signifikanten Befunde



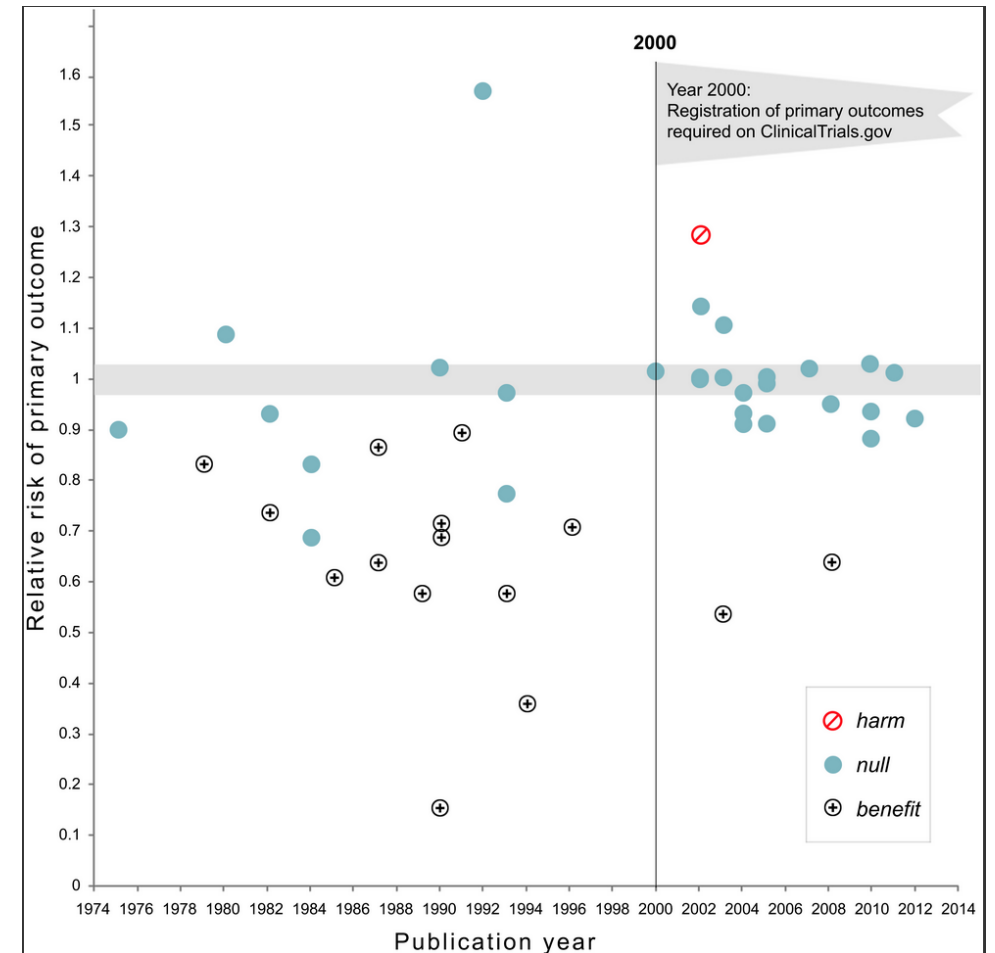
Präregistrierung

- Auf Homepage des Open Science Framework registrieren
 - <https://osf.io>
- Projekt eröffnen
- Präregistrierung der geplanten Studie durchführen
 - Vorlage für Präregistrierung ist verfügbar
- Veröffentlichung vor Erhebung
- Tutorials vorhanden
 - <http://www.osc.uni-muenchen.de/toolbox/index.html>



Präregistrierung

- Verpflichtende Präregistrierung kann sich deutlich auf die Studienlage auswirken
- Im Jahr 2000 fing das amerikanische „National Heart, Lung, and Blood Institute“ (NHLB) an, Analysepläne für Medikamente bei clinicaltrials.gov im Vorfeld zu veröffentlichen
- Die Erfolgsrate vergleichbarer Studien zu Medikamenten gegen kardiovaskuläre Erkrankungen fiel von 57% auf 8%



Registered Report

- Studie wird vorläufig akzeptiert, bevor sie durchgeführt wird
 - Studie wird final akzeptiert, wenn registrierte Methoden befolgt werden
 - Fragestellung und Analysequalität stehen im Mittelpunkt
 - Ergebnis steht nicht im Mittelpunkt



- Gegenwärtig 108 teilnehmende Journals (<https://cos.io/rr/>):
 - U.a. Cortex, European Journal of Personality, Journal of Cognition, Psychological Science, Canadian Journal of School Psychology


Open Access

Weitere Möglichkeiten, Artikel und Daten hochzuladen

- ArXiv.org oder PsyXarchiv.com
 - Preprints
- Bibliothek Uni Regensburg
 - Publikationsserver
- Research Gate
 - Je nach Open Access Politik der Journals
- PsychData.de
 - Scientific Use Files (nur Wissenschaftler haben Zugriff)

Was kann hochgeladen werden?

- Research Gate System
 - <https://www.researchgate.net>
- Sherpa/Romeo
 - <http://www.sherpa.ac.uk/romeo/>



A free preprint service for the psychological sciences
 Maintained by [The Society for the Improvement of Psychological Science](#)
 Powered by [OSF Preprints](#)

Search preprints...



Publisher copyright policies & self-archiving

Search

Journal titles or ISSNs Publisher names

Exact title starts with contains ISSN

[Advanced Search](#)

Zukunft der Open Science-Bewegung

- Challenges
 - Präregistrierung verbessern (Qualitätssicherung, kein Open Washing)
- Europa- oder besser weltweite Standards
- Empirische Überprüfung, ob Open Science die Qualität erhöht
- Anreizstruktur verbessern
 - Ausschreibungen von Stellen
 - Verlage einbinden
 - Mehr Badges!
- Spread the Word



Informationen über Open Science

- Open Science Framework Homepage
 - <https://cos.io>
- Open Science Center der LMU München
 - <http://www.osc.uni-muenchen.de/toolbox/index.html>
 - Toolbox
 - Präregistrierung
 - Repository Webseiten
- DataWiz Knowledge Base (Datendokumentation)
 - <https://datawizkb.leibniz-psychology.org/>
 - Hilfe bei Datenmanagement
 - Hilfe bei transparentem Forschungsprozess
- Informationsseite der Uni Münster
 - <https://www.uni-muenster.de/Forschungsdaten/publizieren/publikationswege/>
 - Information über Publikationswege



Journals, die Open Data verlangen

- **Advances in Methods and Practices in Psychological Science** (<http://www.psychologicalscience.org/publications/ampps/ampps-submission-guidelines#DISC>)
- **Collabra: Psychology** (<https://www.collabra.org/about/research-integrity/>)
- **Journal of Cognition** (<https://www.journalofcognition.org/about/editorialpolicies/>)
- **Journal of Research in Personality** (<http://www.sciencedirect.com/science/article/pii/S0092656617300211>)
- **Judgment and Decision Making** (<http://journal.sjdm.org/>)
- **Experimental Psychology** (<http://econtent.hogrefe.com/doi/10.1027/1618-3169/a000355>)
- **PLOS ONE** (<http://blogs.plos.org/everyone/2017/05/08/making-progress-toward-open-data/>)
- **Royal Society Open Science** (http://rsos.royalsocietypublishing.org/author-information#Open_data)

- Open Science: What, Why, and How?
 - <https://psyarxiv.com/ak6jr>
- Die statistische Krise in der Wissenschaft
 - <https://www.americanscientist.org/article/the-statistical-crisis-in-science>
- How it all began: False-Positive Psychology
 - <http://journals.sagepub.com.emedien.ub.uni-muenchen.de/doi/pdf/10.1177/0956797611417632>

Literatur

- Abele-Brehm, A. E., & Bühner, M. (2016). Wer soll die Professur bekommen?. *Psychologische Rundschau*.
- Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455-463.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, 315(7109), 629-634.
- Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891-904.
- Francis, G., Tanzman, J., & Matthews, W. J. (2014). Excess success for psychology articles in the journal Science. *PloS one*, 9(12), e114255.
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., ... & Cemalcilar, Z. (2014). Investigating variation in replicability: A “many labs” replication project. *Social psychology*, 45(3), 142.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science*, 23(5), 524-532.
- Lakens, D., Adolphi, F. G., Albers, C. J., Anvari, F., Apps, M. A., Argamon, S. E., ... & Buchanan, E. M. (2018). Justify your alpha. *Nature Human Behaviour*, 2(3), 168.
- Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical trials*, 4(3), 245-253.
- Rohrer, D., Pashler, H., & Harris, C. R. (2015). Do subtle reminders of money change people’s political views?. *Journal of Experimental Psychology: General*, 144(4), e73.
- Shanks, D. R., Vadillo, M. A., Riedel, B., Clymo, A., Govind, S., Hickin, N., ... & Puhmann, L. (2015). Romance, risk, and replication: Can consumer choices and risk-taking be primed by mating motives?. *Journal of Experimental Psychology: General*, 144(6), e142.
- Simmons, Joseph P. and Nelson, Leif D. and Simonsohn, Uri, False-Positive Citations (March 27, 2017). Perspectives on Psychological Science, Forthcoming.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359-1366.

Vielen Dank!

...und helfen Sie mit – es gibt noch viel zu tun!

