

Open Science: Wie sich die Wissenschaft öffnet

Lukas Röseler

Table of contents

1 Willkommen	6
Willkommen	7
1.1 Literatur	7
2 Über das Buch	8
I Einleitung	10
Einleitung	11
II Überschrift passend zum Beispiel	12
Literatur	13
3 Was ist Open Science?	14
3.1 Literatur	15
4 Wie ist mit Open Science umzugehen?	16
III Die Geschichte der Open Science Bewegung	18
Vertiefende Informationen	19
Literatur	19
5 Anfänge einer Revolution	20
5.0.1 Stapel-Affäre	20
5.0.2 Bem: Die Zukunft erfüllen	20
5.0.3 Bargh: Beeinflussen durch <i>Priming</i>	21
5.0.4 Literatur	21
6 Bestandsaufnahme	22
6.0.1 Definition von Replizierbarkeit	24
6.0.2 Weiterführende Literatur	24
6.0.3 „Eine Schwalbe macht noch keinen Sommer“	25
6.0.4 Phänomen-zentrierte Replikationsprojekte	27

6.0.5	Disziplin-zentrierte Replikationsprojekte	28
	Literatur	28
7	Wandel im System	29
7.0.1	Hat sich die Replizierbarkeitsrate erhöht?	31
7.0.2	Die Open Science Revolution als Paradigmenwechsel	31
7.0.3	Literatur	34
8	Struktur der Vertrauenskrise	35
8.0.1	Literatur	36
IV	Probleme	37
V	Probleme, die im Rahmen der Revolution identifiziert wurden	38
9	Probleme des Wissenschaftlichen Systems	40
9.1	Probleme des Wissenschaftlichen Systems	40
9.1.1	Wissenschaft versus Academia	40
9.1.2	Prekäre Arbeitsbedingungen	41
9.1.3	Zu viel Forschung	42
9.1.4	Publish or Perish – Veröffentlichen oder Verenden	43
9.1.5	Flaschenhals-Hypothese und Innovationsdrang	45
9.1.6	Schubladen-Problem	46
9.1.7	Zugängigkeit von Wissen	47
9.2	Vertiefende Informationen	48
9.3	Literatur	48
10	Karriere in der Wissenschaft	49
10.0.1	Doppelabhängigkeit	49
10.0.2	Depressionen, Burnout, und #IchbinHanna	50
10.0.3	Top-Down-Wandel	51
10.0.4	Literatur	51
11	Probleme der wissenschaftlichen Methoden	52
11.0.1	Exploratorische versus konfirmatorische Forschung	52
11.0.2	Freiheitsgrade von Forschenden (<i>Researchers' Degrees of Freedom</i>) . . .	57
11.0.3	Tippfehler	61
11.0.4	P-Hacking	61
11.0.5	Selektives Berichten (<i>Selective Reporting</i>)	64
11.0.6	Optionales Stoppen (Optional Stopping)	65
11.0.7	Darstellung kalibrierter Modelle als geplante Modelle (<i>Overfitting</i>) . . .	67
11.0.8	Tendenz von Menschen, sich selbst zu bestätigen (<i>Confirmation Bias</i>) .	68

11.0.9 Datenfälschung	68
11.1 Weiterführende Informationen	69
11.1.1 Literatur	69
12 Theorien	70
12.0.1 Deduktion und Induktion	72
12.0.2 Hilfshypothesen	73
12.1 Weiterführende Informationen	74
12.1.1 Literatur	74
13 Epistemische Probleme	75
13.0.1 Robustheit und Historizität von Phänomenen	75
13.1 Weiterführende Informationen	76
13.1.1 Literatur	76
VI Lösungen	77
VII Lösungen und Ansätze zur Verbesserung der Lage der Psychologie	78
.	79
14 Das System	80
14.0.1 Politik	81
14.0.2 Universitäten[LR9]	82
14.0.3 Institute und Vereinigungen	83
14.0.4 Zeitschriften	84
14.0.5 Forschende	89
14.0.6 Bewertungskriterien	89
14.0.7 Infrastruktur (Open Infrastructure)	91
14.0.8 Open Access Publikationen	91
14.0.9 Pre-Prints	96
14.0.10 Ansätze gegen die Selektion spannender Ergebnisse	97
14.0.11 Open Science Incentives FORRT NOSI Badges Open Science in Beru- fungskommission	101
14.0.12 Replikationsforschung[LR49] : Ansehen und Kompetenz (Normalität) im Umgang mit Replikationsstudien steigern	101
14.0.13 Modellierung von Replizierbarkeit	111
14.0.14 Veröffentlichung aller Ergebnisse	112
14.0.15 Umgang mit Fehlern	113
14.0.16 Offene Lehre / Open Teaching / Open Educational Resources	114
14.0.17 Literatur	117

15 Methoden	118
15.0.1 Reproducibility Checks / Reproduzierbarkeit	119
15.0.2 Robustness checks	120
15.0.3 Open Data	121
15.0.4 Replicability	124
15.0.5 Qualitative Forschung und Open Science	127
15.0.6 Open Source Software	127
15.0.7 Slow Science	128
15.0.8 Big Team Science	129
15.0.9 Literatur	129
16 Theorien	130
16.0.1 Spezifizierung und Formalisierung	130
16.0.2 Gegnerische Kollaborationen	131
16.0.3 Literatur	131
17 Die Welt	132
17.0.1 Literatur	132
VIII Fazit und Ausblick	133
Was hat sich verbessert?	134
Was ist noch zu tun?	135
Was geschieht jetzt gerade?	137
Literaturempfehlungen	137

1 Willkommen

Willkommen

Warnung

Dieses Projekt befindet sich in Arbeit und ist aktuell noch unfertig! Texte sind unvollständig oder fehlen, Quellenangaben und Abbildungen sind vorläufig, und es wimmelt von Rechtschreibfehlern...

Viele wissenschaftliche Disziplinen haben sich in den letzten Jahren stark verändert. Neue Zeitschriften, Methoden, Initiativen, Gesellschaften, Strukturen, und vieles mehr sind entstanden. Es ist die Rede von einer Krise (Mede et al. 2020), einer Revolution (Nosek et al. 2018) oder einer Renaissance (L. D. Nelson, Simmons, and Simonsohn 2018.). All diese Veränderungen haben gemeinsam zu einem stärkeren Grad an Offenheit und Transparenz geführt und das Wissenschaftssystem maßgeblich verändert.

In diesem Buch werden niedrigschwellig der Begriff der Offenheit, die Ursprünge des Wandels, und die Open Science Bewegung erklärt. Es werden Hintergründe des wissenschaftlichen Systems, neu entwickelte Methoden, und Probleme amtierender Methoden erläutert. Dabei wird vorwiegend auf Beispiele und Entwicklungen aus der Psychologie zurückgegriffen.

1.1 Literatur

2 Über das Buch

Wissenschaft wird zu einem Maßgeblichen Teil von den Mitgliedern einer Gesellschaft finanziert, beispielsweise über Steuergelder. Das bedeutet, eine Gemeinschaft fördert einzelne Personen, um ein tieferes Verständnis von sich und der Welt zu erlangen und im Gegenzug Unterstützung bei der Lösung verschiedener Probleme zu erhalten. Dadurch sind Wissenschaftler*innen der Gesellschaft in einem gewissen Maße Rechenschaft schuldig. Dabei geht es nicht darum, dass jede wissenschaftliche Erkenntnis notwendigerweise Früchte tragen muss, aber, dass sie wissenschaftlichen Qualitätskriterien entspricht und zumindest für andere Forschende nachvollziehbar und nachprüfbar ist. Inwiefern das in verschiedenen Disziplinen der Fall ist und wie Personen sich dafür einsetzen, diese Rechenschaft zu ermöglichen, möchte ich in diesem Buch darlegen.

Dieses Buch soll selbst ein Teil der offenen Wissenschaft sein. Im Kontext der Probleme und Lösungen erkläre ich, wie der Beruf *Wissenschaftler*in* aussieht, wie das Wissenschaftssystem funktioniert, und erörtere an Beispielen auszugsweise wissenschaftliche Debatten. Zur Offenheit gehört auch, dass das Buch frei zugänglich ist.

Über mich

Mein Name ist Lukas Röseler und ich bin Wissenschaftler. Ich habe in Wernigerode Psychologie und Wirtschaftswissenschaften studiert, mich darüber hinaus mit Philosophie und Wissenschaftstheorie beschäftigt, und habe 2021 als Doktorand der ersten Generation¹ in Bamberg meinen Dokortitel in Psychologie erhalten. Seit 2023 bin ich Geschäftsführer des Münster Center for Open Science und darf mich im Rahmen dessen mit transparenter und vertrauenswürdiger Wissenschaft in zahlreichen Disziplinen auseinandersetzen. Dabei betreibe ich Meta-Wissenschaft, also *Wissenschaft über Wissenschaft*.

Ich habe dieses Buch geschrieben, um das Thema *Open Science* anderen Wissenschaftler*innen, Studierenden, und allen anderen Personen auf einem einfachen aber ausführlichen Niveau zu erklären. Ein großer Aspekt von offener Wissenschaft ist nämlich, dass auch Personen, die primär nichts mit Wissenschaft zu tun haben, zumindest teilweise verstehen können, was dort vor sich geht. Das gleicht einem Schloss, das einst nur den reichen Herrschenden zugänglich war, im Zuge eines Gesellschaftlichen Wandels nun offene Tore und Türen hat und für alle frei zur Besichtigung steht.

Angesichts dieses Zieles veröffentliche ich das Buch kostenlos und stelle es online zur Verfügung. Sie können mir [hier](#) anonym Rückmeldung geben (Betreff: “Open Science Buch”).



Figure 2.1: Blick über Wernigerode

¹Das bedeutet, dass meine (Groß-)eltern keine Dokortitel haben.

Part I

Einleitung

Einleitung

Part II

Überschrift passend zum Beispiel

- Beispiel, bei dem Psychologische Wissenschaft wichtig ist
- Befunde aus der Wissenschaft zitieren
- Diese Befunde sind eventuell falsch
- Wie finden wir heraus, ob sie falsch sind?
- Wie stark können wir uns darauf verlassen?
- Wie stellen wir sicher, dass wir uns stärker auf Wissenschaft verlassen können?

[2] Durch die Ungewissheit über genaue Ursachen von geringen Replikationsraten wird teilweise auch eher zu dem Begriff „Vertrauenskrise“ geraten (z.B. Feest, 2023).

[LR1]hier noch Wissenschaftsskepsis erwähnen: <https://osf.io/preprints/psyarxiv/7u4fg>

[LR2]OS Taxonomie: <https://periodicos.ufsc.br/index.php/eb/article/view/91712>

Bild

<https://zenodo.org/records/7940641>

[LR3]https://www.researchgate.net/publication/374381552_What_is_the_Replication_Crisis_a_Crisis_of

[LR4]hier Bild vom Spektrum und der Verortung dieses Buches hintun; evtl. Umfrageergebnisse von Wissenschaftler*innen dazutun und zitieren, da gibt es was

[LR5]Als Infobox

Literatur

3 Was ist Open Science?

In der Wissenschaft dreht es sich oft um Details: Was genau passierte in der Studie? Welche Bilder sahen Versuchspersonen auf was für einem Bildschirm? Was war die Bildwiederholungsrate des Bildschirms, und waren die Farben dabei mittels Spektralphotometer kalibriert? Wissenschaftliche Untersuchungen und ihre Befunde werden heutzutage in Zeitschriftenartikeln beschrieben. Es kommt nicht selten vor, dass darin eines der oben aufgeführten Details fehlt, in den Zusatzmaterialien nicht auffindbar ist, oder die Autor*innen nicht gefragt werden können, weil ihre E-Mail-Adresse nicht mehr aktuell ist. Wenn nun aber der Befund von genauso einem Detail abhängt, werden zukünftige Forschende Probleme haben, ihn zum Vorschein zu bringen. Auf dieser konkreten Ebene bedeutet Open Science die Möglichkeit für jede Person, im Rahmen von ethischen und legalen Einschränkungen (z.B. Anonymität), detaillierte Einsicht in den gesamten wissenschaftlichen Prozess der Erkenntnisgewinnung zu erhalten. Es sollte also genau beschrieben werden, wie die Untersuchung ablief, welche Materialien dafür verwendet werden, welche Daten daraus resultierten, wie diese weiterverarbeitet und ausgewertet wurden, und schließlich, was daraus zu lernen ist.

Auf einer abstrakten Ebene meint Open Science einen höheren Grad an Offenheit und Transparenz in allen Facetten der Wissenschaft. Dazu gehören wie beschrieben Studienmaterialien (z.B. Fragebögen, Bilder, oder Videos), der wissenschaftliche Bericht, oder der Diskurs im Rahmen der Begutachtung wissenschaftlicher Berichte durch ihre Kolleg*innen (Peer-Review). Aber auch auf der systemischen Ebene kann der Grad an Transparenz steigen: Beispielsweise gibt es Listen mit Zeitschriften (<https://doaj.org>) oder Listen mit Professuren in einem bestimmten Wissenschaftsbereich (z.B. für die [Persönlichkeitspsychologie in Deutschland](#)). Schließlich kann mit Offenheit auch die Durchführung einer Konferenz im „hybriden Format“, also an einem bestimmten Ort aber mit der Möglichkeit zur Online-Teilnahme angeboten werden, um Personen, die nicht Anreisen (können) nicht auszuschließen, ihnen gegenüber also *offen* zu sein.

Insgesamt wird Open Science als Lösung für zahlreiche aktuelle, meist zusammenhängende Probleme verstanden, allen vorweg das Problem, dass sich ungefähr 50% aller wissenschaftlichen Studien in der Psychologie nicht replizieren lassen (Open Science Collaboration 2015). Fecher und Friesike ((Fecher and Friesike 2014), S. 19) reden von fünf „schools of thought“, also Denkkollektiven, die verschiedene Ziele verfolgen: Infrastruktur (z.B. Plattformen zum Speichern oder Veröffentlichen von Forschungsmaterialien), Öffentlichkeit und Involvierung der Gesellschaft in den wissenschaftlichen Prozess (z.B. Citizen Science), Messbarkeit wissenschaftlichen Erfolgs (z.B. Alternativen zum Impact Factor), Demokratie (z.B. Zugang zum Wissen), und Pragmatismus (z.B. höhere Effizienz von Wissenschaften

durch öffentliche Forschungsdaten). Eine Übersicht über unsortierte Facetten von Open Science ist unten abgebildet. Genaue Schätzungen sind schwierig, grob lässt sich dennoch sagen: Sucht man sich eine zufällige sozialwissenschaftliche Studie aus einer Fachzeitschrift aus, führt sie nach wissenschaftlichem Goldstandard erneut aus, und prüft die dort getestete Hypothese ein weiteres Mal, dann ist die Wahrscheinlichkeit, zum selben Ergebnis zu kommen, so hoch, wie bei einem Münzwurf “Zahl” (vs. Kopf) zu erhalten. Damit teilweise verbundene Probleme sind Betrug bei wissenschaftlichen Publikationen (Gopalakrishna, Wicherts, et al. 2021), oder psychische Probleme bei Jungwissenschaftler*innen (Satinsky et al. 2021).

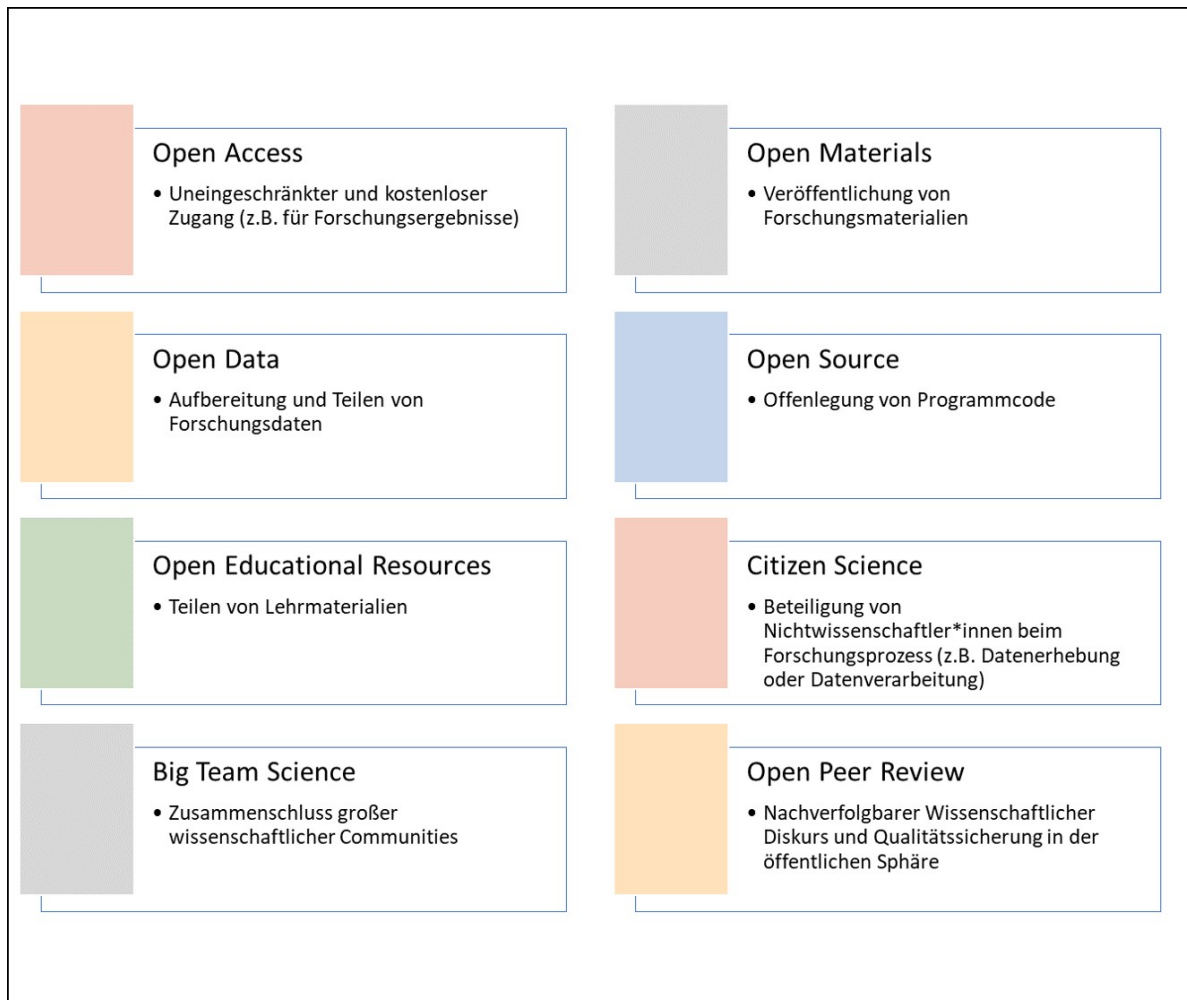


Figure 3.1: Facetten von Open Science

3.1 Literatur

4 Wie ist mit Open Science umzugehen?

Das Thema Open Science und Replikationskrise beziehungsweise Vertrauenskrise ist aus mehreren Gründen schwer kommunizierbar: Zuerst einmal ist es harte Kritik an der Wissenschaft und dem Wissenschaftssystem, die dazu missbraucht werden kann, wissenschaftliche Befunde kleinzureden und das Vertrauen in Wissenschaft insgesamt gefährdet. Mithilfe der hier präsentierten Argumente und Befunde lassen sich zum Beispiel auf Wissenschaft beruhende politische oder individuelle Entscheidungen kritisieren. Dagegen sei gesagt: Weitaus nicht alle wissenschaftlichen Befunde sind “falsch” oder nicht replizierbar. Bei vielen Punkten herrscht immer noch ein großer Konsens.

Auf der anderen Seite gehen die hier diskutierten Probleme über das hinaus, was ganz natürlich in fast jedem wissenschaftlichen Zweig wiederfindet. Denn je genauer man hinschaut, desto wahrscheinlicher ist es, dass ein wissenschaftlicher Befund relativierbar ist. Widersprüche oder Konflikte lassen sich überall finden, treten natürlicherweise auf, und werden von Wissenschaftler*innen ins Visier genommen. Jede*r Forschende kann berichten: Bei genauerem Hinschauen bleibt nichts schwarz oder weiß, sondern zerfließt in viele Grautöne. Bei den im Folgenden behandelten Replikationsfehlschlägen handelt es sich um mehr als die wissenschaftsinhärente Ungewissheit: Sie gefährden gesamte Wissenschaftsstränge.

Bevor wir in die Details fortschreiben, ist es außerdem wichtig festzuhalten, dass aktuell (Herbst, 2023) die meisten der in diesem Buch besprochenen Probleme noch nicht oder erst teilweise gelöst wurden und auf wöchentlicher Basis heiß diskutiert werden. Nach einem Jahrzehnt der Open Science Bewegungen ist ein Punkt erreicht, an dem die meisten Wissenschaftler*innen ein Bewusstsein über das Problem haben. Einige sind mit Lösungsansätzen vertraut, doch haben sich diese weder komplett durchgesetzt, noch ist klar, welche Probleme überhaupt bereits gelöst worden sind.

Ich verstehe Open Science als eine große Trigger-Warnung, die vor die Sozialwissenschaften vorgeschaltet sein sollte und lautet: Achtung, wir haben hier gerade sehr große Probleme. Es ist gut möglich, dass die Hälfte von allem, was wir zu wissen glauben, schlichtweg falsch ist. Meinungen über die Open Science Bewegung sind aber keineswegs homogen: Es lässt sich hier ein Spektrum spannen zwischen “Sollen wir wirklich die nächsten 20 Jahre damit verbringen, herauszufinden, welche Erkenntnisse der letzten 100 Jahre korrekt sind? Fangen wir doch einfach nochmal bei Null an.” und “Eigentlich ist das ganz normal, ich sehe keinen Grund, hier von einer ‘Krise’ zu sprechen”.



Figure 4.1: Spektrum der Reaktionen auf die Replikationskrise

Besonders groß ist das Problem bei Lehrbüchern: Stellen Sie sich vor, sie haben ein Buch über einen Teil der Psychologie (Aushängeschild ist hier oft die Sozialpsychologie) verfasst, das auf Jahrzehnten an Forschung besteht, hunderten Veröffentlichungen, und tausenden Studien. Hier hilft es kaum, das Buch komplett zu ignorieren. Praktikabel, alle Studien selbst zu replizieren, ist es allerdings auch nicht. Klar ist den meisten Wissenschaftler*innen hierbei: So wie es bisher gelaufen ist, kann es nicht weitergehen. Eine fünfzig prozentige Garantie für wissenschaftliche Erkenntnisse sollte nicht der Anspruch von etwas sein, das sich Wissenschaft nennt. Von vorne müssen wir aber auch nicht starten.

Part III

Die Geschichte der Open Science Bewegung

Im Rahmen dieses Buches wird die Open Science Bewegung also als eine Reaktion auf identifizierte Probleme und damit als Selbstkorrektur-Prozess der Wissenschaft verstanden. In im Zentrum steht ein mangelndes Vertrauen in Befunde, die in wissenschaftlichen Fachzeitschriften veröffentlicht wurden. Einige Probleme sind schon seit Jahrzehnten in ähnlicher Form bekannt. Bis sie allerdings öffentlich diskutiert wurden und Lösungsvorschläge erarbeiteten, benötigte es einschneidende Ereignisse. Aufgrund von Problemen, vergangene und für sicher geglaubte Ergebnisse nicht *replizieren* zu können (also bei wiederholten Untersuchungen zum selben Ergebnis zu kommen) ist häufig die Sprache von einer *Replikationskrise*. Durch die Ungewissheit über genaue Ursachen von geringen Replikationsraten wird teilweise auch eher zu dem Begriff „Vertrauenskrise“ geraten (Feest 2019).

Vertiefende Informationen

- Die Geschichte der Replikationskrise beschreiben L. D. Nelson, Simmons, and Simonsohn (2018) und Schimmack (2020). Eine positive Perspektive nehmen Korbmacher et al. (2023) ein.

Literatur

5 Anfänge einer Revolution

Um das Jahr 2010 herum häuften sich in der Psychologie Ereignisse, die für sich genommen als Einzelfälle abgetan werden konnten, gemeinsam aber ein negatives Bild der Wissenschaft zeichneten.

5.0.1 Stapel-Affäre

Durch einen Zufall entdeckten Nachwuchswissenschaftler im Jahr 2011, dass die Daten einer Studie ihres Kollegen, Diederik Stapel, von niemandem jemals erhoben wurden. Sie waren ausgedacht, bzw. fabriziert. Stand Juli 2024 wurden 58 von Stapels Fachartikel identifiziert und zurückgezogen, deren Daten fabriziert oder geschönt wurden.¹ Wissenschaftliche Institutionen wie der Begutachtungsprozess von Artikeln durch Fachkolleg*innen, deren Zweck die Qualitätssicherung war, hatten versagt. Seit dem Vorfall sind einige weitere Fälle bekannt geworden, teilweise durch erneute Analyse von Daten der jeweiligen Studien (O’Grady 2021) und oft durch Whistleblower, also durch Personen, die zu ihrem Schutz anonym bleiben wollen. Umfragen in den Niederlanden unter Forschenden haben ergeben, dass Fälschung oder Schöning von Daten von bis zu 10% aller Personen durchgeführt wird (Gopalakrishna, Riet, et al. 2021). Dabei ist zu beachten, dass Studien durch gefälschte Daten besonders innovativ, überraschend, oder klar werden - Eigenschaften, die die Veröffentlichung in einer Fachzeitschrift wahrscheinlicher machen.

5.0.2 Bem: Die Zukunft erfüllen

Kurze Zeit später veröffentlichte Daryl Bem, bekannt durch grundlegende psychologisch-philosophische Theorien wie der Self-Perception-Theory (Bem 1967), den Befund, dass Personen die Zukunft vorhersagen können (Bem 2011). Genauer gesagt, können manche Personen unter bestimmten Dingen, die Zukunft vorhersagen. In 8 Studien fand die Forschendengruppe, dass Personen Vorhersagen über erotische Bilder machen konnten. Die Ergebnisse wurden in der hoch angesehenen Fachzeitschrift *Journal of Personality and Social Psychology* veröffentlicht. Vielen Psycholog*innen war sofort klar: Entweder, Grundlegende Annahmen ihres Weltbildes waren falsch (“Personen können nicht die Zukunft vorhersagen”) oder es stimmte etwas mit den Ergebnissen nicht. Mehrere Forschende versuchten sich

¹Mittels der Retraction Database lassen sich nach Thema, Autor*in, Zeitschrift, usw. zurückgezogene Artikel durchsuchen: <http://retractiondatabase.org/>

darán, zu erklären, wie es zu den Ergebnissen kam. Analysen mit alternativen statistischen Methoden führten zur selben Schlussfolgerung (Wagenmakers et al. 2011), Replikationen durch unabhängige Forschende schlugen jedoch fehl Roe, Grierson, and Lomas (2012).

5.0.3 Bargh: Beeinflussen durch *Priming*

Seine Studien wurden im Marketing gefeiert und als Neurowissenschaftliche Erkenntnisse verkauft: Wer ein heißes Getränk (im Vergleich zu einem kalten) trinkt, schätzt andere Personen als “wärmer” (großzügig, rücksichtsvoll) ein (Williams and Bargh 2008). Wer Anagramme löst, die etwas mit hohem Alter zu tun haben (z.B. PFLEGEHEIM, GRAU, oder zum selbst probieren GHOSTECK), geht danach in langsamerem Tempo (Bargh, Chen, and Burrows 1996). Viele dieser Studien wurden repliziert: Forschenden fiel in der Anagramme-Studie auf, dass Bargh und Kolleg*innen die Zeit mit Stoppuhren gemessen hatten und dabei wussten, welche Person die “Alt”-Wörter und welche die neutralen Anagramme gelöst hatten - dabei lernt jede*r Psychologie-Studierende im ersten Jahr, dass das nicht der Fall sein sollten und Versuchsleiter*innen “blind” gegenüber dem Untersuchungszweck und der Zuordnung der Personen zu den Gruppen sein sollte. In ihrer Replikation (Doyen et al. 2012) ließen Doyen und Kolleg*innen die Zeit mit Lichtschranken erfassen und maßen selbst wie Bargh et al. in der Originalstudie. Bei der problematischen Messung kam dasselbe raus, die Lichtschranken, denen vorher nicht verraten wurde, welche Hypothese mit ihnen untersucht werden sollten und welche Personen welche Anagramme lösen mussten, konnten den Effekt jedoch nicht replizieren.

5.0.4 Literatur

6 Bestandsaufnahme

Eine ähnliche Vertrauenskrise gab es in der Sozialpsychologie in den 1960er Jahren (Lakens 2023). Ein entscheidender Unterschied war diesmal die Bestandsaufnahme: Parallel zu diesen eigenartigen Befunden oder *Anomalien* vernetzten sich Psycholog*innen um Brian Nosek international und untersuchten die Replizierbarkeit von 100 Studien aus namhaften psychologischen Fachzeitschriften (Open Science Collaboration 2015). Sie fanden heraus, dass sich nur 39 der 100 Originalbefunde replizieren ließen. Bei allen anderen Studien, waren die Replikationsergebnisse anders als die ursprünglichen Ergebnisse. Viele weitere Großprojekte folgten, alle mit ähnlichen Ergebnissen: Die Replikationsraten lagen weit unter den Gewünschten.

Kritische Betrachtung der Open Science Collaboration, 2015

Obgleich dieses „Reproducibility Project Psychology“ die gesamte Fachgemeinschaft zutiefst erschütterte und den Weg für einen Paradigmenwechsel ebnete, bemängeln manche Forschende auch negative Auswirkungen auf nachfolgende Replikationsforschung. Indem 100 Studien gleichzeitig von einer Gruppe aus über 100 Forschenden veröffentlicht wurden, setzte das Projekt unrealistische Maßstäbe für Replikationsforschung. Gleichzeitig war die Qualitätskontrolle dabei weniger streng, da die einzelnen Studien nicht alle in dem Maße begutachtet werden konnten, wie es bei einer traditionellen Veröffentlichung der Fall gewesen wäre (z.B. Röseler et al. (2022)). Einige gleichermaßen ambitionierte Vorhaben wurden veröffentlicht, wie zum Beispiel die ManyLabs Studien (z.B. Klein et al. (2014); Klein et al. (2018)) oder Versuche, bei denen unabhängige Gruppen dieselben Hypothesen testeten und replizierten (Landy et al. 2020). Oft beschränken sich diese Vorhaben auf Studien, die sich im Rahmen einer Online-Befragung replizieren lassen. Formate wie Längsschnittstudien oder Verhaltensbeobachtungen sind dabei unterrepräsentiert.

Zahlreiche Verbünde folgten. Einige Projekte konzentrierten sich auf einzelne Phänomene. Beispielsweise haben sich 17 Forschungsgruppen zusammengetan, um den Befund des *Facial Feedback* (Strack, Martin, and Stepper 1988) zu replizieren (Wagenmakers et al. 2016). Dabei geht darum, dass Personen einen Stift mit den Zähnen festhalten und dabei je nach Ausrichtung des Stiftes entweder diejenigen Muskeln anspannen, die sie auch zum Lachen benötigen oder eben nicht. In der „Lachen“-Bedingung fanden die Versuchspersonen im Anschluss Comics witziger. Die Replikation schlug fehl. 2022 wurde eine weitere Studie mit über 3000 Versuchspersonen aus 19 Ländern veröffentlicht - diesmal auch mit direkter Beteiligung von Fritz Strack, der die Originalstudie durchgeführt hatte (Coles et al. 2022). Wieder zeigte sich, dass

die Position eines Stiftes im Mund sich nicht auf die Bewertung von Stimuli auswirkt. Jenseits von sozialpsychologischen Befunden konzentrierten sich Forschende auch auf Bereiche wie Forschung mit Babys (Byers-Heinlein et al. 2020) oder auf ganze Zeitschriften (Camerer et al. 2018).

Liste großer Replikationsprojekte (siehe auch [FORRT Replication Hub](#))

Projekt	Thema	Link
Reproducibility Project: Psychology	Psychologie	https://osf.io/ezcu/
CORE Data Replicada	Entscheidungsforschung	https://osf.io/5z4a8/
	Konsumentenverhalten und Entscheidungs-forschung	https://datacolada.org/archives/category/replication
Many Labs 1	Psychologie	https://osf.io/wx7ck/
Many Labs 2	Psychologie	https://osf.io/8cd4r/
Many Labs 3	Psychologie	https://osf.io/ct89g/
Many Labs 4	Psychologie	https://osf.io/8ccnw/
Many Labs 5	Psychologie	https://osf.io/7a6rd/
Soto Social Sciences Replication Project	Persönlichkeitspsychologie	https://doi.org/10.1177/0956797619831612
Registered Replication Reports	Verhaltensforschung	http://www.socialsciencesreplicationproject.com
	Verschiedene	—
Many Babies 1	Entwicklungspsychologie	https://manybabies.org
Sports Sciences Replications	Sportwissenschaften	https://ssreplicationcentre.com
Hagen Cumulative Science Project	Psychologie	https://osf.io/d7za8/
I4R Replications	Politikwissenschaften	https://i4replication.org/reports.html
Experimental Philosophy	Experimentelle Philosophie	https://doi.org/10.1007/s13164-018-0400-9
Reproducibility Project: Cancer	Krebsforschung (Medizin)	https://www.cos.io/rpcb
SCORE	Sozialwissenschaften	https://www.cos.io/score
REPEAT	Gesundheitssystem	https://www.repeatinitiative.org
CREP	Psychologie	https://www.crep-psych.org
Boyce et al., 2023	Psychologie	https://doi.org/10.1098/rsos.231240

ReproSci	Biologie	https://reprosci.epfl.ch
Boyce et al. 2024	Psychologie	https://doi.org/10.31234/osf.io/an3yb

6.0.1 Definition von Replizierbarkeit

Zu sagen, was repliziert werden konnte und was nicht, ist erst nach einer Definition möglich. Im Sprachgebrauch von Forschenden wird mit „wurde repliziert“ gemeint, dass ein Replikationsversuch zu gleichen Ergebnissen wie eine Originalstudie gekommen ist. Zu Replikationsfehlschläge wird „konnte nicht repliziert werden“ gesagt, subtil davon abweichend kann „wurde nicht repliziert“ meinen, dass keine Replikationsversuche existieren oder sie fehlschlugen. Für eine Wissenschaft, die über 100 Jahre alt ist, scheint es überraschend, dass noch immer keine klare Definition wichtiger Konzepte rund um das Thema Replikation vorliegt, geschweige denn es zur Routine gehört, Studien zu replizieren. Während sich in verschiedenen Feldern abweichende Taxonomien durchgesetzt haben, sieht die Verwendung in diesem Buch wie in der Tabelle beschrieben aus.

Table 6.2: Replikations-Taxonomie nach Turing Way (The Turing Way Community and Scriberia 2024)

		Daten	
Analyse	gleich	gleich	unterschiedlich
	unterschiedlich	reproduzierbar	replizierbar
		robust	verallgemeinerbar

6.0.2 Weiterführende Literatur

Für eine systematischere, in den Informationswissenschaften verankerte Taxonomie zur Art der Replikation siehe (Plesser 2018). Eine an den statistischen Methoden angelehnte Taxonomie für die Ergebnisse von Replikationsstudien haben LeBel et al. (LeBel et al. 2019) vorgeschlagen. Philosophisch diskutiert wird Replikationsnähe zum Beispiel von (Choi 2023) und (Leonelli 2023).

Table 6.3: Replikationstaxonomie

Unterscheidungskriterium	Ausprägungen
Ergebnisse einer Replikationsstudie	Erfolgreich Fehlgeschlagen Unklar oder gemischt

Unterscheidungskriterium	Ausprägungen
Nähe einer Replikationsstudie zur Originalstudie (in Anlehnung an Lebel REF und Hüffmeier et al REF)	Direkte Replikation (selbe Versuchsleiter*innen, selbe Versuchsmaterialien, neue Versuchspersonen) Nahe Replikation (andere Versuchsleiter*innen, möglichst ähnliche Versuchsmaterialien, neue Versuchspersonen) Konzeptuelle oder konstruktive Replikation (andere Versuchsleiter*innen, andere Versuchsmaterialien, neue Versuchspersonen)
Ziel der Replikation	Reproduktion Mit selben Daten und selbem Programmiercode zu denselben Ergebnissen gelangen Replikation Mit anderen Daten zu denselben Ergebnissen gelangen

6.0.3 „Eine Schwalbe macht noch keinen Sommer“

Ob an einem wissenschaftlichen Befund „etwas dran ist“, er also einen Wahrheitsanspruch hat, hängt – neben seiner eigentlichen Art der Etablierung – bei der Replikationsforschung von vielen Faktoren ab. Was waren die Ergebnisse der Replikationsstudie? Wie viele und wie unterschiedliche Studien wurden durchgeführt? Wie sahen die genauen Methoden aus? Was waren die Unterschiede zwischen Replikationen und Originalstudie? Während Einzelstudien immer einen Erkenntnisgewinn liefern (mindestens, ob eine bestimmte Methode praktikabel ist, (Sikorski and Andreoletti 2023), können sie je nach Forschungsgebiet stark variieren Landy et al. (2020). Für das Gesamtbild braucht es mehr, wie zum Beispiel eine statistische Aggregation aller Einzelbefunde im Rahmen einer Meta-Analyse. Ein Beispiel mit Fantasiedaten befindet sich dazu in der folgenden Abbildung.

Betrachtet man viele Studien, die den Zusammenhang zweier Dinge, wie zum Beispiel Einkommen und Bildungsabschluss, untersucht haben, so werden sich die Studien hinsichtlich der Details unterscheiden: Was wurde alles zum Einkommen gezählt (Netto, Brutto, Sozialleistungen, Einkommen von Familienangehörigen, Werte über einen Zeitraum oder von einem bestimmten Zeitpunkt aus der Vergangenheit, usw.) oder auch welche Personen befragt wurden (Studierende, Berufstätige, wurden Befragte bezahlt, usw.). Alle diese Unterschiede wirken

sich möglicherweise auf den Zusammenhang aus, und selbst wenn sie es nicht tun, unterliegen Zusammenhänge oft Schwankungen, die sich durch die Messmethoden ergeben. In dem Beispiel lässt sich die Stärke des Zusammenhangs auf eine Zahl und eine dazugehörige Präzision herunterbrechen. Die Zahl heißt hier “Korrelation” und die Schärfe “Fehlerbalken”. In dem Wald-Diagramm (*Forest Plot / Blobbogram*) sind mögliche Korrelationen aus verschiedenen Studien abgebildet. Im Rahmen einer Meta-Analyse können Studienergebnisse anschließend kombiniert und Unterschiede untersucht werden.

```
library(ggplot2)
library(metafor)
```

Lade nötiges Paket: Matrix

Lade nötiges Paket: metadat

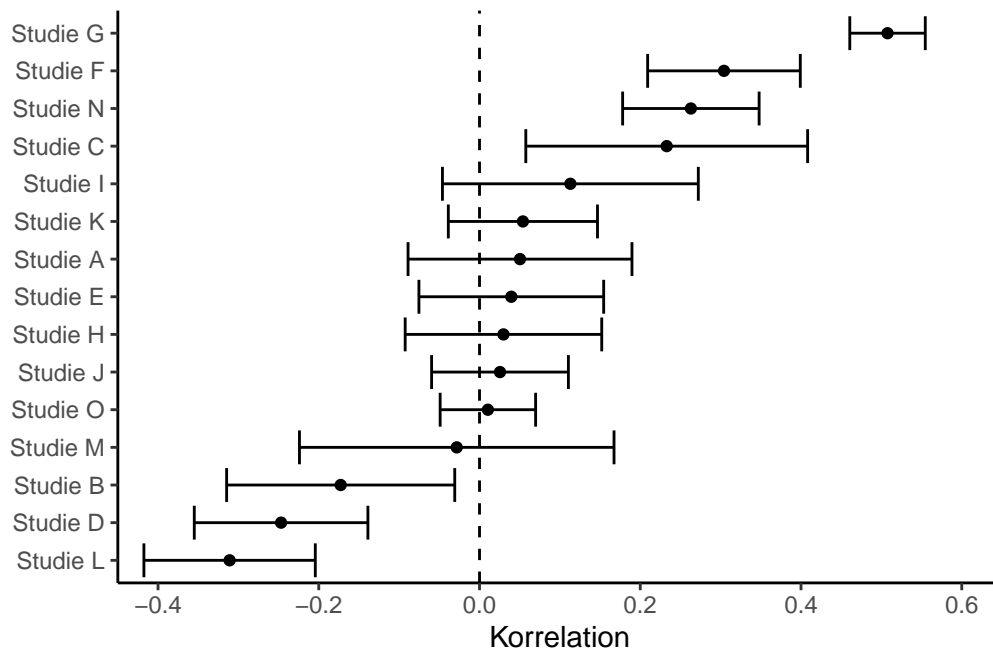
Lade nötiges Paket: numDeriv

Loading the 'metafor' package (version 4.6-0). For an introduction to the package please type: help(metafor)

```
set.seed(10)
k <- 15
cors <- data.frame("Stichprobenumfang" = round(rchisq(n = k, df = 2, ncp = 0)*5+10, digits = 0),
  "Korrelation" = rnorm(n = k, mean = .05, sd = .3),
  "Studie" = paste("Studie", toupper(letters[1:k]), sep = " "),
  "yi" = NA,
  "vi" = NA
)

cors[, 4:5] <- metafor::escalc(ni = cors$Stichprobenumfang, ri = cors$Korrelation, measure = "I",
  cors$ucb <- cors$yi + qnorm(.975)*cors$vi
  cors$lcb <- cors$yi - qnorm(.975)*cors$vi

ggplot(cors, aes(x = Korrelation, y = reorder(Studie, Korrelation))) +
  geom_point() + geom_errorbar(xmin = cors$lcb, xmax = cors$ucb) +
  geom_vline(xintercept = 0, lty = 2) + theme_classic() + ylab("") +
  xlim(c(-.4, .6))
```



6.0.4 Phänomen-zentrierte Replikationsprojekte

Im Gegensatz zu dem breit gefächerten RPP und anderen Versuchen, die Replikationsrate zu schätzen, haben sich andere Versuche auf grundlegende Phänomene fokussiert. Dutzende Gruppen auf der ganzen Welt haben sich in solchen Fällen zusammengeschlossen, auf einen Versuchsaufbau geeinigt, und führen die Studien mit einer enormen Anzahl an Versuchspersonen durch. Die meisten dieser Vorhaben stammen aus der Psychologie. Während die dabei gefundenen Effektstärken, also sozusagen die Deutlichkeit eines Zusammenhanges oder Befundes, in fast allen Fällen weit unter denen bisheriger Studien lagen (Kvarven, Strømmland, and Johannesson 2020), waren sie zudem beim Großteil der Studien null, die Phänomene waren also „nicht sichtbar“ (Alogna et al. 2014; Eerland et al. 2016; Bouwmeester et al. 2017; O'Donnell et al. 2018; Wagenmakers et al. 2016.; Cheung et al. 2016; Vaidis et al. 2024; Rife et al. 2024). So konnte beispielsweise mit einer enormen Präzision gezeigt werden, dass eine Geschichte über einen Professor Versuchspersonen in einem anschließenden Leistungstest *nicht* schlauer macht (O'Donnell et al. 2018).

i Effiziente Nutzung von Ressourcen?

Wie geht man mit Ressourcen bei Replikationen um? Bei Zusammenschlüssen vieler Forschender stellt sich diese Frage unweigerlich. Erstellen alle Gruppen unabhängig voneinander die Studie? Halten sich alle an ein zuvor abgestimmtes Protokoll? Führen sie die Studie nacheinander durch um voneinander zu lernen? Bei *Registered Replica-*

tion Reports wird für gewöhnlich von einem zuvor mit anderen Forschenden (z.B. den Autor*innen der Originalstudie) ein Versuchsaufbau abgestimmt. In anderen Fällen wird gemeinsam ein Versuchsaufbau erarbeitet, der zum Testen der Theorie ideal sein sollte (*Creative Destruction Approach*, Tierney et al. (2020)). Teams in verschiedenen Ländern übersetzen das Protokoll dann und halten sich bei der Durchführung eng daran. Diese Protokolle sind manchmal nicht im Vorhinein getestet (Buttliere 2024), basieren oft aber auf erfolgreichen, namhaften Studien. Das hat den Vorteil, dass Unterschiede zwischen den Gruppen nicht auf Unterschiede in der Durchführung zurückzuführen sind und sich Kulturen vergleichen lassen (Kakinohana, Pilati, and Klein 2022). Ein Nachteil dabei ist jedoch, dass, wenn an einem, zwei, oder fünf Standorten das Experiment schon nicht funktioniert, es fraglich ist, ob die übrigen 30 Gruppen es auch probieren sollten. In den Worten von Buttliere (2024): “Wer bekommt bessere Ergebnisse? 39 Personen, die etwas zum ersten Mal tun, oder eine Person, die etwas 39 Mal tut?”

6.0.5 Disziplin-zentrierte Replikationsprojekte

Ungefähr die Hälfte aller psychologischen Befunde ist also nicht replizierbar. Heißt das, alle Sozialwissenschaftlichen Lehrbücher aus allen Disziplinen sind zur Hälfte falsch? Die klare Antwort heißt *nein*. Die akkurate Antwort lautet *kommt darauf an*.

6.0.5.1 Jenseits der Psychologie

Inwiefern es auf die Disziplin innerhalb der Sozialwissenschaften ankommt wurde bisher vor allem in der Psychologie untersucht. Aktuelle Tendenzen weisen darauf hin, dass Replikationsraten in der Persönlichkeitspsychologie und kognitiven Psychologie (Soto 2019) höher liegen als die in der Sozialpsychologie (Open Science Collaboration 2015) oder im Marketing (Charlton 2022). Während schon hunderte Replikationsversuche für sozialpsychologische Studien veröffentlicht sind, sind es in anderen Bereichen wie dem Marketing aktuell weniger - Stand Oktober 2022 sogar nur 9. Bereiche außerhalb der Psychologie sind von Replikationsproblemen ebenfalls betroffen. Von Problemen der Replizierbarkeit, Reproduzierbarkeit, und Nachvollziehbarkeit sind fast alle Disziplinen betroffen. Neue Lösungsansätze werden in Medizin, Biologie, Chemie, Physik, Geschichtswissenschaften, Politikwissenschaften, Erziehungswissenschaften, Informatik, und vielen weiteren Bereichen diskutiert.

Literatur

7 Wandel im System

Seit dem Bekanntwerden der geringen Replizierbarkeit psychologischer Studien wurde das wissenschaftliche System in vielen Aspekten hinsichtlich seiner Offenheit und Transparenz verändert. Einige Zeitschriften setzen für die Veröffentlichung wissenschaftlicher Artikel voraus, dass die Daten öffentlich zugänglich sind beziehungsweise erklärt ist, weshalb das nicht der Fall ist (z.B. bei Daten, die sich schwierig anonymisieren lassen). In seltenen Fällen wie bei der Zeitschrift *Meta-Psychology* rechnen Wissenschaftler*innen alle Ergebnisse nach. Als Antithese zum “Impact Factor”, einer Kennzahl, die angibt wie oft eine Zeitschrift zitiert wird und nachweislich nichts mit der Qualität der darin enthaltenen Forschung zu tun hat (Brembs 2018), wurde der TOP-Factor eingeführt (TOP: Transparency and Openness Promotion). Dieser gibt für eine Liste von Kriterien an, in welchem Maße sie von verschiedenen Zeitschriften erfüllt werden. In anderen Feldern wie Betriebswirtschaftslehre oder Marketing werden Zeitschriften sogar von wenigen ausgewählten Forschenden über Ratingskalen bewertet, die jährlich als Rankings veröffentlicht werden. TOP-Factors hingegen sind objektiv, werden stetig aktualisiert, und sind unter topfactor.org öffentlich einsehbar und nachvollziehbar. Für jeden Faktor gibt es vier Stufen: Keine erwähnung, Level 1, Level 2, und Level 3. Level 3 ist dabei das Ideal, zum Beispiel hieße das im Hinblick auf Transparenz von Daten, dass ein Artikel erst dann veröffentlicht wird, wenn die Daten öffentlich verfügbar sind und die Analysen von einer unabhängigen Person erfolgreich nachgerechnet (*reproduziert*) wurden. Zum Scoring einer Zeitschrift werden für die Levels 1-3 entsprechende Punkte vergeben und aufsummiert.¹

Table 7.1: Übersicht über TOP Richtlinien

Faktor	Erklärung
Zitation von Daten	Den meisten Forschungsartikeln liegen Daten zugrunde. Spezifisch geht es hier um die Möglichkeit, Daten unabhängig von den Artikeln zitieren zu können (z.B. über eine eigene Kennung, wie ein <i>Digital Object Identifier</i> [DOI]).
Transparenz von Daten	Wenn möglich sollten Daten veröffentlicht werden. Das ist direkt über Fachzeitschriften oder über sogenannte <i>Forschungsdaten-Repositorien</i> möglich.

¹ Sozialwissenschaftler*innen wissen, dass das Aufsummieren der Werte unsinnig ist, da Level 2 nicht “doppelt so gut” wie Level 1 ist, und die verschiedenen Faktoren nicht immer gleich gewichtet werden sollten - rechnerisch wird aber genau das angenommen. Als heuristische Schätzung der Offenheit von Zeitschriften ist dieses Vorgehen jedenfalls sinnvoller als bibliometrische Maße.

Faktor	Erklärung
Transparenz vom Analyse-Code	Mit dem Analyse-Code werden die Forschungsdaten ausgewertet. Andere Forschende sollten die Möglichkeit haben, den Code auszuführen und die Ergebnisse zu <i>reproduzieren</i> bzw. zu prüfen.
Transparenz der Forschungsmaterialien	Bei Forschungsmaterialien kann es sich um Fragebögen, gezeigte Bilder oder Filme, aber auch präsentierte Gerüche, verkostete Früchte, oder Programme handeln. Wenn möglich, sollten auch sie (in digitaler Form) in einem Repositorium abgelegt werden. Das ist bei Gegenständen selbstverständlich nicht möglich. Bei Genen gibt es beispielsweise <i>alphanumerische Codes</i> , mittels welchen sich Forschende verständigen.
Richtlinien zum Beschreiben des Versuchsaufbaus und der Analysen	Zum Verständnis aber auch zur Nachbildung einer Studie ist es wichtig, den Versuchsaufbau genau zu beschreiben. Einige Zeitschriften haben in den letzten Jahren beispielsweise die Wortbegrenzung für die entsprechende Abschnitte in Forschungsartikeln aufgehoben.
Präregistrierung von Studien	Siehe auch Abschnitt zu Präregistrierungen in diesem Buch: Sofern eine Studie Hypothesen testet (also z.B. Erwartungen oder Vorhersagen für das Ergebnis), sollten diese im Vorhinein festgelegt sein und nach dem Sehen der Ergebnisse nicht an diese angepasst werden. Idealerweise fordern Fachzeitschriften für solche Studien Präregistrierungen und verpflichten Forschende den Link zu ihnen anzugeben.
Präregistrierung des Analyseplans	Der Weg von den Daten zu den Ergebnissen ist ein langer. Die Ergebnisse hängen von vielen Entscheidungen ab. Um sich dieser Flexibilität zu berauben, sollten Forschende den Analyseplan in der Präregistrierung beschreiben.
Replikation	Trotz ihres Wertes gibt es noch immer viele Zeitschriften, die keine Replikationsstudien veröffentlichen. Im Mindestfall ermutigen Zeitschriften Forschende dazu, Replikationsstudien bei ihnen einzureichen.

Am Wandel beteiligt sind vor allem Jungwissenschaftler*innen oder Forschende im frühen Karrierestadium (*Early Career Researchers*, ECR). An vielen Universitäten haben sich in den letzten Jahren Open Science Initiativen oder regelmäßige Treffen („Reproducibili-Tea“) herausgebildet, die fast ausschließlich aus “Post Docs” (Personen nach der Promotion ohne Professur), Promovierenden (Doktoranden), und Studierenden bestehen. In Deutschland haben sie sich zum NOSI (Netzwerk der Open Science Initiativen Deutschland) vernetzt (Schönbrodt, Baumert, et al. 2022).

7.0.1 Hat sich die Replizierbarkeitsrate erhöht?

Die Replikationskrise dauert nun schon über ein Jahrzehnt an und einiges hat sich geändert. Ist dadurch auch das Problem der Replizierbarkeit gelöst? Zum einen ist für viele Bereiche noch gar nicht klar, was sich replizieren lässt. Im Marketing führte die Zeitschrift *Journal of Business Research* kurzzeitig eine Replikationsecke ein (Easley and Madden 2013), welche dann jedoch in eine andere Zeitschrift verlagert wurde. Aktuell sind Projekte in Arbeit, die die Replizierbarkeit für verschiedene Disziplinen schätzen. Deren Ergebnisse sind größtenteils noch vorläufig und unklar. In einem eigenen Projekt, sammeln wir Replikationsergebnisse, um langfristig je Disziplin und über die Zeit zu schauen, wie sich Replikationsraten verändert haben. Tagesaktuelle Werte sind online verfügbar (https://forrr-replications.shinyapps.io/fred_explorer/). Eine Evaluation über mehrere Disziplinen und Jahre erfordert noch hunderte weitere Replikationsstudien.

7.0.2 Die Open Science Revolution als Paradigmenwechsel

Wissenschaftshistoriker oder -theoretiker beschreiben die Entwicklung der Wissenschaft als nicht-stetig. Lehrbücher werden nicht immer dicker, stattdessen werden manche Kapitel kürzer, weil das darin beschriebene Wissen verworfen wird, andere werden dicker, weil neue Erkenntnisse hinzukommen. Zeitweise verschwinden Kapitel sogar vollständig. Eines der bekanntesten Wissenschaftsmodelle stammt von Thomas Kuhn (1970/1996), der selbst Psychologe war und große Teile vom Mediziner und Soziologen Ludwik Fleck (1935/2015) übernommen hat. Darin wird angenommen, dass zeitweise das Wissen wächst, sich jedoch dabei Befunde anhäufen, die mit allem anderen Wissen nicht vereinbar sind. Diese sogenannten Anomalien lassen sich ab einem bestimmten Punkt nicht mehr ignorieren. Ab dort kippt das wissenschaftliche Weltbild: Neue Theorien werden entworfen, die die Anomalien erklären können und altes Wissen wird verworfen oder in die neuen Theorien integriert. Dieses Kippen wird als Paradigmenwechsel oder wissenschaftliche Revolution bezeichnet. Paradebeispiel für so einen Paradigmenwechsel ist der Übergang vom geozentrischen zum heliozentrischen Weltbild in der Astronomie, die Prospect Theory (Kahneman and Tversky 1979) in den Wirtschaftswissenschaften, oder, so die Behauptung hier im Buch und auch anderorts (Sönning and Werner 2021): Die Replikationskrise in der Psychologie. Anomalien sind in diesem Fall die Befunde von Bem oder Bargh oder vereinzelte Replikationsfehlschläge. Sie waren mit bisherigem Wissen nicht vereinbar und nachdem sich viele solcher Befunde häuften, ließen sie sich nicht mehr ignorieren oder abtun. Dass die Replikationsforscher*innen etwas falsch gemacht hatten, nicht qualifiziert waren, oder Pech hatten, war keine gute Erklärung mehr. Im Gegensatz zu klassischen Kuhnschen Revolutionen steht bei der Open Science Revolution keine bestimmte Theorie oder Forschungsdisziplin die verworfen wird im Fokus, sondern die wissenschaftliche Methode und das Wissenschaftssystem der Sozialwissenschaften. Sozialwissenschaften haben darüber hinaus nicht jeweils nur ein Paradigma sondern mehrere unabhängige (Hoyningen-Huene and Kincaid 2023). In Anlehnung an die wissenschaftstheoretische Terminologie von Kuhn wird neben Replikationskrise auch der Begriff *Glaubwürdigkeits-Revolution* [Credibility revolution;

Korbmacher et al. (2023)] verwendet. Für philosophische Betrachtungen siehe auch Rubin (2023).

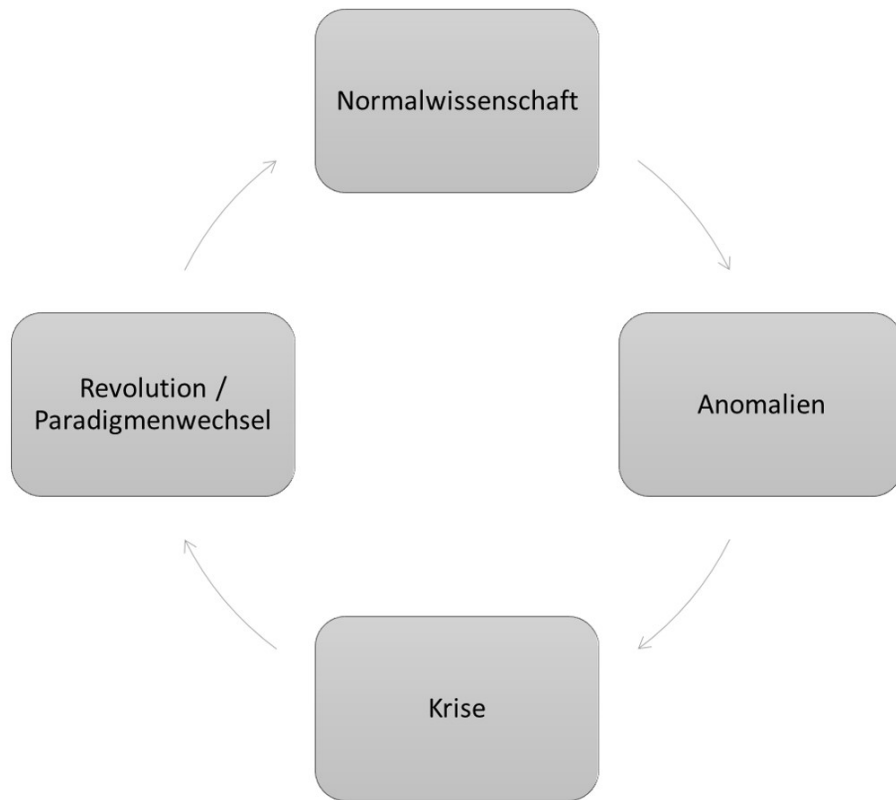


Figure 7.1: Wissenschaftliche Revolution nach (Kuhn 1970/1996), Darstellung in Anlehnung an (Fiorentino and Montana Hoyos 2014)

Ein Paradigmenwechsel ist vergleichbar mit einem Kippbild die dem Hase-Ente-Bild (**Abbildung 1**) wie es zum Beispiel Wittgenstein (1968) gezeichnet hat. Bis zu einem gewissen Punkt sind sich alle einig, dass es ein Hase ist. Doch nach und nach kommen Erkenntnisse und Perspektiven hinzu. Der Punkt wird überschritten, die Ente ist anerkannt, und niemand würde es mehr für einen Hasen halten. Beim Hase-Ente-Bild lässt sich natürlich beliebig hin und her springen. Beim wissenschaftlichen *Fortschritt* kommt neues Wissen hinzu und eine Rückkehr ist nur noch schwer möglich.

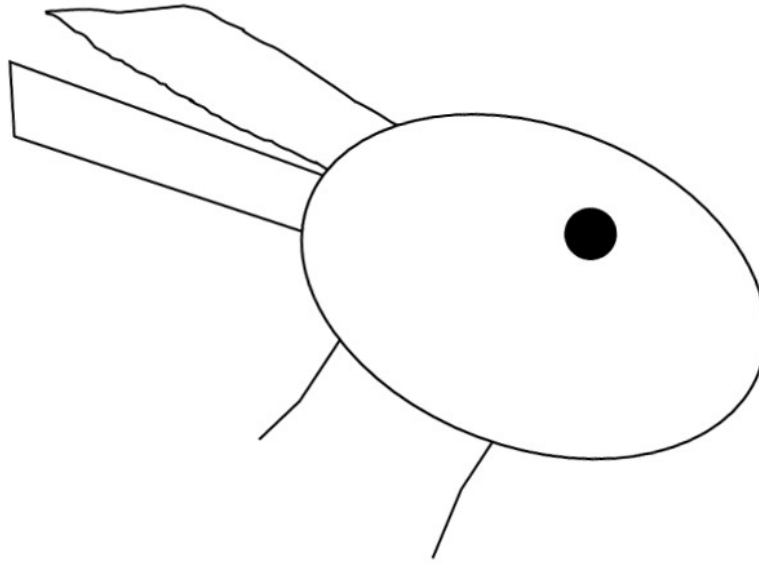


Figure 7.2: Hase-Ente-Kippbild frei nach Wittgenstein: Die Ausstülpungen auf der linken Seite können entweder als Schnabel einer nach links blickenden Ente oder als Ohren eines nach rechts blickenden Hasen interpretiert werden.

Forschende werden im Rahmen ihres Studiums oder einer Promotion von Beginn an darin geschult, entsprechend des amtierenden Paradigmas mit Befunden umzugehen.

💡 Beim ersten Versuch klappt es nie: Paradigmenwechsel in der Psychologie

Schon im Rahmen meines Studiums wurde ich geschult, mit dem amtierenden Paradigma konform mit fehlgeschlagenen Replikationen umzugehen. Das fand noch statt bevor sich die Replikationskrise herauskristallisierte. Bei der ersten Studie, an der ich beteiligt war, replizierten wir beispielsweise den Befund, dass bunte Mengen in ihrer Anzahl weniger aussahen als einfarbige Mengen. Im Rahmen der Konsumentenpsychologie ist das hinsichtlich Slogans wie “viele viele bunte Smarties” etwas kontraintuitiv, es lässt sich aber gestaltpsychologisch plausibel darlegen (Redden and Hoch 2009). Nachdem eine Gruppe von Kommilitoninnen das Gegenteil herausfand, nämlich dass *bunte Smarties tatsächlich nach mehr aussahen als einfarbige Smarties*, konnten wir in zwei eigenen Folgestudien nichts von beiden nachweisen. Egal ob sich die Smarties in Teller, Tassen, oder Schalen, befanden, egal ob es blaue, rote, gelbe, oder bunte Schokolinsen waren, egal ob Mengen geschätzt oder Smarties aus großen Flaschen in Gefäße geschüttet wurden: Unsere

Versuchspersonen ließen sich von der “Buntheit” nicht beeinflussen. In den folgenden Jahren durfte ich als Tutor dann weitere Replikationsversuche durchführen. Der betreuende Professor und mein Mentor erklärte: Ich habe es eigentlich noch nie erlebt, dass die Hypothese gleich beim ersten Versuch bestätigt wird. Nach jedem Experiment ist man schlauer und weiß, was man nächstes Mal besser machen muss. Es ist ganz natürlich, dass es ein paar Versuche dauert, bis man weiß, wie sich die Hypothese bestätigen lässt. Nachdem wir sechs Studien mit insgesamt 1383 Versuchspersonen durchgeführt hatten, die Autoren der Originalstudie um Rat gebeten hatten, haufenweise Schokolinsen verzehrt hatten, und die Hypothese über alle Studien hinweg nicht bestätigt wurde, hatte ich das Vertrauen in den Befund verloren.

Ungefähr sechs Jahre später dachte ich während meiner Promotionszeit an die Studien zurück und diskutierte mit dem Professor. Im Rahmen der Replikationskrise war klar: Wenn man mehrere Experimente durchführt und die Hypothese eigentlich falsch ist, wird alleine durch den Zufall ab und zu trotzdem die Hypothese bestätigt. Das ist vergleichbar damit, dass selbst eine faire Münze sechs Mal hintereinander auf derselben Seite landen kann. Wenn man jedoch 10 Münzen jeweils sechs Mal wirft, ist es nicht selten, dass eine der 10 Münzen sechs Mal auf derselben Seite landet. Aus dieser Perspektive hat das Zitat, dass die Ergebnisse beim ersten Versuch eigentlich nie so rauskommen, wie man es sich wünscht, einen bitteren Beigeschmack: Wenn die Hypothese falsch ist, ist es tatsächlich unwahrscheinlich, dass sie dennoch bestätigt wird. Nicht aber, wenn viele Studien durchführt. Dann ist es sogar zu erwarten, dass irgendwann eine Studie die Hypothese - auch wenn sie eigentlich falsch ist (!) - bestätigt. Die Studien haben wir gemeinsam mit einigen Beteiligten schließlich veröffentlicht (Röseler, Felser, et al. 2020).

7.0.3 Literatur

8 Struktur der Vertrauenskrise

Im Gespräch über Präregistrierungen – eine Methode zur Erhöhung der Replizierbarkeit eines Befundes – entgegnete ein Kollege: “Schön und gut, aber solange sich das System nicht ändert, wird sich an Replikationsraten nichts ändern.” Wie lässt sich dieser Einwand verstehen? Zäumen (bzw. trensen) wir dazu das Pferd von hinten auf: Spätestens seit dem Reproducibility Project Psychology (Open Science Collaboration, 2015) ist den meisten Psycholog*innen klar, dass es große Probleme bei der Replizierbarkeit von Befunden gibt. Woher kommen diese genau? Ein wichtiges Problem ist hierbei der *Publikationsbias*, womit gemeint ist, dass bei der Veröffentlichung von wissenschaftlichen Berichten eine Auswahl getroffen werden muss und dadurch nur bestimmte Ergebnisse publiziert werden (z.B. Ergebnisse, die eine bestimmte Theorie stützen). Dadurch steht ein verzerrtes Bild der Realität. Oft schreiben Wissenschaftler*innen sogar nur bestimmte Befunde zu Berichten zusammen. “Fehlgeschlagene Experimente”, also solche, bei denen eine Hypothese nicht bestätigt oder eine Theorie nicht gestützt werden konnte, landen in der Schublade (engl. *File-Drawer-Problem*; Rosenthal (1979); Theodore D. Sterling (1959)). In extremeren Fällen bedienen sich Wissenschaftler*innen verschiedener größtenteils anerkannter Methoden, die Daten so darzustellen, dass sie die Hypothese stützen oder tun so, als ob das, was sich aus den Daten lesen lässt, von Beginn an die Hypothese gewesen wäre (HARKing, Parsons et al. (2022)). Aber was treibt Personen, die sich vor allem für den Beruf wegen ihres Interesse an der Funktionsweise der Welt (oder im Falle der Psychologie der Funktionsweise des Menschen) und wegen Ihrer Suche nach Wahrheit entschieden haben dazu, die Wahrheit zu unbewusst zu schönen oder sogar bewusst zu manipulieren? Am Anfang des komplexen Prozesses, welcher zur Replikationskrise geführt hat, steht das aktuelle wissenschaftliche System: Ein Großteil der in der Wissenschaft beschäftigten arbeitet unter extremem Druck und prekären Arbeitsbedingungen. Um nach ein bis zwei Jahren eine Vertragsverlängerung zu erhalten, müssen Publikationen von Artikeln in möglichst angesehenen Fachzeitschriften nachgewiesen werden. Diese kriegt man durch besonders aussagekräftige und spannende Ergebnisse. Das Anreizsystem der Wissenschaft belohnt also nicht Wahrheit, Genauigkeit, Bescheidenheit, oder Transparenz, sondern vor allem diejenigen Dinge, die nicht in der Hand eine*r Wissenschaftler*in liegen: Spannende und eindeutige Ergebnisse (Bakker, van Dijk, and Wicherts 2012).

Der Prozess von prekären Arbeitsbedingungen bis zur niedrigen Replikationsrate ist hier veranschaulicht. In den folgenden Kapiteln werden die Probleme und Lösungsansätze im Detail diskutiert.

Abbildung 2

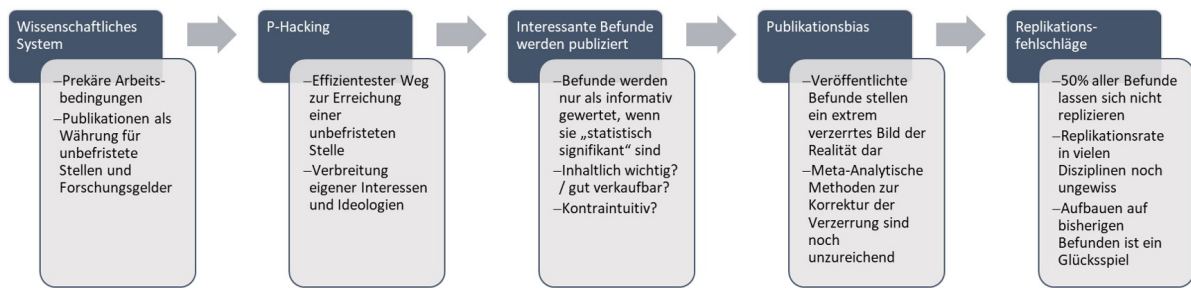


Figure 8.1: Das Anreizsystem der Wissenschaft als Ursache für Replikationsfehlschläge

8.0.1 Literatur

Part IV

Probleme

Part V

Probleme, die im Rahmen der Revolution identifiziert wurden

Im folgenden werden zahlreiche Probleme des Wissenschaftssystems erklärt und aufgelistet. Manche der Probleme sind seit vielen Jahrzehnten bekannt und andere erst seit wenigen Jahren. Es sei bei der Lektüre zu beachten, dass es zu *allen* diesen Problemen bereits ausgearbeitete und teilweise auch ausgeführte Lösungsansätze gibt. Wie die Lösungsansätze aussehen und welche Lösungen auf welche Probleme zugeschnitten sind, wird ausführlich im Kapitel zu Lösungen erläutert.

9 Probleme des Wissenschaftlichen Systems

9.1 Probleme des Wissenschaftlichen Systems

Neben dem Idealbild davon, was Wissenschaft sein sollte oder wie sie funktionieren sollte, existiert die Wissenschaft, wie sie in unser Gesellschaftssystem integriert ist. Besonderheiten sind dabei, dass Wissenschaftler*innen ihre Tätigkeit als Beruf ausüben, also Geld dabei verdienen. Wie das Geld, das größtenteils aus Steuergeldern stammt, verteilt werden soll entscheiden Gremien, die wiederum selbst aus Wissenschaftler*innen bestehen (*akademische Selbstverwaltung*). Durch die hohe Arbeitsbelastung, gleichzeitig Wissenschaft zu betreiben und zu verwalten, vereinfachen sich die Entscheider*innen die Arbeit und verwenden zur Auswahl hochqualifizierter Personen Abkürzungen. So konnte es passieren, dass die Währung in der Wissenschaft zum Großteil die Anzahl der in Fachzeitschriften veröffentlichten Forschungsartikel (*Paper*) ist. Ein weiterer verwendeter Indikator ist die Anzahl, in wie vielen weiteren wissenschaftlichen Artikeln auf die Forschung einer Person verwiesen wird, also die Zitationszahlen. Vorbilder wie Charles Darwin oder William James und selbst aktuelle Nobelpreisträger hätten nach dem heutigen Maßstab keine Chance auf eine unbefristete Stelle in der Wissenschaft - sie haben einfach nicht genug Paper geschrieben. Viele der hier diskutierten Problemen, sind beispielsweise in der Psychologie seit mehreren Jahrzehnten bekannt, sodass eine Replikationskrise unabwendbar erschienen haben muss (Cronbach, Snow, and Wiley 1991; Greenwald 1976).

9.1.1 Wissenschaft versus Academia

Während einige Natur- und Sozialwissenschaften in ihrer Anfangszeit oft von Buchveröffentlichungen lebten und darin eine umfangreiche Basis von meist einzelnen Personen erarbeitet wurde (z.B. Galileo Galilei für die Physik oder William James für die Psychologie) stehen heutzutage wissenschaftliche Fachzeitschriften im Fokus. Diese Zeitschriften sind vergleichbar mit solchen, die es im Kiosk und im Supermarkt gibt, nur bestehen sie halt aus (meist englischsprachigen) Artikeln, die Wissenschaftler*innen verfasst haben und sind in vielen Fällen nur noch online oder in Hochschulbibliotheken erhältlich. Forschende laden sich dann einzelne Artikel aus den Zeitschriften aus dem Internet über das Hochschulnetzwerk herunter und Bibliotheken haben Verträge mit Verlagen und zahlen Geld, damit Universitätsangehörige Zugriff zu den Katalogen haben. Jeder eingereichte Artikel befasst sich mit einer Fragestellung, die von den Forschenden selbst festgelegt wurde (z.B. wie viele psychologische Studien lassen sich im Mittel replizieren?). Darin werden meistens Studien

mit deren Ergebnissen berichtet. Vor der Veröffentlichung werden die Artikel begutachtet (*Review*) – nicht von Mitarbeitenden der Zeitschrift sondern von Kolleg*innen (*Peers*). Mit diesem *Peer-Review* soll die Qualität von Forschung sichergestellt werden. Üblicherweise wird dabei darauf geachtet, dass die Schlussfolgerungen auf Basis der erhobenen Daten gerechtfertigt sind, die Fragestellung klar beantwortet wird, der Artikel verständlich ist, und die Befunde spannend oder überraschend sind. Zeitschriften unterscheiden sich darin, welche Themen sie abdecken (z.B. Sozialpsychologie, Konsumentenverhalten, Angewandte Sportwissenschaft, usw.), wie streng das Peer-Review ist, und von wie vielen Forschenden sie gelesen und zitiert werden. Wissenschaften sind also stark integriert in ein System, das Forschenden und Verlagen erlaubt, ein festes Gehalt zu verdienen.

9.1.2 Prekäre Arbeitsbedingungen

Soweit der Rahmen: Wissenschaftler*innen forschen und teilen ihre Ergebnisse meist in Form von *Publikationen* wie Zeitschriftenartikel- oder Buch-Veröffentlichungen. Was die Jobs in der Wissenschaft (also vorwiegend an Hochschulen) angeht, so sind sie hierarchisch strukturiert. Es gibt wenige unbefristete Stellen, meist Professuren, und darunter befristete Stellen für Promovierende und bereits promovierte *Post Docs*. Wer in der Wissenschaft arbeitet, befindet sich meistens in einem harten Konkurrenzkampf um eine der wenigen unbefristeten Stellen (Rahal, Fiedler, Adetula, Berntsson, Dirnagl, Feld, Fiebach, Himi, Horner, Lonsdorf, Schönbrodt, et al. 2023). Der Hintergedanke ist dabei, dass Konkurrenz zwischen Forschenden die Produktivität steigern möge. Vom Start der Promotion¹ bis zur Berufung auf eine Professur, also eine der raren unbefristeten Stellen, dauern Verträge meistens nur ein bis drei Jahre und haben oft einen Umfang von weniger als 100%. Gleichzeitig ist es unüblich, mit weniger als 40 Stunden pro Woche innerhalb von den typischen drei Jahren die Promotion erfolgreich abzuschließen. Während ein Großteil aller wissenschaftlichen Veröffentlichungen auf Studien beruht, die im Rahmen von Doktorarbeiten durchgeführt wurden, sind Doktorand*innen gleichzeitig diejenigen Personen im System, die den geringsten Wert haben bzw. deren Arbeitskraft am günstigsten ist. Forschende auf befristeten Stellen sind also einem enormen Leistungsdruck ausgesetzt. Psychische Probleme wie Burnout oder Depressionen sind unter Promovierenden weit verbreitet (Jaremka et al. 2020; Liu et al. 2019). Den Weg zur Professur schaffen vor allem die Personen erfolgreich, die viele Artikel in prestigeträchtigen Zeitschriften veröffentlichen. Durch die immense Arbeitsbelastung und große Zahl an Artikeln, die bei Fachzeitschriften eingereicht werden, ist keine Zeit mehr, Ergebnisse genau zu prüfen und nachzurechnen (Michele B. Nuijten et al. 2017), sondern es wird vor allem darauf geachtet, wie eindeutig die Ergebnisse die Fragestellung beantworten - oder genauer gesagt: bestätigen (Giner-Sorolla 2012; Mynatt, Doherty, and Tweney 1977). Mit anderen Worten: Es wird ausgerechnet der Teil der wissenschaftlichen Arbeit belohnt, der nicht in der Hand der Forschenden liegt, nämlich die Ergebnisse von Untersuchungen. Veröffentlichte Artikel und Prestige statt Qualität (Brembs 2018) sind ab dort die Währung der Wissenschaft: Auf ihrer Basis wird

¹Prozess der Erlangung eines Doktorgrades/Dokortitels, je nach Disziplin und Hochschule durch das Verfassen eines Buches (Monografie) oder mehrerer Zeitschriftenartikel (kumulative Promotion)

entschieden, wer Forschungsgelder erhält und auf Basis von Forschungsgeldern und Publikationen werden Professuren vergeben. In den darüber entscheidenden Berufungskommissionen lesen die Beteiligten üblicherweise nicht die Artikel der Bewerbenden, sie zählen bloß, wie viele in welchen Zeitschriften aufgelistet werden. Teilweise werden die Bewerber*innen gebeten, Zitationszahlen anzugeben. Manche dieser Zahlen (z.B. Impact Factor) gehören Unternehmen an und der Zugang muss über die Institution erkaufte werden.

Zu diesen verheerenden Problemen kommen außerdem systemische Probleme der sexuellen Belästigung (Hoebel et al. 2022) und des Machtmissbrauches, die in dem aktuell streng hierarchischen Aufbau des Systems nur schwierig zu lösen sind (Forster and Lund (2018), siehe auch <https://www.netzwerk-mawi.de/> und <https://www.jmwiarda.de/2023/11/20/das-stille-leiden-der-betroffenen/>). Berichten des Netzwerkes gegen Machtmissbrauch in der Wissenschaft werden Fälle verschwiegen, und die schuldigen wechseln stillschweigend die Universität, sodass das Problem nicht gelöst wird. Wer es dabei besonders leicht hat, erklären Elsherif et al. (2022) anschaulich an dem *Academic Wheel of Privilege* (“Akademisches Rad der Privilegierten”; S. 85; siehe auch <https://www.psychologicalscience.org/observer/gs-navigating-academia-as-neurodivergent-researchers>). Beispielsweise haben Doktorandinnen in den Niederlanden vor allem dann schlechtere Noten als Doktoranden bekommen, wenn der Promotionsausschuss, also die Gruppe an Professor*innen, die die Promotion beurteilt, nur aus Männern bestand (Bol 2023). Schwerbehindertenquoten weit unter den [Quoten](#) anderer Berufe kommen ebenfalls hinzu.

9.1.3 Zu viel Forschung

Im Rahmen von Promotionen müssen Forschende in insgesamt 3-6 Jahren zuzüglich Elternzeiten üblicherweise drei wissenschaftliche Artikel veröffentlichen (bzw. in fairen Fällen drei veröffentlichungs-würdige Artikel vorweisen). Bei Post Docs (siehe Section 9.1.2) müssen es noch mehr sein. Dass Personen vor und nach der Promotion jeweils maximal 6 Jahre an Hochschulen angestellt sein dürfen ist gesetzlich festgelegt. Für die weitere Qualifikation *Habilitation*, für die eine ähnliche Zeit angesetzt ist, sind zum Beispiel in der Psychologie circa 6 Artikel die Daumenregel. Dabei spielt es eine nachrangige Rolle, wie umfangreich die Artikel sind. Beispielsweise dauert die Durchführung einer Meta-Analyse, in der bisherige Befunde zu einem bestimmten Thema systematisch gesammelt und statistisch zusammengefasst bzw. verglichen werden, oft mehrere Jahre. Eine Längsschnitterhebung kann je nach Forschungsfrage sogar Jahrzehnte dauern. Im Kontrast dazu lässt sich eine Querschnittserhebung über einen Online-Fragebogen in wenigen Wochen durchführen. Eine Doktorandin, die eine einzige Meta-Analyse durchführt, könnte damit nicht promovieren. Hätte sie stattdessen drei einfache Online-Studien durchgeführt und einzeln veröffentlicht, wäre es kein.² Diese

²Hier ließe sich einwenden, dass einige Zeitschriften nur Artikel veröffentlichen, in denen mehrere Studien durchgeführt wurden. Das Ziel, nämlich die *internale* Replikation der eigenen Befunde, verfehlen diese Zeitschriften damit deutlich. Stattdessen reizt es Forschende dazu an, mehrere Studien mit wenigen Versuchspersonen durchzuführen, statt eine Studie mit vielen Befragten.

willkürlichen Vorgaben haben dazu geführt, dass sich Wissenschaftler*innen alleine durch die Begutachtung der Artikel ihrer Kolleg*innen einen enormen Arbeitsaufwand auferlegen, der den wissenschaftlichen Fortschritt behindert (Hanson et al. 2023).

Zur Veranschaulichung des Aufgabenpensums nun ein Gedankenpiel: Angenommen es gäbe 10 Wissenschaftler*innen, die gemeinsam 10 Artikel im Jahr veröffentlichen würden - manchmal alleine, manchmal in einer Gruppe - und jeder der Artikel würde von zwei Personen begutachtet, so müsste jede*r zwei Artikel begutachten. Damit das System funktioniert, müsste jede Person die Anzahl der im Schnitt veröffentlichten Artikel mal die Anzahl der benötigten Gutachtenden begutachten. Bei 10 Veröffentlichungen pro Person und drei Gutachtenden wären es $10 \times 3 = 30$ Gutachten. Nun werden aber nicht alle Artikel von der Zeitschrift, bei der sie eingereicht werden, veröffentlicht, noch werden sie sofort veröffentlicht. Wissenschaftler*innen reichen ihre Artikel oft bei den "hochrangigsten" Zeitschriften ein. Nachdem dort mehrere Gutachter*innen den Artikel geprüft haben, wird er abgelehnt (Jaremka et al. 2020). Im Mindestfall werden Revisionen angefordert, welche oft eine weitere Runde Peer Review auslösen und nicht immer werden Artikel danach veröffentlicht. Unsere Rechnung geht also nicht auf: Nehmen wir *vor-sichtshalber* an, ein Artikel würde neun Mal begutachtet (z.B. einmal drei Gutachtende, dann Ablehnung, dann erneut drei Gutachtende, Revision, zweites Gutachten, Akzeptanz). Aus 10×3 wird 10×9 , bei etwas Urlaubszeit also etwas mehr als zwei Gutachten pro Woche, idealerweise bis zu zwei Arbeitstage. Bei dieser Rechnung bleibt weniger Zeit für Lehre, Wissenstransfer, Betreuung von Studierenden oder Promovierenden, Einwerbung von Forschungsgeldern, universitäre Selbstverwaltung, usw. Durch die vielen zu publizierenden Artikel und das strenge Review-System bürgt sich die Wissenschaft einen großen Berg Arbeit auf – einen der realistisch nicht machbar ist und unter dem am Ende die Qualität der Forschung leidet. Beispielsweise fiel es weder den Gutachtenden, noch den Herausgebern von Zeitschriften auf, dass in über 30 Artikeln mitten im Text „Regenerate response“ stand – ein Satz, der in OpenAIs ChatGPT Programm auf einem Knopf erlaubt, einen von einer künstlichen Intelligenz erstellten Text umzuformulieren (<https://retractionwatch.com/2023/10/06/signs-of-undeclared-chatgpt-use-in-papers-mounting/>). In manchen Artikeln hieß es sogar „As an AI language model, I ...“ (<https://pubpeer.com/search?q=As+an+AI+language+model%2C+I>). In einem Fall wurde der Artikel von dem Verlag Elsevier geändert, und zwar nicht auf dem empfohlenen Weg³ mittels eines transparenten *Errandum* oder *Corrigendum*, also einer öffentlichen Mitteilung über die Änderung und ihre Gründe, sondern ohne Erklärung oder Zustimmung der Autor*innen (<https://predatory-publishing.com/elsevier-changed-a-published-paper-without-any-explanation/>).

9.1.4 Publish or Perish – Veröffentlichen oder Verenden

Wer in der Wissenschaft arbeitet sollte die wichtigste Spielregel kennen: Wer überleben will, muss Artikel veröffentlichen. Kurz: Veröffentlichen oder Verenden (engl. *publish or perish*).

³Ethische Richtlinien im Publikationsprozess sind zum Beispiel verfügbar über das Committee on Publication Ethics (<https://publicationethics.org/guidance/Guidelines>).

Zur Promotion, Habilitation, Einwerbung von Forschungsgeldern, und zur Berufung auf eine Professur sind Veröffentlichungen das oberste Kriterium. Kennzeichen einer Währung ist, dass sich Dingen ein Wert zuweisen lässt. Wie sieht der Wert in der Forschung aus? Bibliometriker*innen entwarfen zur Beschreibung und zur Auswahl von Zeitschriftenabonnements (also explizit *nicht zur Bewertung*) verschiedener Forschungsgebiete verschiedene Kennzahlen, wie den Impact Factor, oder Hirsch-Index. Beide Zahlen bestehen aus Verrechnungen davon, wie oft Artikel je nach Zeitschrift oder je nach Person zitiert wurden. Beispielsweise werden beim Journal Impact Factor die Gesamtzahl der Zitationen in einem bestimmten Jahr durch die Anzahl der zitierbaren Veröffentlichungen in Bezugsjahren (z.B. den 3 vorangegangenen Jahren) geteilt. Zeitschriften werden mit hohen Impact Factors beworben und erlauben teilweise nicht die Zitation von Kommentaren (bzw. veröffentlichen Kommentare unter derselben Referenz wie Originalstudien), um möglichst hohe Impact Factors zu erhalten. Sie können natürlich auch entscheiden, ob sie den 3-Jahres- oder 2-Jahres Impact Factor berichten. Berechnungsweisen unterscheiden sich außerdem darin, auf Basis welcher Daten sie errechnet werden. Die Datenbank des Journal Impact Factors gehört dem Unternehmen Clarivate Analytics und ist nicht öffentlich zugänglich. Das heißt, die genauen Zahlen lassen sich nicht einfach nachrechnen und prüfen. Durch die wichtige und intransparente Rolle von Zitationsmetriken ist wenig überraschend, dass sie [manipuliert](#) werden. Klar ist auch, dass Zitationsmetriken entweder gar nicht oder negativ mit wissenschaftlicher Qualität zusammenhängen [Brembs (2018); Etzel et al. (2024); Table 3]. Zum Beispiel gab es das Problem, dass das Programm Microsoft Excel Genomnamen als Datumsangaben erkannt hat, umformatiert hat, und die eigentlichen Namen nicht mehr erkennbar waren. Somit waren Teile der jeweiligen Ergebnisse nutzlos (Ziemann, Eren, and El-Osta 2016). Dieses Problem kam vor allem bei angesehenen Zeitschriften vor. Sie haben nichts dagegen unternommen, stattdessen wurde Excel nach ein paar Jahren angepasst.

Neben Zeitschriften können auch Ranglisten von Personen erstellt werden. Forschende aus Stanford veröffentlichten eine solche Liste der *20 meistzitierten Forschenden*. Abgesehen davon, dass sie zum Missbrauch verführt, enthält zahlreiche Fehler (Abduh 2023) wie zum Beispiel Forschende, die hunderte Jahre lang Artikel veröffentlicht haben – vor und nach ihrem Tod. Unternehmen, denen Verlage und andere Werkzeuge für die Wissenschaft (z.B. Programme zur Verwaltung von Literatur) gehören, sammeln darüber hinaus Daten über die Forschenden (z.B. welche Artikel wie lange aufgerufen werden, welche Textpassagen markiert werden). Teilweise werden Bibliotheken aufgefordert, Überwachungsprogramme von Verlagen zu installieren. Die gesammelten Daten verkaufen die Verlage dann zurück an die Forschenden. Diese Praxis gefährdet die Wissenschaftsfreiheit, da Staaten Verlage auffordern können, die Namen von Forschenden zu nennen, die zu politisch brisanten Themen forschen. Die Initiative [Stop Tracking Science](#) setzt sich gegen das Verhalten ein.

Seit längerem wird für die verantwortungsvolle Verwendung dieser Metriken plädiert (Hicks et al. 2015) und zahlreiche Universitäten und Forschende tun sich zusammen um Bewertung von Forschung sinnvoll zu gestalten (z.B. mittels der [San Francisco Erklärung zur Forschungsbewertung](#)). Wie sich Anreizstrukturen und Karrierestatus auf wissenschaftliches Fehlverhalten auswirken, wird mit gemischten Ergebnissen untersucht. Das Problem: Sobald es in einem Sys-

tem ein klares Bewertungskriterium gibt, wird alles darauf ausgerichtet (*gaming the system*). Anreizstrukturen, die schlechte Forschung zur Folge haben herrschen in Bezug auf Bildmanipulationen in der Biologie laut Fanelli, Schleicher, et al. (2022) vor allem in China, weniger jedoch in USA, Großbritannien, und Kanada.

Ein weiteres Produkt der *Publish or Perish Struktur* ist, dass die meisten in der Psychologie entwickelten Instrumente zur Messung von Persönlichkeitseigenschaften nur wenige Male verwendet werden – und das auch hauptsächlich von ihren eigenen Entwickler*innen (Elson et al. 2023). Elson et al. (2023) plädieren: Psychologische Messinstrumente sind keine Zahnbürsten! Ein noch extremeres Ausmaß ist bei sogenannten *Paper Mills* zu sehen (van Noorden 2023): Personen erstellen dabei automatisiert große Mengen von wissenschaftlichen Artikeln, ohne die darin beschriebenen Untersuchungen wirklich durchzuführen. Wissenschaftler*innen können dann Ko-Autor*innenschaften kaufen, um ihre Anzahl an veröffentlichten Artikeln zu erhöhen und mehr Zitationen zu bekommen. Je nach Zeitschrift werden diese Artikel nicht im Peer-Review entlarvt. Es wird befürchtet, dass Käufer*innen solcher Artikel selbst sehr erfolgreich werden können, selbst Herausgeber von Zeitschriften werden, und sich dadurch selbst schützen, ertappt zu werden. Der genaue Ausmaß des Paper-Mill-Problems ist unklar und weitgehend unerforscht (Byrne and Christopher 2020). Ein Indiz, mithilfe künstlicher Intelligenz erstellte Artikel zu erkennen sind sogenannte *tortured phrases* (Cabanac, Labbé, and Magazinov 2021), welche grammatisch korrekt sind, im üblichen Sprachgebrauch jedoch selten vorkommen oder wenig Sinn machen.

9.1.5 Flaschenhals-Hypothese und Innovationsdrang

Namhafte wissenschaftliche Zeitschriften erhalten täglich unzählige Einreichungen, veröffentlichen aber nur eine begrenzte Anzahl an Artikeln. Sie müssen also streng selektieren, was begutachtet und gegebenenfalls veröffentlicht wird. Weil das Ziel einer Zeitschrift ist, viel gelesen zu werden, wählen Herausgeber*innen von Zeitschriften diejenigen Artikel, welche möglichst großes Potenzial haben, bekannt und viel zitiert zu werden (Giner-Sorolla 2012). Das betrifft zum Beispiel Beiträge mit besonderen praktischen Implikationen, überraschenden Befunden, oder besonders konsistenten Befunden. Studien, deren Ergebnisse keine eindeutigen Schlüsse zulassen – oder deren Autor*innen mit zu großer Vorsicht Schlüsse ziehen – kommen also nicht infrage. In den Neurowissenschaften kommunizieren manche Zeitschriften beispielsweise öffentlich, dass sie keine Replikationsstudien veröffentlichen und nach Neuheit selektieren, während die meisten keine Stellung dazu nehmen (Yeung 2017). In der Psychologie nahmen 2017 nur 33 von 1151 Zeitschriften Stellung dazu, dass sie Replikationen akzeptieren (Martin and Clarke 2017). Zwar werden innovative Befunde dann häufiger zitiert, die Zeitschrift erhält also mehr Leser*innen, mehr Einreichungen, und damit mehr Geld über Abonnements und Veröffentlichungskosten. Vielzitierte Artikel lassen sich jedoch schlechter replizieren als weniger zitiert (Serra-Garcia and Gneezy 2021) und prestigereiche Zeitschriften sind Magneten für fragwürdige Forschungspraktiken (Kepes et al. 2022) und nachweisbar gleichwertige oder sogar qualitativ schlechtere Forschung (Brembs 2018).

Wie kommt es dazu? Es ist die Rede von einem Flaschenhals (*Bottleneck*; viele Einreichungen aber wenige Veröffentlichungen). Das führt gemeinsam mit dem Anreiz in eben solchen Zeitschriften zu publizieren dazu, dass Forschende alle möglichen Mittel nutzen, um eine Chance auf eine Publikation zu erhalten. In manchen Instituten gilt, wer in einer bestimmten Zeitschrift veröffentlicht, erhält automatisch die Bestnote auf die Promotion. Andere Institute erklären schon in der Stellenausschreibung für eine Promotionsstelle, in welcher Zeitschrift die Ergebnisse des Forschungsprojektes veröffentlicht werden müssen. Die Tatsache, dass prestigereiche Zeitschriften *Nature* oder *Science* vor allem Artikel mit klaren Botschaften veröffentlichen, spornt also Forschende an, klare Ergebnisse zu erschaffen. Passt mal ein Befund nicht zu der geprüften Hypothese, wird er entweder manipuliert oder gar nicht veröffentlicht und landet in der Schublade.

9.1.6 Schubladen-Problem

Seit mehreren Jahrzehnten ist bekannt, dass Wissenschaftler*innen vor allem diejenigen Studien veröffentlichen, die ihre Theorien stützen (Rosenthal 1979; T. D. Sterling 1959). Im Extremfall hat jemand zum Prüfen einer Theorie fünf Studien durchgeführt, in nur einer davon die Theorie bestätigt, und nur diese veröffentlicht. Andere Forschende, die dann die (veröffentlichte) Literatur durchsuchen, sehen nur die “erfolgreiche” Studie. Es entsteht der Eindruck, dass die Theorie stimmt, während die Mehrheit der Studien diesen Schluss eigentlich nicht nahelegt. Dass Studienergebnisse natürlichen, statistischen Schwankungen unterliegen führt dazu, dass bei vielen Studien auch eine dabei sein kann, die das gewünschte Ergebnis zeigt. Durch das Schubladen-Problem konnten sich ganze Forschungsstränge entwickeln, die seit dem Bewusstsein für Replikationsstudien komplett ausgestorben sind (Brockman, 2022).

Für *Meta-Analysen*, also Studien, die bisherige Befunde zusammenfassen, wurden bereits verschiedene Methoden entwickelt, die Stärke des Schubladen-Problems (engl. *File-Drawer-Problem*) zu prüfen. Auch Methoden, diese Verzerrung zu korrigieren, existieren bereits vielzählige (Fisher, Tipton, and Zhipeng 2017; Hedges and Vevea 1996; Schimmack 2020; Simonsohn, Nelson, and Simmons 2014; van Aert and van Assen 2018). Allerdings funktioniert keine der Methoden in allen möglichen Szenarien (Carter et al. 2019). Um eine Veröffentlichung der fehlgeschlagenen Studien kommen Forschende also nicht herum.

In der Medizin gibt es den besonderen Fall, dass alle dort durchgeführten Studien öffentlich registriert werden müssen. Bei einer Veröffentlichung muss dann eine Registrierungsnummer angegeben werden. Über öffentliche Angaben zu registrierten Studien lässt sich somit nachverfolgen, welche Personen, Institutionen, oder Länder wie viele ihrer tatsächlich durchgeführten Studien veröffentlichen. Forschende in Berlin haben dazu eine Website mit einem sogenannten interaktiven *Dashboard* entwickelt, um die darüber gesammelten Daten durchsuchen und abbilden zu können (Franzen et al. 2023; Riedel et al. 2022). Auf <https://questcttd.bihealth.org/> ist nach aktuellem Stand (Juli 2024) sichtbar, dass von allen registrierten Studien nur 46% innerhalb der folgenden zwei Jahre und 74% innerhalb der folgenden fünf Jahre veröffentlicht wurden. Personen, die sich für die Studien als Versuchspersonen melden

oder Drittmittelgeber erhalten somit Aufschluss über die Größe der Schublade, in der die fehlgeschlagenen Studien und die “nicht so spannenden Ergebnisse landen”.

9.1.7 Zugängigkeit von Wissen

Durch die Einbindung von Forschung in das kommerzielle Verlagssystem befindet sich ein Großteil der Wissenschaften in einem sozialen Dilemma, das einen massiv eingeschränkten Zugang zum Wissen zur Folge hat. Das vorherrschende Modell bei wissenschaftlichen Zeitschriften, die zum Großteil Verlagen wie Springer, Elsevier, Sage, oder Taylor and Francis angehören, ist ein Abonnement-Modell. Hochschul-Bibliotheken zahlen Regelmäßig Geld an die Verlage, damit die Hochschul-Angehörigen (also Studierende und Mitarbeitende) Zugriff auf die darin veröffentlichten Arbeiten haben. Wer kein Abonnement hat, kann Artikel einzeln kaufen. Versucht man, einen Artikel online herunterzuladen, ohne dass man sich in einem Hochschulnetzwerk befindet, geht das nicht kostenlos: Der Artikel befindet sich hinter einer Bezahlschranke (*Paywall*). Soll ein Artikel für alle frei zugänglich veröffentlicht werden (*Open Access*, öffentlicher Zugang), kostet das extra, nämlich je Artikel zwischen 2000€ und 9000€. Um Forschung also lesen zu können, müssen Hochschulen Abonnements oder Artikel kaufen. Infolge dessen sind finanziell schlechter ausgestattete Institutionen, Länder, oder sogar gesamte Regionen wie der globale Süden in ihrer Teilnahme am internationalen Wissenschaftsdiskurs systematisch benachteiligt.

Das soziale Dilemma besteht nun in der Schwierigkeit, dieses System zu verändern. Brembs et al. (2023) beschreiben es wie folgt: Bibliotheken schließen die Abos mit Verlagen ab, um Forschung an ihren Hochschulen zu ermöglichen. Würden sie die Verträge kündigen, könnte das die Forschung verlangsamen und die Stellung einer Hochschule verschlechtern. Forschende können die namhaften Zeitschriften der kommerziellen Verlage nicht boykottieren, da sie sonst ihre Karriere gefährden würden. Sie sind darauf angewiesen, in prestigereichen Zeitschriften zu publizieren. Zudem ist es extrem schwierig, das Verhalten von Millionen von Forschenden, von denen die meisten viele Jahre lang mit dem aktuellen System gelebt haben, schlagartig zu verändern. Die Zeitschriften, als dritter Akteur in diesem Dilemma, profitieren von der Abhängigkeit und können Preise beliebig steigen lassen. Abgesehen vom Markenwert der Zeitschriften, also dem Ansehen und dem Vertrauen, das ihnen fälschlicherweise (Brembs 2018) entgegengebracht wird, tragen sie zu diesem System fast nichts bei. Universitäten und Länder haben bereits jetzt die Möglichkeit, Forschungsartikel online zu verwalten; die Begutachtung und Qualitätssicherung geschieht durch freiwillige Forschende und nicht durch Angestellte der Zeitschrift; und die Formatierung der Artikel können Forschende wie an bestehenden ⁴ ersichtlich ist (Carlsson et al. 2017), mit geringem Aufwand selbst übernehmen. Entsprechende Lösungen werden in Section 14.0.8 erläutert.

⁴Damit sind Zeitschriften gemeint, die Artikel ohne Abonnement-Zugang und ohne Paywall veröffentlichen, und zwar ohne Kosten für die Autor*innen.

9.2 Vertiefende Informationen

Der kostenlose Film “Paywall: The Business of Scholarship” behandelt das Thema Paywalls sowie den öffentlichen Zugang von wissenschaftlichem Wissen: <https://paywallthemovie.com/paywall>

9.3 Literatur

10 Karriere in der Wissenschaft

Ein Beruf in der Wissenschaft setzt üblicherweise einen einschlägigen Studienabschluss (Bachelor und Master) voraus und beginnt mit einer Promotion. Im Rahmen der Promotion wird das wissenschaftliche Handwerk erlernt und der Doktor*ingrad erlangt: Studien werden durchgeführt, Daten analysiert, und Forschungsartikel veröffentlicht. In vielen Fällen arbeiten Promovierende mit einer 50 – 66% Stelle als wissenschaftliche Mitarbeiter*innen, begutachten für ihre Vorgesetzten Artikel und Forschungsgelder-Anträge, schreiben eigene Anträge, verbringen mehrere Monate im Ausland, erstellen, beaufsichtigen, und korrigieren Klausuren, betreuen Abschlussarbeiten, und beteiligen sich an der universitären Selbstverwaltung durch die Teilnahme an Sitzungen und Mitgliedschaften bei Kommissionen. Zeitlich sind dabei bei Stellen oder Stipendien üblicherweise drei Jahre angesetzt. Am Anfang meiner Promotion sagte man mir, mit einer 40-Stunden-Woche schafft man es nicht, in der vorgesehenen Zeit zu promovieren – 50% Gehalt für über 100% Arbeitszeit und einem befristeten Vertrag also. Verträge sind dabei jedoch nicht immer auf die Zeit der Promotion befristet. Viele wissen erst wenige Monate vor Vertragsende, wie viel Prozent Gehalt sie erhalten, wie umfangreich die Lehrverpflichtung ist, und wie lange der nächste Vertrag geht. Kurz: Die Arbeitsbedingungen sind nicht optimal und man muss ein hohes Maß an Flexibilität mitbringen, wenn man sich für den Weg in die Wissenschaft entscheidet.

10.0.1 Doppelabhängigkeit

Für die Promotionsphase gibt es verschiedene Finanzierungsmöglichkeiten: Über Unternehmen lässt sich berufsbegleitend promovieren oder Stipendien zahlen über eine begrenzte Zeit Geld, das den Grunderhalt sichert (z.B. 1100€ über 36 Monate bei der Graduiertenförderung des Landes Sachsen-Anhalt, dazu kommen noch zusätzliche Kosten für die Sozialversicherung). Der häufigste Weg ist jedoch über eine Stelle als wissenschaftliche*r Mitarbeiter*in. Dabei ist die vorgesetzte Person diejenige Person, die auch die Promotionsarbeit bewertet. Die einem auferlegte Korrektur der 120 Erstsemester-Klausuren steht dann im Extremfall in Konflikt mit der Zeit, die für die Arbeit an der wissenschaftlichen Studie benötigt wird. Promovierende hängen also meistens von den betreuenden Professor*innen in Form der Beschäftigung *und* über die Bewertung ihrer Arbeit ab, was also als *Doppelabhängigkeit* bezeichnet wird.

Familienfeindliche Stipendien

Unbefristete Stellen, Doppelabhängigkeit oder geringfügige Stipendien, sowie Leistungsdruck haben klare negative Auswirkungen auf die Gesundheit der Forschenden und damit auf die Qualität der Wissenschaft. Wie sieht es mit Promotionsstipendien aus? Stipendien laufen oft 1 Jahr mit Option auf Verlängerung oder bis zu 3 Jahre. Forschende, die direkt nach der Schule und dem Studium promovieren, sind ca. 24 Jahre alt, haben eventuell einen Partner, und wünschen sich eine Familie. Im Rahmen ungewisser Vertragsverhältnisse ist das schwierig. Stipendien haben oft keine Möglichkeit zur Elternzeit und müssen dann pausiert werden – oder sie laufen weiter und es wird keine zusätzliche Zeit angehängt. Sie verbieten zusätzliches Einkommen, weshalb das Elterngeld dann auf den dafür angesetzten Mindestbetrag fällt. Zusätzlich schreiben Sie vor, dass die geförderte Person *inklusive Ehepartner* ein bestimmtes Maximalvermögen nicht überschreiten darf, sodass auch keine Finanzierung durch mögliche Rücklagen klappt.

10.0.2 Depressionen, Burnout, und #IchbinHanna

Fächerübergreifend hat sich als Antwort auf ein inzwischen gelöscht Erklärvideo vom Bundesministerium für Bildung und Forschung zum Wissenschaftszeitvertragsgesetz (WissZeitVG) eine Bewegung unter dem Hashtag #IchbinHanna entwickelt, die die dortige sachliche Erklärung (<https://www.youtube.com/watch?v=PIq5GIY4h4E>) und die Arbeitsbedingungen in der Wissenschaft stark kritisiert. Es heißt „damit auch nachrückende Wissenschaftlerinnen und Wissenschaftler die Chance auf den Erwerb dieser Qualifizierung haben und nicht eine Generation alle Stellen verstopft, dürfen Hochschulen und Forschungseinrichtungen befristete Verträge nach den besonderen Regeln des WissZeitVG abschließen. So kommt es zur Fluktuation und die fördert die Innovationskraft“. 90% aller Wissenschaftler*innen sind unbefristet angestellt (<https://www.youtube.com/watch?v=H1wJmqpGhJc>).

Planungsunsicherheit und massiver Konkurrenzdruck fördern die Innovationskraft? Das scheint unwahrscheinlich: Unter Forschenden mentale Erkrankungen wie Burnout (Jaremka et al. 2020) und Anzeichen für Angststörungen und Depression stark verbreitet. Einer Überblicksstudie von Satinsky et al. (2021) zufolge, leiden zwischen 18 und 31% der Promovierenden unter Depressionen.

Wer hilft?

Schnelle Hilfe bei psychischen Problemen leistet beispielsweise das Krisentelefon der [TelefonSeelsorge](#). Manche Universitäten haben spezifische Anlaufstellen, Städte haben häufig örtliche Psychotherapie-Ambulanzen und Beratungsstellen.

10.0.3 Top-Down-Wandel

Diejenigen, die das System ändern können, also alle mit unbefristeten Verträgen, leiden nicht mehr unter ihm. Für diejenigen, die unter dem System leiden, ist es unklug, das System ändern zu wollen und sich an die Professor*innen zu wenden, denn das sind die Leute, die über ihren späteren Verbleib in der Wissenschaft im Rahmen von Berufungskommissionen bei der Entscheidung der Vergabe von Professuren entscheiden. Im Extremfall kann das dazu führen, dass jemand Kritik an Arbeiten von Wissenschaftler*innen in höheren Positionen aus Angst, die eigenen Chancen auf eine Professur zu schmälern, zurückfährt. Auf der anderen Seite haben Professor*innen bewiesen, dass sie sehr gut nach den Spielregeln spielen können. Zuzugeben, dass sie eine der seltenen und heiß begehrten Stellen nicht über Qualität sondern Quantität ihrer Forschung erhalten haben, hieße, sich selbst abzuwerten.

10.0.4 Literatur

11 Probleme der wissenschaftlichen Methoden

Im Alltagsdenken herrscht noch oft der Mythos, dass Wissenschaft sich von Nicht-Wissenschaft durch *die wissenschaftliche Methode* unterscheide. Das ist falsch (Feyerabend 1975/2002). Zwar unterscheidet sich wissenschaftliches Wissen von alltäglichem Wissen (und auch Religion) durch einen höheren Grad an Systematizität (Hoyningen-Huene 2013), allerdings gibt es weder eine einzige noch eine konstante wissenschaftliche Methode. Methoden haben sich stattdessen über die Zeit gewandelt, und das ist auch gut so. Neue Technologien ermöglichen beispielsweise in der Physik hochpräzise Messungen mittels Elektronenlaser, in den Geschichtswissenschaften 3D-Scans von Artefakten, die sonst nur wenige sehen würden, oder in den Sozialwissenschaften Datenbanken mit freiwillig bereitgestellten und anonymisierten Chatverläufen (<https://db.mocoda2.de/c/home>).

So wie ein Hammer und andere Werkzeuge nicht gut oder schlecht sind, so sind auch Methoden nicht gut oder schlecht, nicht richtig oder falsch, sondern sie werden *angemessen und korrekt* verwendet oder missbraucht. Statt Missbrauch ist in den Sozialwissenschaften von fragwürdigen Forschungspraktiken (*Questionable Research Practices*), kurz QRP, die Rede. Sie erlauben Wissenschaftler*innen Befunde zu generieren, die sie wollen. Im Folgenden werden verbreitete und oft angewandte (John, Loewenstein, and Prelec 2012) Techniken vorgestellt (für einen Überblick über die Forschung dazu in den letzten 50 Jahren, siehe auch Neoh et al. (2023)).

11.0.1 Exploratorische versus konfirmatorische Forschung

Zum Verständnis der fragwürdigen Forschungspraktiken (QRP) ist eine wichtige forschungstechnische Unterscheidung unabdingbar: Wie bei einem Spaziergang kann eine wissenschaftliche Untersuchung erkundend oder zielgerichtet sein. Mal wird frei durch die Gegend spaziert und dabei neue Entdeckungen gemacht, mal ist das Ziel und die Route klar und im Vorhinein bestimmt. Im wissenschaftlichen Kontext ist die Rede von exploratorischer und konfirmatorischer Forschung. Bei der Exploration stehen höchstens die Forschungsfrage und grobe Züge der Methode fest, bei einem konfirmatorischen Test ist alles durchdacht: Vorgehen, mögliche Ergebnisse, sowie Erklärungsansätze für jedes mögliche Resultat. Es wird dann eine spezifizierte Hypothese mit der dazugehörigen Theorie bestätigt oder eben nicht. Kein Vorgehen ist dem anderen überlegen. Üblicherweise beginnen Untersuchungen in einem bisher wenig erschlossenen Gebiet mit Exploration, während mehr vorhergehende Forschung mit klareren Erwartungen einhergeht. Es sei dazu gesagt, dass es sich hier um Extremtypen von Forschung handelt, die ein Spektrum bilden. Es erlauben außerdem erst beide Ansätze zusammen Erkenntnisgewinn. Im Rahmen des hermeneutischen Zirkel, (einfach gesagt “dem

Kreis des Verstehens”) wird aus einzelnen Beobachtungen eine allgemeine Gesetzmäßigkeit formuliert (*Induktion*) und diese Gesetzmäßigkeit wird im Anschluss bei weiteren Einzelbeobachtungen geprüft (*Deduktion*). Die Deduktion kann je nach Gesetzmäßigkeit logisch notwendig sein, denn es wird von Prämissen auf eine Konklusion geschlossen. Wenn beide Prämissen korrekt sind, muss also auch die Schlussfolgerung korrekt sein. Die Induktion ist keine Notwendigkeit (siehe *Induktionsproblem*, Hume (1748/2011)).

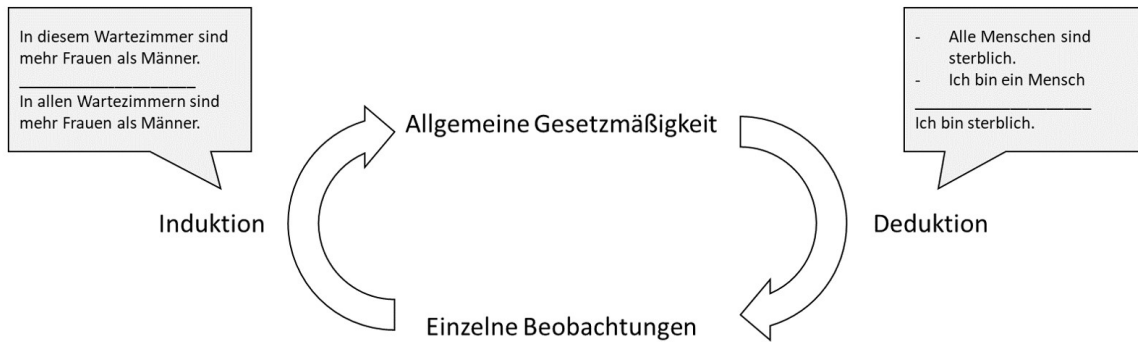


Figure 11.1: Hermeneutischer Zirkel

Problematisch wird es, wenn exploratorische Forschung als konfirmatorische kommuniziert wird, also so getan wird, als hätte eine Einzelbeobachtung eine bereits formulierte Gesetzmäßigkeit bestätigt, statt sie bloß inspiriert. Diese Art Unlogik heißt Zirkelschluss: Die Gesetzmäßigkeit gilt wegen der Beobachtung. Und die Beobachtung entspricht der Gesetzmäßigkeit.

Abbildung Z

Skizziertes Vorgehen bei explorativer (a) versus konfirmatorischer (b) Forschung. Exploratives Vorgehen ist nicht zielgerichtet, die Richtung kann sich ändern, und ist manchmal mit unvorhergesehenen Ergebnissen verbunden. Konfirmatorisches Vorgehen bildet oft einen engen und kontrollierten Ausschnitt eines Sachverhaltes ab.

11.0.1.1 Methoden der Datengenerierung

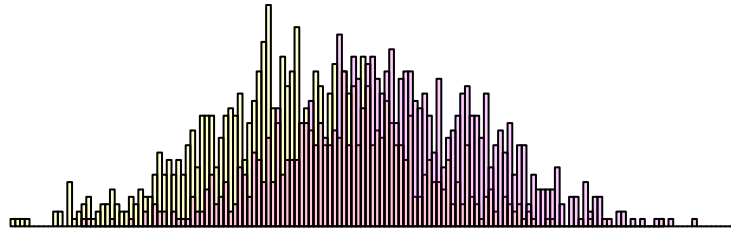
Wissenschaftliche Disziplinen bedienen sich für gewöhnlich vieler verschiedener Methoden. Idealerweise sind Erkenntnisse unabhängig von der Methode, die zu ihrer Entdeckung geführt hat und verschiedene Methoden führen zur selben Erkenntnis. Typische sozialwissenschaftliche Methoden sind die Befragung mittels standardisiertem Fragebogen, Verhaltensbeobachtung mittels Kameraaufzeichnungen und anschließender Kodierung von Verhaltensweisen durch

mehrere Personen, die den Untersuchungszweck nicht kennen, indirekte Methoden (Schimmack 2019), bei welchen etwas anderes gefragt wird, als was gemessen werden soll, Verhaltensmessungen wie Blickrichtungsmessung (*eye tracking*), oder Simulationsstudien, mittels derer zum Beispiel Verkehrsflüsse auf Basis vorher festgelegter Prinzipien per Computer berechnet werden oder Panikattacken vorhergesagt werden (Robinaugh et al. 2021). Diese Methoden generieren fast immer Daten, also beispielsweise eine Tabelle, in der je Beobachtungseinheit (z.B. je Versuchsperson) in mehreren Spalten Daten festgehalten werden und diese Daten werden fast immer statistisch ausgewertet. Die Notwendigkeit einer solchen Auswertung ergibt sich daraus, dass die Beobachteten Gesetzmäßigkeiten keine absoluten Gesetze im Sinne von „alle männlichen Babys wiegen mehr als alle weiblichen Babys“ sind, sondern statistische Regelmäßigkeiten im Sinne von „im Mittel wiegen kurz nach der Geburt männliche Babys ein paar hundert Gramm mehr als weibliche Babys, aber nicht jedes männliche Baby wiegt mehr als jedes weibliche Baby“ sind. Ein Vergleich von Körpergrößen nach Geschlecht ist über [Statista](#) zu sehen.

In der folgenden Abbildung sind die Häufigkeiten verschiedener Werte zu sehen (*Histogramm*). Je weiter rechts ein Wert ist, desto höher ist der Wert (z.B. Geburtsgewicht) und je höher der Balken ist, desto häufiger kommt der Wert vor. Die gelbe und die lilane Verteilung überlappen einander, das heißt, dass nicht alle gelben Werte niedriger als alle lilanen sind. Im Mittel, sind die gelben Werte jedoch geringer.

```
set.seed(1)
x <- rnorm(1000, mean = 0)
y <- rnorm(1000, mean = 1)
hist(x, xlim = c(-3, 5), col = rgb(1,1,0,.2), ylim = c(0, 40), breaks = 100, xaxt = "n", yaxt = "n",
     , xlab = "Wert", ylab = "Häufigkeit", main = "")
hist(y, add = T, col = rgb(1,0,1,.2), breaks = 100)
```

Häufigkeit



Wert

Statistische Signifikanz

Eine der am weitesten verbreiteten Methoden in den Sozialwissenschaften (und auch darüber hinaus) ist Statistik, genauer *Inferenzstatistik*. Dabei wird von einer begrenzten Menge von Beobachtungen (z.B. ausgefüllte Fragebogen von 100 Personen) auf alle möglichen Beobachtungen (z.B. alle Menschen) verallgemeinert. Untersuchte Zusammenhänge sind selten eindeutig, es gibt aber häufig statistische Regelmäßigkeiten. Charakteristisch ist dabei ein gewisses *Zufallselement*. Wiegt man ein kürzlich geborenes männliches und weibliches Baby, dann ist die Wahrscheinlichkeit sehr hoch, dass das männliche Baby mehr wiegt. Es kommt aber auch häufig vor, dass das nicht der Fall ist. Ähnlich verhält es sich bei einer fairen Münze, also einer die im Mittel gleich häufig auf Kopf und auf Zahl landet: Dass sie bei insgesamt vier Würfeln immer auf Kopf landet ist unwahrscheinlich, dass sie 1 oder 2 Mal auf Kopf landet, es kommt aber auch vor (nämlich in 12,5% aller Fälle, in denen eine faire Münze vier Mal hintereinander geworfen wird).

Inferenzstatistische Tests gehen nun davon aus, dass bei der Betrachtung eines statistischen Zusammenhanges (z.B. Geschlecht und Geburtsgewicht, Körpergewicht und Größe, Bildungsniveau der Eltern und Bildungsniveau der Kinder) „nur der Zufall am Werk ist“ (Röseler and Schütz 2022). Unter der Annahme wird berechnet, wie häufig ein beobachteter Zusammenhang mit der beobachteten Stärke vorkommen würde, wenn *eigentlich* kein Zusammenhang vorliegen würde. Also zum Beispiel „dass eine faire Münze vier Mal auf Kopf landet, passiert in 12,5% aller Fälle“. Bei sechs Würfeln wären

es 1,5625%. Die Kunst des statistischen Schließens besteht nun darin, den Punkt zu finden, ab dem Forschende davon ausgehen, dass der Zufall nicht am Werk war, weil die berechnete Wahrscheinlichkeit so gering ist. Konventionell liegt dieser bei 5%, für neue Befunde manchmal bei 0,5% (Benjamin et al., 2018), und in besonders prekären Fällen noch niedriger. Fachtechnisch wird von einem *Alpha-Niveau* oder einem *Signifikanzniveau* gesprochen und die berechnete Wahrscheinlichkeit heißt *p*-Wert. P-Werte unter 5% werden *statistisch signifikant* oder *auf dem 5%-Niveau signifikant* bezeichnet. Forschende würden also sagen, dass eine Münze *nicht* fair ist, wenn sie sechs Mal hintereinander auf Kopf landet (sogar schon bei fünf Mal, was in 3,125% der Fälle vorkommt). Dabei nehmen sie in Kauf, dass sie, wenn die Münze eigentlich doch fair ist, in 5% aller Fälle falsche Schlüsse ziehen.

Auf der anderen Seite ist es durchaus möglich, dass eine Münze nicht fair ist, zum Beispiel in 60% der Fälle auf Kopf landet und in 40% auf Zahl.

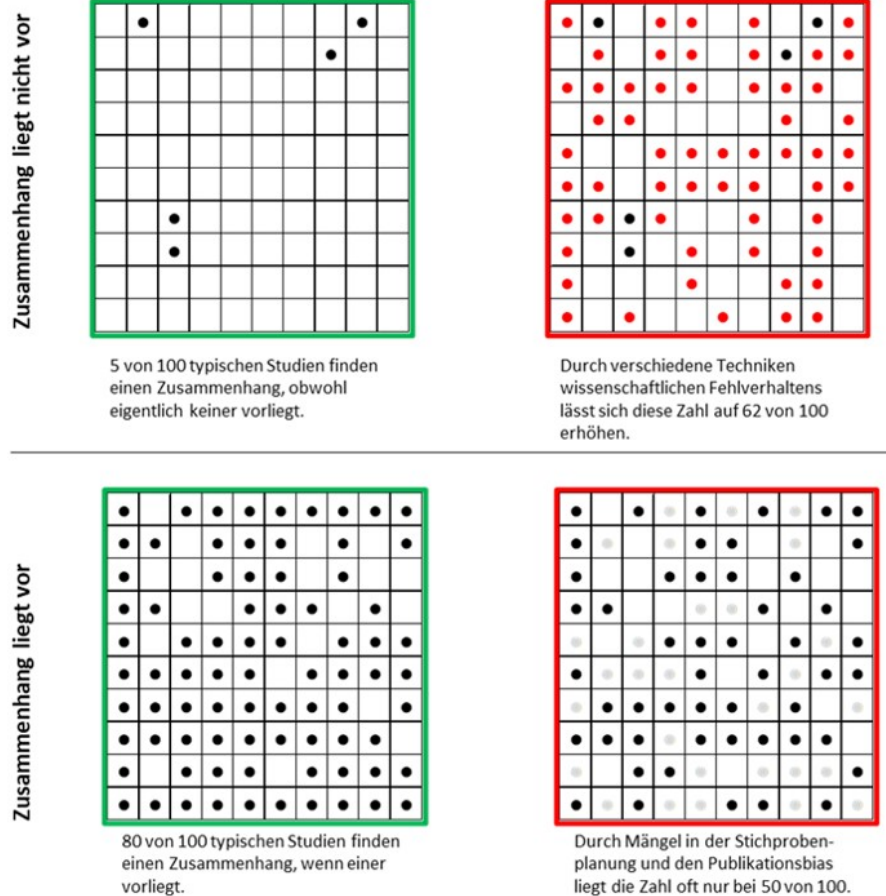


Figure 11.2: Eine einzelne Studie führt noch nicht zu sicherer Erkenntnis. Auch, wenn ein untersuchter Zusammenhang nicht vorliegt, kann er in Daten zufällig aufscheinen. Und auch, wenn eigentlich ein Zusammenhang vorliegt, kann dieser in den Daten durch Zufallsschwankungen nicht zu erkennen sein. Mithilfe von Open Science Praktiken soll der Zustand in den linken Kästen wiederhergestellt werden. Aus Röseler, L. (2021). Wissenschaftliches Fehlverhalten [Abbildung]. <https://osf.io/uf7gz/>. Lizenziert unter CC BY-Attribution International 4.0.

11.0.2 Freiheitsgrade von Forschenden (*Researchers' Degrees of Freedom*)

Vollständige Studien mehrfach durchzuführen ist sehr aufwändig. Obwohl es ein relativ sicherer Weg zu signifikanten P-Werten ist, gibt es weitaus sparsamere Lösungen. Die meisten Analysen sind um ein vielfaches komplexer als die oben beschriebene Münzwurfstudie. Betrachten

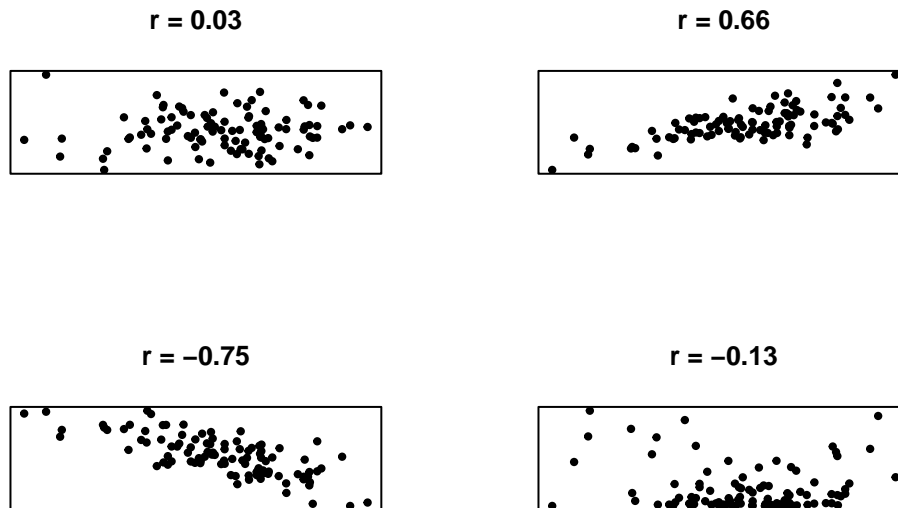
wir den immer noch sehr einfachen Signifikanztest für einen Korrelationskoeffizienten. Der Koeffizient ist eine Zahl zwischen -1 und 1 und beschreibt die Art des Zusammenhanges zwischen zwei Variablen (z.B. Einkommen und Lebenszufriedenheit). 0 bedeutet, dass kein Zusammenhang vorliegt, positive Werte bedeuten, dass wenn die eine Variable hohe Werte hat, dann hat auch die andere hohe Werte, und negative Korrelationen bedeuten, dass wenn eine Variable hohe Werte hat, dann hat die andere Variable eher niedrige Werte. In Abbildung X sind verschiedene Korrelationen dargestellt.

Verschiedene Zusammenhänge zwischen zwei Variablen und deren Korrelationskoeffizienten (simulierte Daten).

```
layout(matrix(c(1, 1, 2, 2, 3, 3, 4, 4), nrow = 2, byrow = TRUE))
set.seed(42)
n <- 100

x1 <- rnorm(n)
y1 <- rnorm(n)
y2 <- x1+rnorm(n)
y3 <- -x1+rnorm(n)
y4 <- abs(x1+rnorm(n))^2

plot(x1, y1, main = paste("r = ", round(cor(x1, y1), 2), sep = ""),
     , xaxt = 'n', yaxt = 'n', xlab = "", ylab = "", pch = 20)
plot(x1, y2, main = paste("r = ", round(cor(x1, y2), 2), sep = ""),
     , xaxt = 'n', yaxt = 'n', xlab = "", ylab = "", pch = 20)
plot(x1, y3, main = paste("r = ", round(cor(x1, y3), 2), sep = ""),
     , xaxt = 'n', yaxt = 'n', xlab = "", ylab = "", pch = 20)
plot(x1, y4, main = paste("r = ", round(cor(x1, y4), 2), sep = ""),
     , xaxt = 'n', yaxt = 'n', xlab = "", ylab = "", pch = 20)
```



layout(1)

i Korrelation

r bedeutet bei statistischen Berichten, dass es sich um einen *Korrelationskoeffizienten* handelt, hier nämlich die Produkt-Moment-Korrelation (oder auch Bravais-Pearson Korrelation). Diese sind standardisierte Werte für jeweils zwei Variablen. Das heißt, dass sich für eine möglichst große Anzahl an paarweisen Beobachtungen die Korrelation aller möglichen Dinge miteinander berechnen lässt. Korrelationskoeffizienten liegen zwischen -1 und 1. Werte unter 0 bedeuten, je höher das eine, desto niedriger das andere (negativer Zusammenhang). Werte über 0 bedeuten, je höher das eine, desto höher das andere (positiver Zusammenhang). 0 bedeutet, dass beide beteiligten Variablen unabhängig voneinander sind. Die 98 in der Klammer ist die Anzahl der Beobachtungen minus 2 (Freiheitsgrade). Der Korrelationswert von .420 (bzw 0,42) bedeutet, dass ein positiver Zusammenhang beobachtet wurde. Wichtig ist außerdem, dass es nur um insgesamt positive oder negative Zusammenhang geht (*Linearität*). Falls also beispielsweise ein u-förmiger Zusammenhang vorliegt (unten rechts in der Abbildung), wird sich dieser nicht in der Korrelation niederschlagen.

Obwohl es sich hierbei um einen sehr einfachen Test handelt, bringt er viele Entscheidungen mit sich. Selbst nach der Datenerhebung muss entschieden werden: Welche der befragten Personen werden für den Test verwendet? Sollen Personen ausgeschlossen werden und falls ja, warum (z.B. extreme Werte oder unplausible Werte)? Wie werden die Werte der Variablen

berechnet? Welche Art der Korrelation soll verwendet werden (z.B. Bravais-Pearson, Kendall, oder Spearman)? Gibt es eine Erwartung der Richtung der Korrelation (Gerichtetheit der Hypothese)?

Diese Fragen entsprechen Freiheitsgraden – Forschende sind also dahingehend flexibel, welche Optionen sie wählen. Keine der Optionen ist per se allen anderen überlegen und jede Entscheidung lässt sich in einem gewissen Rahmen rechtfertigen. Das Problem dieser Flexibilität ist, dass die Ergebnisse von ihr abhängen und je nach den Entscheidungen kann das Ergebnis eine positive, negative, oder keine Korrelation bedeuten. Je komplexer die Untersuchung und das statistische Verfahren ist, desto größer ist auch die Flexibilität bei der Datenanalyse. An sich sind diese Freiheitsgrade nichts Schlechtes, problematisch wird es bloß dann, wenn nur diejenigen Ergebnisse dargestellt werden, die sich gut veröffentlichen lassen oder zu den Überzeugungen der Forschenden passen. Dieses Vorgehen heißt HARKing (hypothesizing after the results are known = Hypothesen aufstellen, nachdem die Ergebnisse bekannt sind) und stellt einen Zirkelschluss dar. Die Hypothese, die geprüft wurde, stammt aus den Daten, die sich natürlich bestätigen. Verschiedene Lösungswege erlauben auch die Reduktion oder komplette Auslöschung von Freiheitsgraden (z.B. Präregistrierung, siehe Kapitel XXX). Auch ist es möglich, das Vorgehen als explorativ, also nicht vorher durchdacht und vorbestimmt, zu kommunizieren.

Im Datenanalyseprozess wird die Analogie des „garden of forking paths“ verwendet. In einem vereinfachten (!) Beispiel in Abbildung 4 haben wir $3 \times 4 \times 4 \times 4 = 192$ verschiedene Ergebnisse, die das gesamte Spektrum der Schlussfolgerungen abdecken werden – egal, ob unsere Hypothese stimmt oder nicht.

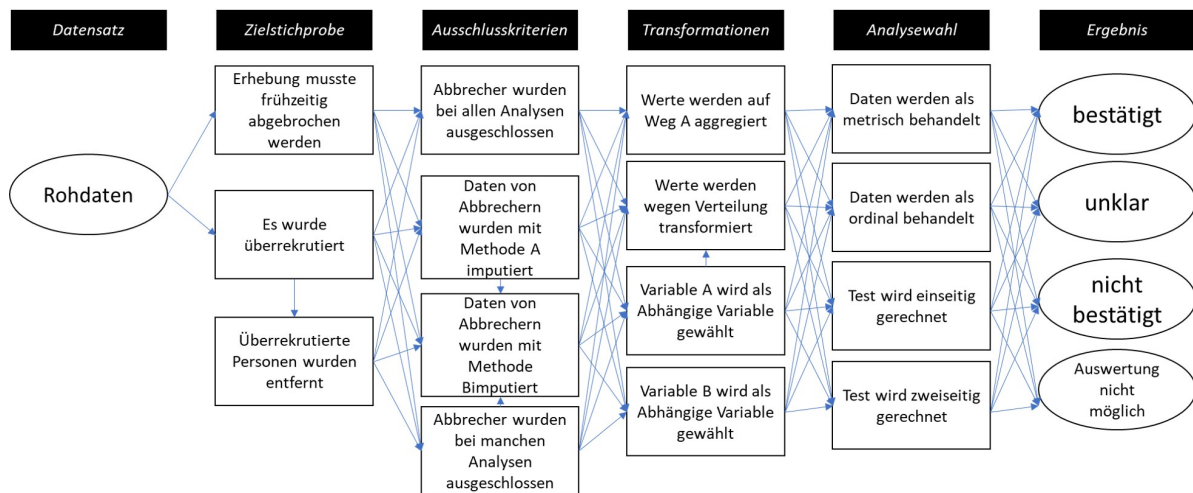


Figure 11.3: 192 verschiedene Wege von einem Datensatz zum (gewünschten) Ergebnis

Demonstrationen des *garden of forking paths* existieren für verschiedenste Felder und wurden bereits für Evolutionsbiologie (Gould et al. 2023), Sozialpolitik (Brenzau et al. 2022),

Strukturgleichungsmodelle (Sarstedt et al. 2024), und Sprachanalysen (Coretta et al. 2023) nachgewiesen.

11.0.3 Tippfehler

Häufig werden Daten mittels fortgeschrittener Programme ausgewertet und die Ergebnisse werden dann mühselig in den Bericht übertragen. Hierbei entstehen schnell Tippfehler. Michèle B. Nuijten et al. (2016) erstellten einen Algorithmus, der automatisch berichtete Signifikanztests erkennt, nachrechnet, und Inkonsistenzen zurückmeldet. Dabei fanden sie heraus, dass in großen Psychologiezeitschriften zwischen 1985 und 2013 in ungefähr der Hälfte aller Artikel mindestens ein Fehler war. Diese “Tippfehler” waren dabei nicht völlig zufällig, sondern fehlerhafte Werte waren meistens zugunsten positiver Befunde. Solche Übertragungsfehler passieren auch bei Meta-Analysen (Lopez-Nicolas et al. 2022). Und selbst bei Zitationen sind Fehler nicht selten: Über verschiedene wissenschaftliche Disziplinen fanden (Smith and Cumberledge 2020), dass in 25% aller untersuchten Zitationen, die zitierten Behauptungen in den Originalartikeln nicht vertreten wurden.

11.0.4 P-Hacking

Der P-Wert bei statistischen Tests gibt an, wie hoch die Wahrscheinlichkeit für das beobachtete Muster ist, gegeben eines vorausgesetzten Musters. Für eine Korrelation heißt das: Wie wahrscheinlich ist es, eine Korrelation der vorgefundenen Höhe zu beobachten, wenn eigentlich kein Zusammenhang (also $r = 0$) zwischen den untersuchten Variablen besteht. Konkret könnte das heißen: Wie wahrscheinlich ist es, dass in meinem Datensatz von 100 Personen die Korrelation zwischen Intelligenz und Alter genau $r(98) = .420$ ist, wenn ich eigentlich davon ausgehen, dass beide Variablen nicht zusammenhängen.

Die im Signifikanztest mitformulierte Annahme des fehlenden Zusammenhanges heißt *Nullhypothese*. Wenn das beobachtete Muster gegeben der Nullhypothese extrem unwahrscheinlich ist (oft unter 5%) wird von einem statistisch signifikanten Zusammenhang gesprochen. Wichtig ist dabei, dass Signifikanz (also „Bedeutsamkeit“) hier wirklich nur im statistischen Sinne zu verstehen ist. Die Frage, wie bedeutsam ein Befund für die Welt und das Leben ist, lässt sich mit Statistik in diesem Rahmen nicht beantworten. Weil P-Werte Wahrscheinlichkeiten sind, liegen sie zwischen 0 und 100%.

Unter den QRPs (fragwürdigen Forschungspraktiken) ist p-hacking eine weitere Kategorie, die wiederum selbst verschiedene Techniken beinhaltet. Mit p-hacking ist gemeint, dass Forschende ihre Freiheitsgrade nutzen, um den P-Wert „signifikant zu machen“, also unter 5% zu bringen. Eine oft fälschlicherweise gemachte Annahme zu P-Werten ist, dass hohe P-Werte für die *Abwesenheit* eines Zusammenhanges sprächen, oder dass P-Werte nur dann niedrig sind, wenn tatsächlich ein Zusammenhang vorliegt. Stattdessen sind P-Werte tendenziell klein, wenn ein Zusammenhang vorliegt, der auch mit der Menge der erhobenen Daten

nachgewiesen werden kann. Wenn kein Zusammenhang vorliegt, sind P-Werte gleichverteilt, das heißt, alle P-Werte kommen gleich häufig vor. Im Sinne der oben genannten Definition ist a priori klar, dass bei 100 durchgeführten Studien tendenziell 5 einen signifikanten Zusammenhang aufweisen, *wenn eigentlich keiner vorliegt*. Diese Tatsache erlaubt diverse P-Hacking Methoden. Simonsohn et al. (2014) zeigten, die Wahrscheinlichkeit, ein signifikantes Ergebnis zu kriegen, wenn eigentlich kein Zusammenhang in den Daten herrscht, von 5% auf ca. 60% steigen kann. Abbildung X zeigt die Verteilung von P-Werten bei verschiedenen hoher Teststärke (bzw. Power: der Wahrscheinlichkeit, einen Zusammenhang einer bestimmten Größe zu finden, wenn es ihn tatsächlich gibt).

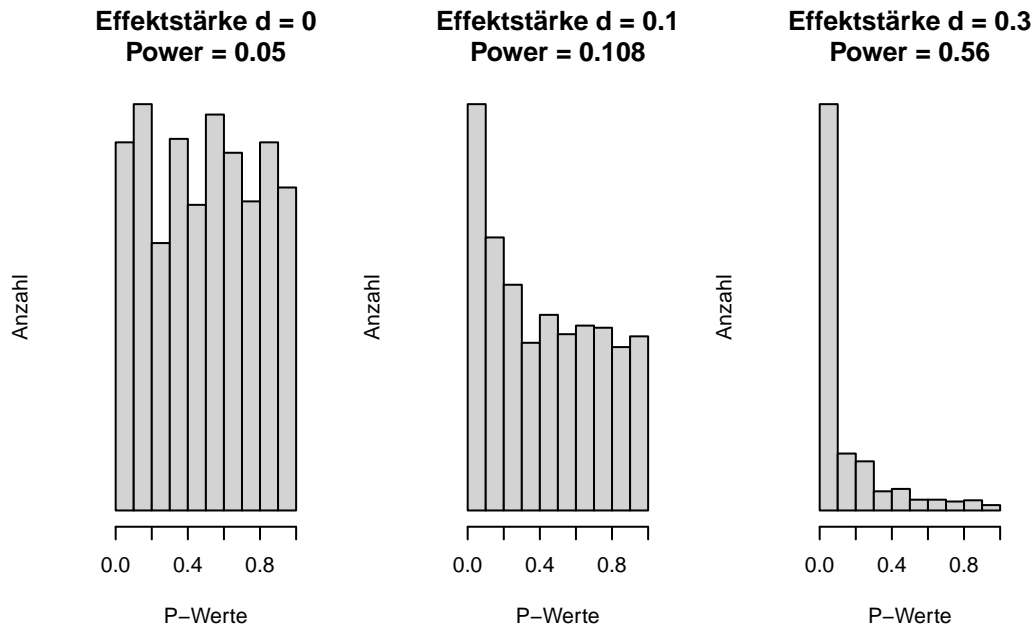
P-Werte sind bei Abwesenheit von Unterschieden oder Zusammenhängen, also beim Gelten der Nullhypothese gleichverteilt. Je höher die statistische Teststärke (Power), desto weiter verschiebt sich die Verteilung in den Bereich statistischer Signifikanz.

```
# P-Value distribution -----
layout(matrix(c(1,2,3), nrow = 1))

effects <- c(0, .1, .3)

for (i in effects) {
  effect <- i
  n <- 100
  pvalues <- (replicate(1000, t.test(rnorm(100), rnorm(100, i))$p.value))
  power <- round(pwr::pwr.t.test(n = n, d = i, power = NULL, alternative = "two.sided")$power)

  hist(pvalues
    , xlab = "P-Werte"
    , ylab = "Anzahl"
    , main = paste("Effektstärke d = ", i, "\nPower = ", power, sep = "")
    , xlim = c(0, 1)
    , yaxt = 'n')
}
```



layout(1)

Die Chance, signifikante P-Werte zu kriegen, ohne, dass die getestete Hypothese überhaupt stimmt, lässt sich durch „zerschneiden“ der Stichprobe machen (z.B. werden nur Frauen analysiert), durch das Erheben zusätzlicher Daten („optional stopping“), oder durch die Verwendung mehrerer zentraler Variablen (zum Beispiel wird Intelligenz mit 3 verschiedenen Tests erfasst und alle werden einzeln mit Alter korreliert). Selbst das verändern kleiner Parameter in den statistischen Tests (z.B. Verwendung einer nicht-parametrischen Spearman Korrelation statt der Bravais-Pearson Korrelation) erhöhen die Chancen auf ein signifikantes Ergebnis (siehe Tabelle Y). Einige Formen des *p-hacking* lassen sich zum Beispiel hier ausprobieren: <https://shinyapps.org/apps/p-hacker/> (Schönbrodt 2016). Eine Checkliste, um P-Hacking zu vermeiden, schlägt Wicherts et al. (2016) vor.

Wahrscheinlichkeit für ein signifikantes Ergebnis durch die Anwendung verschiedener P-Hacking Techniken nach Simmons, Nelson, and Simonsohn (2011), Tabelle 1. Der Anteil signifikanter Ergebnisse sollte hier den festgelegten 5% des Alpha-Fehlers entsprechen.

Technik	Anteil signifikanter Ergebnisse
Mehrere abhängige Variablen, deren Korrelation $r = .5$ ist	9,5%
10 weitere Beobachtungen je Gruppe erheben	7,7%

Technik	Anteil signifikanter Ergebnisse
Eine weitere Variable (z.B. Geschlecht) mit in das Modell aufnehmen	11,7%
Ausschließen (bzw. Beibehalten) einer von drei Gruppen	12,6%
Alle Techniken gleichzeitig	60,7%

💡 Forschung kann nicht witzig sein?

Hussey and Hughes (2018) (und darauf aufbauend Sarstedt and Adler (2023)) schlugen bereits Methoden vor, um P-Hacking noch stärker zu erleichtern. Auf dieser Website können Benutzer*innen Zufallszahlen in dem gewünschten Bereich generieren: https://mktg.shinyapps.io/extra-p_ointless/

11.0.5 Selektives Berichten (*Selective Reporting*)

Im Rahmen der Planung einer sozialwissenschaftlichen Studie stellt sich oft die Frage, wie ein bestimmtes Konstrukt gemessen werden soll. Für Intelligenz, politischer Ansicht, Lebenszufriedenheit, und viele andere Variable gibt es nicht *den* Test sondern viele Maße, die teilweise gering miteinander zusammenhängen. Gleichzeitig sind die zu testenden Theorien meist vage und diktieren nicht, mit welchem Maß ein Konstrukt gemessen werden sollte. Theorien sind den Messmethoden gegenüber also oft agnostisch. Werden in einer Studie dann verschiedene Messmethoden für ein Konstrukt gewählt, müsste die Theorie über alle Tests gleichermaßen bestätigt werden. Falls das nicht der Fall ist, sollte die Theorie angepasst werden. Entgegen dieser Empfehlung und um die Chance der Publikation der Ergebnisse zu maximieren, berichten Forschende Ergebnisse oft *selektiv*. Statt aller Ergebnisse werden also nur die „passendsten“ oder „spannendsten“ berichtet. Wie oben im Thema P-Hacking und Freiheitsgrade von Forschenden klar geworden ist, führt das dazu, dass Zusammenhänge gefunden werden, die eigentlich nicht existieren. Werden zum Beispiel drei verschiedene und unabhängige Maße zum Testen einer Hypothese verwendet steigt Wahrscheinlichkeit für mindestens ein signifikantes Ergebnis von 5% auf 14%.

Abbildung K

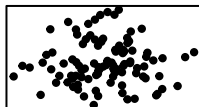
Selektives Berichten: Von den sechs geprüften Korrelationen ist nur eine signifikant - und das rein zufällig. Alle gemeinsam sind nicht signifikant. Um die Ergebnisse zu veröffentlichen berichten Forschende nur den spannendsten Teil der Ergebnisse und verzerren damit das Bild. Die Chance, aus sechs Tests - bei denen keine Zusammenhänge vorliegen - einen signifikanten zu erhalten liegt bei ca. 27%.


```
# mean(replicate(1000, mean(replicate(6, cor.test(rnorm(50), rnorm(50))$p.value) < .05) > 0))
set.seed(1)
n <- 100

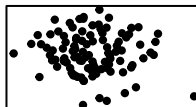
layout(matrix(c(1:6), nrow = 2))
for (i in 1:6) {
  x <- rnorm(n)
  y <- rnorm(n)

  plot(x, y
       , main = paste("r(", n-2, ") = ", round(cor(x,y), 3), ", p = ", round((cor.test(x,y)$
       , pch = 20, yaxt = 'n', xaxt = 'n', xlab = "", ylab = ""))
}
```

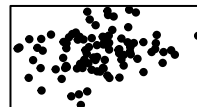
$r(98) = -0.001, p = 0.992$



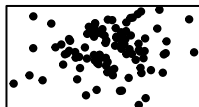
$r(98) = -0.091, p = 0.369$



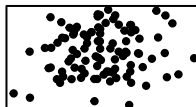
$r(98) = 0.199, p = 0.047$



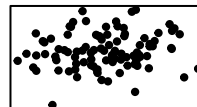
$r(98) = 0.111, p = 0.273$



$r(98) = 0.099, p = 0.327$



$r(98) = 0.119, p = 0.24$



```
layout(1)
```

11.0.6 Optionales Stoppen (Optional Stopping)

Führt man bei der Durchführung einer Studie nach jeder Beobachtung den Test erneut aus und betrachtet den P-Wert, dann gibt es zwei Möglichkeiten zum Verlauf der P-Werte: Falls ein Zusammenhang zwischen den erhobenen Variablen besteht, wird der P-Wert *konvergieren*, also

sich einem bestimmten Wert annähern, nämlich 0. Die Wahrscheinlichkeit für das beobachtete Ergebnis wird mit größerer Stichprobe immer geringer. Dass eine Münze nur auf “Kopf” landet ist ungewöhnlicher, wenn sie das 100 Mal getan hat, als wenn sie das 3 Mal tat. Falls kein Zusammenhang vorliegt, wird der P-Wert nicht wie oft erwartet gegen 1 gehen, sondern *nicht konvergieren*. Er wird dann chaotisch mal hoch und mal niedrig sein – und auch öfter mal signifikant. Diese Tatsache machen sich Forschende beim optionalen Stoppen zu Nutzen: Sie erheben so lange Daten, bis ihre Hypothese bestätigt wird. Das Problem besteht übrigens nicht für Effektstärkemaße wie zum Beispiel Korrelationen. Diese konvergieren je nach Größe ab ungefähr 250 Beobachtungen (Schönbrodt & Perugini, 2013).

Abbildung P

Konvergenz von P-Werten und Effektstärken je nach Effektgröße: Effektstärken (hier: Korrelationskoeffizienten) konvergieren bei großen Stichproben, P-Werte konvergieren nur, wenn die Korrelation nicht 0 ist.

```
# P-Value Convergence -----
imax <- seq(5, 1000, 10)
p0 <- NULL
p1 <- NULL
p2 <- NULL

for (i in imax) {

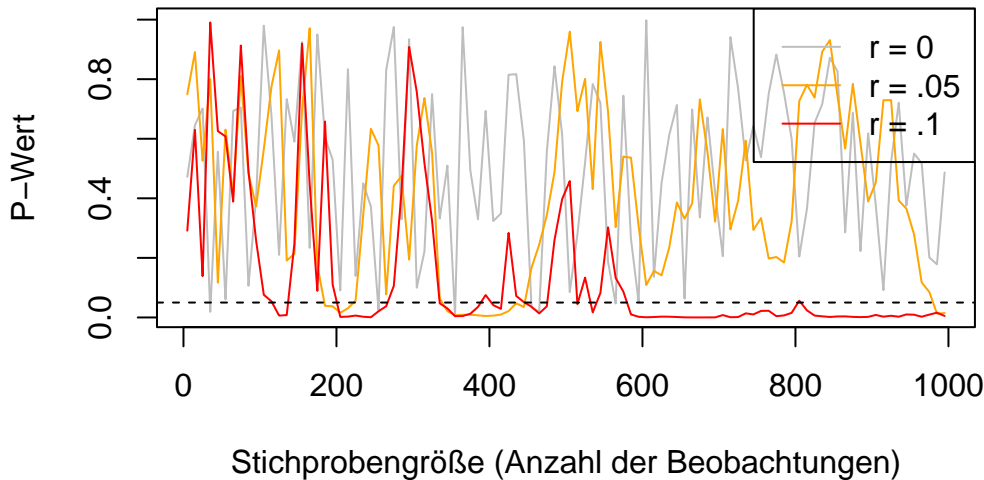
  set.seed(42)

  ds0 <- MASS::mvrnorm(n = i, mu = c(0,0), Sigma = matrix(c(1, 0, 0, 1), nrow = 2))
  ds1 <- MASS::mvrnorm(n = i, mu = c(0,0), Sigma = matrix(c(1, .05, .05, 1), nrow = 2))
  ds2 <- MASS::mvrnorm(n = i, mu = c(0,0), Sigma = matrix(c(1, .1, .1, 1), nrow = 2))

  p0 <- c(p0, cor.test(ds0[, 1], ds0[, 2])$p.value)
  p1 <- c(p1, cor.test(ds1[, 1], ds1[, 2])$p.value)
  p2 <- c(p2, cor.test(ds2[, 1], ds2[, 2])$p.value)

}

plot( y = p0, x = imax, type = "l", col = "grey", xlab = "Stichprobengröße (Anzahl der Beobachtungen)", ylab = "P-Wert")
lines(y = p1, x = imax, col = "orange")
lines(y = p2, x = imax, col = "red")
abline(h = .05, lty = 2)
legend("topright", c("r = 0", "r = .05", "r = .1"), col = c("grey", "orange", "red"), lty = c(1, 2, 1))
```



11.0.7 Darstellung kalibrierter Modelle als geplante Modelle (*Overfitting*)

Komplexe statistische Modelle haben viele Stellschrauben. Es ist möglich, die unzähligen Entscheidungen vor Anwendung eines Modells auf Daten zu treffen, für gewöhnlich werden aber andere Kalibrationen ausprobiert und eine andere als die geplante hat eine bessere Passung. Damit ist gemeint, dass beispielsweise bestimmte Variablen mit in ein Modell aufgenommen werden, um die Vorhersagekraft zu maximieren. Zu vielen Modellen gehören sogar verschiedene Algorithmen, die auf Basis festgelegter Regeln entscheiden, wie das Modell aussehen soll. Ein Modell wird also an ein Datenmuster angepasst. Wird das Vorgehen transparent offengelegt, ist das absolut in Ordnung. Problematisch wird es, wenn das beste gefundene Modell als geplantes Modell dargestellt wird. Das in den Daten vorliegende Muster beinhaltet in sozialwissenschaftlichen Untersuchungen nämlich fast immer auch ein Rauschen, also Schwankungen, die auf Messungenauigkeiten oder andere unbekannte Einflüsse zurückzuführen sind. Diese Einflüsse schwanken definitionsgemäß (in der psychologischen Testtheorie ist zum Beispiel die Rede vom *Error*, einer unsystematischen Schwankung, die sich bei häufiger Messung herausmittelt). Bei zukünftigen Untersuchungen wird das an die vergangenen Daten und das darin enthaltene Rauschen angepasste Modell dann notwendigerweise schlechter abschneiden, weil das Rauschen in den neuen Daten ein anderes ist. Man spricht dann von einem „überangepassten“ Modell oder *Overfitting*.

11.0.8 Tendenz von Menschen, sich selbst zu bestätigen (*Confirmation Bias*)

Ein besonderes Problem wissenschaftlicher Methoden ist der Confirmation Bias. Das Phänomen ist in der wissenschaftlichen Literatur nicht klar definiert (Nickerson 1998), hier meine ich damit die Tendenz von Menschen (oder in diesem Kontext: Forschenden), diejenigen Muster zu finden, die sie erwarten. Der Confirmation Bias basiert auf wissenschaftlichen Befunden (Oswald and Grosjean 2004) und wurde von den Wissenschaftler*innen auf sie selbst übertragen (Mynatt, Doherty, and Tweney 1977; Yu et al. 2014). Diese Gedanken führen nah an logischen Unsinnigkeiten und Paradoxa entlang, selbstironisch bemerkt zum Beispiel Nickerson (1998) die Möglichkeit, dass alle Befunde zu Confirmation Biases selbst nur Produkte desselben sein könnten, was die Existenz des Confirmation Biases dann wieder bestätigen würde (S. 211). Praktisch besteht die Gefahr, dass Wissenschaftler*innen nicht Wahrheiten herausfinden, sondern alles so drehen, dass ihre Vorahnungen bestätigt werden. Ludwik Fleck (1935/2015) geht in seiner Wissenschaftssoziologie, die die Grundlage für Thomas Kuhns Arbeit zu wissenschaftlichen Revolutionen (Kuhn 1970/1996) bildet, noch ein paar Schritte weiter: Er argumentiert für ein Modell des wissenschaftlichen Fortschritts, bei dem es nicht darum geht, der Wahrheit näher zu kommen, sondern nach bestem Wissen Probleme vor dem gesellschaftlichen Hintergrund zu verstehen. Das heißt nicht, dass es keine Wahrheit gibt, nur dass Wahrheit eben nicht bloß die Übereinstimmung von Aussagen mit Tatsachen ist. Statt dieser oft von Wissenschaftler*innen vertretenen *Korrespondenztheorie von Wahrheit*, findet sich bei Fleck eine *Konsens Theorie* von Wahrheit wieder: Die Übereinstimmung vieler Leute ist wichtig. Wissenschaftliche Tatsachen werden nicht von einzelnen Personen „entdeckt“, sondern von einem Kollektiv erschaffen. Der Confirmation Bias findet sich dabei so wieder, dass dem Konsens widersprechende Befunde ausgeblendet werden und auf den aktuellen Auffassungen so lange wie möglich beharrt wird. Wenngleich philosophische Wahrheitstheorien den Rahmen dieses Buches sprengen, sei darauf hingewiesen, dass keine der drei Wahrheitstheorien (Korrespondenz, Konsens, und Kohärenz) haltbar ist [Albert (2010); *Münchhausen Trilemma*].

11.0.9 Datenfälschung

Die bisher diskutierten Praktiken werden oft als fragwürdig (questionable) dargestellt. Manche Wissenschaftler*innen halten das für ein Euphemismus, denn in der Verantwortung als Forscher*in sollte genügend Wissen vorliegen, um zu erkennen, dass die oben beschriebenen Techniken *nicht* wissenschaftlich sind und eindeutig nicht der Generierung von Wissen dienen. Sie behindern deutlich den wissenschaftlichen Fortschritt, gefährden das Vertrauen in Wissenschaft, und führen zu enorm hohen Kosten. Unglücklicherweise sind diese Problematiken vielen Wissenschaftler*innen heute immer noch nicht bekannt. „Das haben wir halt so gelernt und schon immer so gemacht“ heißt es zum Beispiel. Dass bestimmte Studien sich nicht replizieren ließen, war teilweise schon vielen Personen bewusst, sie hielten es nur nicht für möglich, das im Scientific Record festzuhalten. Jedenfalls legt der Begriff der fragwürdigen Forschungspraktiken nahe, dass sich Forschende damit in einer Grauzone bewegen würden.

Meiner Ansicht nach, ist das nur der Fall, da, wenn Forschende ihren Job verlieren würden, weil sie P-Hacking betrieben haben, nicht mehr viele Forschende übrig wären.

Anders ist es beim Fälschen und Manipulieren von Daten. Wie häufig Datenmanipulationen oder -fälschungen vorkommen ist ungewiss und Schätzungen sind schwierig. In einer Meta-Analyse von Umfragen zu dem Thema wurde geschätzt, dass zwischen 0,86 und 4,45% aller Wissenschaftler*innen zugaben, Daten manipuliert zu haben. 72% gaben an, fragwürdige Forschungspraktiken anzuwenden (Fanelli 2009). Stroebe, Postmes, and Spears (2012) stellten später Beispiele von Datenfälschung zusammen und empfahlen Peer Review und Replikationen als Betrugs-Detektoren. Eine neuere und extrem umfangreiche Studie von Gopalakrishna, Wicherts, et al. (2021) berichtete, dass 8,3% aller Befragten Daten manipuliert oder gefälscht hätten und 51,3% fragwürdige Forschungspraktiken angewandt hätten (Tabelle 2) und bestätigte den Ausmaß der Probleme. Je nach Disziplinen kommen weitere Probleme hinzu, wie zum Beispiel die Verwendung bereits veröffentlichter biomedizinischer Bilder, die in ungefähr 3,8% aller veröffentlichten Artikel angewandt wurde (Bik, Casadevall, and Fang 2016). Es wird davon ausgegangen, dass Datenfälschung nur in sehr seltenen Fällen aufgedeckt wird. Diejenigen Fälle, die ans Licht kamen, hatten die Zurückziehung (*Retraction*) der jeweiligen wissenschaftlichen Artikel zur Folge und oft Konsequenzen für die wissenschaftliche Karriere der Verantwortlichen. Retractionwatch.org verwaltet die weltweit größte Datenbank zu zurückgezogenen Artikeln (Stand Dezember 2023: 49.628 Artikel): <http://retractiondatabase.org/>.

Sehr düster ist dabei die Tatsache, dass Methoden zur Datenfälschung einerseits immer einfacher werden (Naddaf 2023) und Wissenschaftler*innen, die Fehler aufdecken, gelegentlich verklagt werden. Das betrifft beispielsweise wurden die Autoren von Datacolada.org, die bereits häufiger Probleme aufgezeigt haben, von Francesca Gino für die Veröffentlichung verklagt (<https://datacolada.org/109>), woraufhin tausende Wissenschaftler*innen Gelder für die finanzielle Unterstützung des Gerichtsprozesses sammelten (<https://www.gofundme.com/f/uhbka-support-data-coladas-legal-defense>).

11.1 Weiterführende Informationen

- Bei diesem englischsprachigen Podcast wird diskutiert, wie und ob sich Betrug in der Wissenschaft stoppen lässt: <https://freakonomics.com/podcast/can-academic-fraud-be-stopped/>
- Wie sich die Replikationskrise aus Perspektive eines jungen Forschers angefühlt hat beschreibt Daniel Lakens: <http://daniellakens.blogspot.com/2020/11/why-i-care-about-replication-studies.html>

11.1.1 Literatur

12 Theorien

Wissenschaft arbeitet mit Theorien. Wie diese genau aussehen, unterscheidet sich zwischen Disziplinen deutlich. Während naturwissenschaftliche Bereiche häufig mit mathematischen Modellen, also Formeln, arbeiten, die den Zusammenhang zwischen Variablen explizit und unmissverständlich beschreiben und Vorhersagen erlauben, arbeiten Sozialwissenschaften häufig mit verbalen Theorien im Stile von „X und Y hängen positiv miteinander zusammen“ oder „je höher X, desto höher Y“ und traditionelle Geisteswissenschaften arbeiten beispielsweise mit verbalen Erklärungen. Verbale Theorien haben den Vorteil, dass sie tendenziell leicht verständlich und allgemein anwendbar sind, allerdings unterliegen die verwendeten Begriffe häufig individuellen, kulturellen, oder zeitlichen Einflüssen und Diskutant*innen droht, im wissenschaftlichen Diskurs aneinander vorbei zu reden. Für formale Theorien werden alle beteiligten Variablen genau definiert und die Theorien haben häufig einen stark eingeschränkten Geltungsbereich (z.B. gelten viele physikalische Gesetze nur unter streng kontrollierten Bedingungen wie im Vakuum, bei einer bestimmten Temperatur, usw.). Die Sorge im Rahmen der Replikationskrise ist, dass Theorien nicht klar genug sind, um vorherzusagen, wann Replikationen erfolgreich sind und damit eine der Ursachen für geringe Replikationsraten sind (Buzbas and Devezzer 2023; P. Smaldino 2019). Eine Theorie über die Konsequenzen von der Identifikation mit Geschlechterrollen muss beispielsweise die Veränderung von Geschlechterrollen und Besonderheiten von Geschlechterrollen in verschiedenen Ländern berücksichtigen. Dass ein und dasselbe Experiment zu diesem Thema in den USA im Jahre 1980 andere Ergebnisse hat als in Deutschland im Jahr 2020 ist wenig überraschend. Problematisch ist allerdings, dass – auch wenn solche Ergänzungen für viele sozialwissenschaftliche Theorien sinnvoll und nötig erscheinen – nur selten Aussagen darüber gemacht werden.

Verbale Theorien sind per se nicht weniger wissenschaftlich: Im Kontext der jeweiligen Bereiche heben sich wissenschaftliche Theorien stets durch ihren besonders hohen Grad an Systematizität (Hoyningen-Huene and Kincaid 2023) von alltagswissenschaftlichen Erklärungen ab. Bereiche, die Wert auf Vorhersage von Geschehnissen legen, kommen jedoch nicht ohne formale Theorien aus (Muthukrishna and Henrich 2019). Dabei sei hervorgehoben, dass bestimmte Wissenschaften eben *keinen Wert* auf Vorhersage legen (z.B. Geschichtswissenschaften oder Disziplinen, die vorwiegend hermeneutisch vorgehen). Sozialwissenschaften wie die Psychologie, quantitative Soziologie, oder Teile der Geisteswissenschaften („Digital Humanities“) nähern sich aktuell formalen Modellen an – in der Sozialpsychologie gab es den Aufruf, Theorien zu formalisieren beispielsweise schon einmal bei einer Krise in den 1970er Jahren (Lakens 2023). Dadurch, dass sich Theorien durch ihren Mangel an Objektivität selten von verschiedenen Forschenden verwendet werden und sich durch ihre flexible Auslegung nur schwer wieder-

legen lassen ist dort eine enorm große Menge an nutzlosen Theorien entstanden (Ferguson and Heene 2012). Darunter sind auch einander widersprechende Theorien: Beispielsweise argumentierten Banker et al. (2017), dass „ego depletion“, also die Erschöpfung von Selbstkontrollressourcen, dazu führt, dass Personen sich eher an Hinweise anderer Leute orientieren (S. 2) während Francis et al. (2018) gegenteilig vermuteten, dass die Erschöpfung verhindert, dass Hinweise überhaupt verarbeitet werden können. Beide lieferten Daten, die die jeweiligen Theorien bestätigten, jedoch fand eine Folgeuntersuchung, dass vermutlich beide falsch lagen (Röseler, Schütz, et al. 2020).

Robinaugh et al. (2021) diskutieren Beispiele der Umwandlung verbaler Theorien in formale. Dieser Prozess hat zur Folge, dass sich neue und spezifischere Vorhersagen ableiten lassen. Wenn eine Theorie genauere Vorhersagen macht und die Menge an möglichen Ereignissen, die der Theorie widersprechen, steigt, bedeutet das einen gestiegenen *empirischen Gehalt* (Glöckner and Betsch 2011; Popper 1959/2008).

💡 Empirischer Gehalt und Strong Inference

Theorien können sich in ihrem empirischen Gehalt unterscheiden. Damit ist konkret gemeint, wie spezifisch ihre Vorhersagen sind. Je mehr *mögliche* Beobachtungen eine Theorie widerlegen *würden*, desto höher ist ihr empirische Gehalt.

Nehmen wir den Fall, dass unsere Theorie uns erlaubt, Vorhersagen darüber zu machen, was für ein Auto zu einer bestimmten Zeit an einer bestimmten Straße entlang fährt. In Abbildung sind alle möglichen Autos abgebildet. Zur Vereinfachung gibt es in unserer Beispielwelt nur 9 verschiedene Autos, die sich hinsichtlich der Merkmale *Farbe* (grün, schwarz, blau), *Heckflügel* (mit, ohne), und *Radfarbe* (grau, gelb) unterscheiden.

- Die lila Theorie sagt: *Das beobachtete Auto hat graue Räder*. Ohne Theorie wären für uns alle Autos gleich wahrscheinlich, die lila Theorie „verbietet“, dass das Auto gelbe Räder hat. Sie verbietet 3/9 Autos.
- Die rote Theorie sagt: *Das beobachtete Auto ist blau*. Die Wahrscheinlichkeit, sie zu widerlegen wäre in unserer Musterwelt höher, nämlich 6/9. Weil die rote Theorie *a priori*, also ohne weiteres Vorwissen, sozusagen eine riskantere Wette ist, hat sie höheren empirischen Gehalt.
- Den höchstmöglichen empirischen Gehalt hat die orangene Theorie: *Das beobachtete Auto ist grün, ohne Heckflügel, und hat graue Räder*. Sie verbietet alle außer einen Fall (8/9).

Das Beispiel mit den neun möglichen Auto-Typen ist natürlich stark vereinfacht. In bestimmten Bereichen schaffen es Forschende jedoch gelegentlich, Resultate von Experimenten auf wenige mögliche Ergebnisse herunterzubrechen und damit zwischen Theorien abzuwägen. Platt (1964). nennt das die *Methode der starken Inferenz* (Strong Inference)

und argumentiert, dass Bereiche, in denen so vorgegangen wird, schnellen Fortschritt erleben. In Anlehnung daran fordert P. E. Smaldino (2017), dass wir mehr Theorien bzw. Modelle benötigen und Forschende immer mehrere Erklärungen gleichzeitig anbieten sollten. Das kann den Vorteil bringen, dass Forschende sich nicht auf eine Möglichkeit festlegen und Theorien nicht als Besitztum von jemandem behandelt werden. Solange sich eine Theorie klar einer Person zuordnen lässt, besteht die Gefahr das Kritik an der Theorie mit Kritik an der Person verwechselt wird.

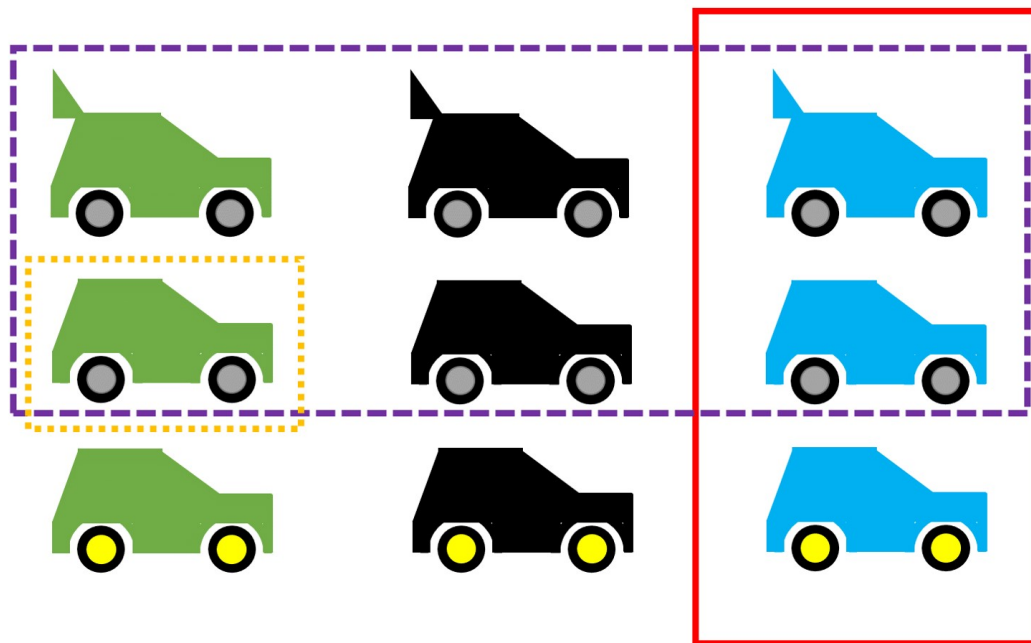


Figure 12.1: Visualisierung von Theorien mit unterschiedlichem empirischem Gehalt

12.0.1 Deduktion und Induktion

Methoden werden reformiert und Wissenschaftler*innen diskutieren, wie Wissenschaft funktioniert, ablaufen sollte, und welche Methoden sinnvoll und unsinnig sind. Wie am hermeneutischen Zirkel klar wird, führt ein Erkenntnisweg darüber, eine Menge von Beobachtungen zu einer Regelmäßigkeit oder Gesetzmäßigkeit zusammenzufassen (*Induktion*) und ein weiterer besteht daraus, aus einer Gesetzmäßigkeit bzw. Theorie Vorhersagen über noch nicht angestellte Beobachtungen zu machen (*Deduktion*). Immer wieder wird diese Unterscheidung im wissenschaftlichen Diskurs vernachlässigt oder ausgeblendet. Beispielsweise drehte sich ein Dialog in der Konsumentenpsychologie jahrelang darum, welcher Weg besser sei, obwohl beide Wege gleichermaßen legitim sind und einander ergänzen (Calder, Phillips, and Tybout 1981). Ähnlich verhält es sich bei Konflikten zwischen qualitativer und quantitativer Vorgehensweise,

die formal betrachtet jeweils eher induktiv oder deduktiv vorgehen (Borgstede and Scholz 2021). Bei Replikationsforschung hat traditionell die induktive Seite mehr Beachtung erfahren (Hüffmeier, Mazei, and Schultze 2016; Yamashita and Neiriz 2024): Jeder Unterschied zwischen Replikation- und Originalstudie wird als mögliche Ursache für ein Scheitern des Replikationsversuches herangezogen um die Vertrauenswürdigkeit der Originalbefunde aufrechtzuerhalten (Baumeister and Vohs 2016). Dabei gerät außer Acht, dass kleinere Unterschiede zwischen Original- und Replikationsstudie (z.B. Verwendung der Maße, durchschnittliches Alter der Versuchspersonen, Sprache der Instruktion) von Theorien nicht erfasst werden – ihnen zufolge also unerheblich sein sollten – und eine fehlgeschlagene Replikation klar die Grenzen der Theorie aufzeigt und sich aus ihr Empfehlungen für die Modifikation von Theorien ableiten lassen (Cesario 2014; Dijksterhuis 2014). Ein Überblick über die Vorgehensweisen ist in der folgenden Tabelle.

Facette	Deduktives Vorgehen (Theorie-geleitet)	Induktives Vorgehen (Phänomen-geleitet)
Verallgemeinerung steckt in...	der Theorie: Sie ist a priori maximal allgemein (z.B. gilt sie, bis anderweitig nachgewiesen, für alle Menschen).	den Daten: Erst vielfältige Beobachtungen in verschiedenen Kontexten erlauben die Annahme, dass das Phänomen allgemeingültig ist.
Veränderung von Verallgemeinerbarkeit	Mit mehr Beobachtungen sinkt die Allgemeingültigkeit.	Mit mehr Beobachtungen steigt die Allgemeingültigkeit (sofern sie bestätigender Natur sind).
Art der Prüfung	Vorhersagen der Theorie werden vorwiegend Versuchen der <i>Widerlegung</i> unterzogen.	Wiederholte Beobachtungen <i>bestätigen</i> den ursprünglichen Einzelfall.
Wahl des Studiensettings	Studentische Stichproben aus nur einem Land oder Laboruntersuchungen sind unbedenklich.	Der Kontext der Untersuchung sollte die Zielbedingungen (z.B. bei der Anwendung der Erkenntnisse in der Praxis) möglichst gut widerspiegeln.

Merkmale induktiver und deduktiver Vorgehensweise, entnommen, übersetzt, und angepasst aus einem unveröffentlichten Manuskript von Röseler & Leder.

12.0.2 Hilfhypothesen

Über folgende Wege lassen sich Replikationsfehlschläge erklären:

1. Fehler erster Art der Originalstudie: Der Originalbefund war nur ein Zufallsbefund oder kam durch wissenschaftliches Fehlverhalten zustande (siehe Kapitel „Freiheitsgrade von Forschenden (Researchers’ Degrees of Freedom)“).

2. Fehler erster Art der Replikationsstudie: Die Originalstudie lag richtig, die Replikationsstudie hat einen Fehler gemacht (z.B. zu kleine Stichprobe, schlechte Kalibrierung der Instrumente, oder wissenschaftliches Fehlverhalten).
3. Grenzbereich des Phänomens: Beide Studien sind vertrauenswürdig. Die Replikationsstudie *unterscheidet* sich auf eine für die Theorie wichtige Weise (z.B. wurde die Replikationsstudie mit Personen aus einem anderen Land durchgeführt und die Theorie gilt nur für Menschen aus dem „Original-Land“).

Variante 3 ist konstruktiv und nimmt beide Einzelbefunde für robust hin. Notwendig dafür ist ein theoretisch relevanter Unterschied zwischen der Original- und Replikationsstudie, der durch die unendliche Anzahl möglicher wichtiger Faktoren in den meisten Fällen zutrifft (Smedslund 2015). Über diesen Weg lässt sich die Theorie dann modifizieren oder eine weitere Theorie aufstellen, die für den Kontext der Untersuchung ebenfalls berücksichtigt werden muss. Schwierig wird es, wenn Forschende nach bestem Wissen eine Replikation durchführen, diese „fehlschlägt“ (also nicht das nachgewiesen wird, was nachgewiesen werden sollte), und andere Forschende die Replikation dafür kritisieren, dass sie etwas „falsch“ gemacht hat. Nachdem Hagger et al. (2016) unter Absprache mit Roy Baumeister dessen Ego Depletion Theorie mit einer großangelegten Studie prüften, kritisierten Baumeister and Vohs (2016), dass von Anfang an zu erwarten gewesen wäre, dass die Studie nicht funktioniert und bezeichnete die Studie als fehlgeleitet. Vohs, die Ko-Autorin der Kritik war, führte einige Jahre später eine weitere groß angelegte Replikationsstudie durch. Obwohl sie dieses Mal ihrem eigenen Rat folgen konnten, konnten die Forschenden wieder nicht den erwarteten Effekt finden (Vohs et al. 2021).

12.1 Weiterführende Informationen

- Eine philosophie Perspektive auf den Zusammenhang zwischen Theorie, Messungen, und Replikationen diskutiert @ramminger2023vermessen
- Yarkoni (2019) argumentiert, dass das Replikationsprobleme in der Verallgemeinerung von Ergebnissen zu Theorien ihren Ursprung haben.
- Fanelli diskutiert in einem Vortrag die Komplexität von Forschung als Grund für Replikationsfehlschläge und schlägt eine Theorie zur Messung von Komplexität vor (Fanelli, Tan, et al. 2022), Ein Video zu einem Vortrag ist online verfügbar: <https://www.youtube.com/watch?v=CEAV7420jBk>

12.1.1 Literatur

13 Epistemische Probleme

Epistemologie heißt Erkenntnislehre und ist ein Teilgebiet der Philosophie. Darin wird diskutiert, wie Wissenschaft funktioniert, wie Wissen produziert werden kann, oder worin der Unterschied zwischen wissen und glauben liegt. Epistemologische Erklärungsansätze für Replikationsfehlschläge übersteigen systemische und methodische Faktoren, haben aber auch andere Prüfbarkeitsansprüche. Während wissenschaftstheoretische Ansätze teilweise empirisch (also durch Beobachtungen) prüfbare Vorhersagen erlauben, lässt sich über die philosophischen Probleme nur nachdenken und diskutieren.

13.0.1 Robustheit und Historizität von Phänomenen

Unter welchen Voraussetzungen ist es wenig überraschend, dass Replikationsversuche fehlschlagen? Ein Ausweg ist anzunehmen, dass die untersuchten Phänomene extrem empfindlich oder unstabil seien. Regelmäßigkeiten im menschlichen Verhalten analog zu den Planetenbewegungen zu entdecken könnte schlichtweg nicht möglich sein (Smedslund 2015). Weniger extreme Annahmen über die Existenz von Regelmäßigkeiten, die möglicherweise nicht jede Person ausnahmslos betreffen aber „im Schnitt“ gelten. Lewin (1930) unterscheidet in aristotelische und galileische Gesetzmäßigkeiten, wobei ersteres die *aristotelischen* sind, die in den Sozialwissenschaften relativ unumstritten sind. Eine Regelmäßigkeit bzw. ein Naturgesetz im aristotelischen Sinne kann zum Beispiel sein, dass Männer größer als Frauen sind. Noch extremer ist die Theorie, dass Menschen sich des Wissens über sie bewusst sind und ihr Verhalten dynamisch anpassen und Verhaltenswissenschaften immer historisch bzw. zeitgebunden sind: Wird herausgefunden, dass Menschen in ihren Entscheidungen tendenziell dazu neigen nichts zu ändern, auch wenn sich dadurch ihre Situation verbessern würde, wird ihnen diese Tatsache über die Wissenschaft vor Augen geführt und sie können ihr Verhalten anpassen. Dabei handelt es sich übrigens um den Status Quo Bias, bei dem Menschen die jetzige Situation einer anderen vorziehen (Samuelson and Zeckhauser 1988; Xiao et al. 2021).

Wie stark sich Phänomene durch vermeintlich kleinere Unterschiede im Versuchsaufbau unterscheiden wurde bereits meta-wissenschaftlich untersucht. Landy et al. (2020) ließen mehrere Hypothesen von mehreren Forschenden prüfen und Faktoren, die laut den dahinterliegenden Theorien eigentlich keinen Unterschied machen sollten, führten dazu, dass Gegenteilige Ergebnisse entstanden. Auf Replikationsforschung übertragen ist es also möglich, dass in bestimmten Forschungsbereichen völlig unklar ist, unter welchen Bedingungen welche Zusammenhänge zu beobachten sind.

13.1 Weiterführende Informationen

- Fleck (1935/2015) schlägt eine Theorie des wissenschaftlichen Fortschritts vor, bei der Wissen immer einem sogenannten Denkstil unterliegt, der für die jeweilige gesellschaftliche Situation optimal ist. Dabei bedient er sich Elementen der Evolutionstheorie, Soziologie, und Psychologie.
- Shiffrin, Börner, and Stigler (2018) diskutiert das scheinbare Paradox zwischen Fallibilität und Fortschritt in der Wissenschaft.

13.1.1 Literatur

Part VI

Lösungen

Part VII

Lösungen und Ansätze zur Verbesserung der Lage der Psychologie

Werfen wir ein Blick darauf, was sich seit 2012 in der Psychologie verändert hat. Eingeteilt sind die Veränderungen dahingehend, welchen Teil des Problems sie vor allem betreffen: Ist der Zweck einer neuen Vorgabe, das System, die Methodik, oder die Theorien der Forschung zu verbessern? Einige Lösungen sind mit vielen Problemen gleichzeitig verknüpft und manche sind auf bestimmte Angelegenheiten maßgeschneidert. Die Abbildung enthält eine grobe Unterteilung.

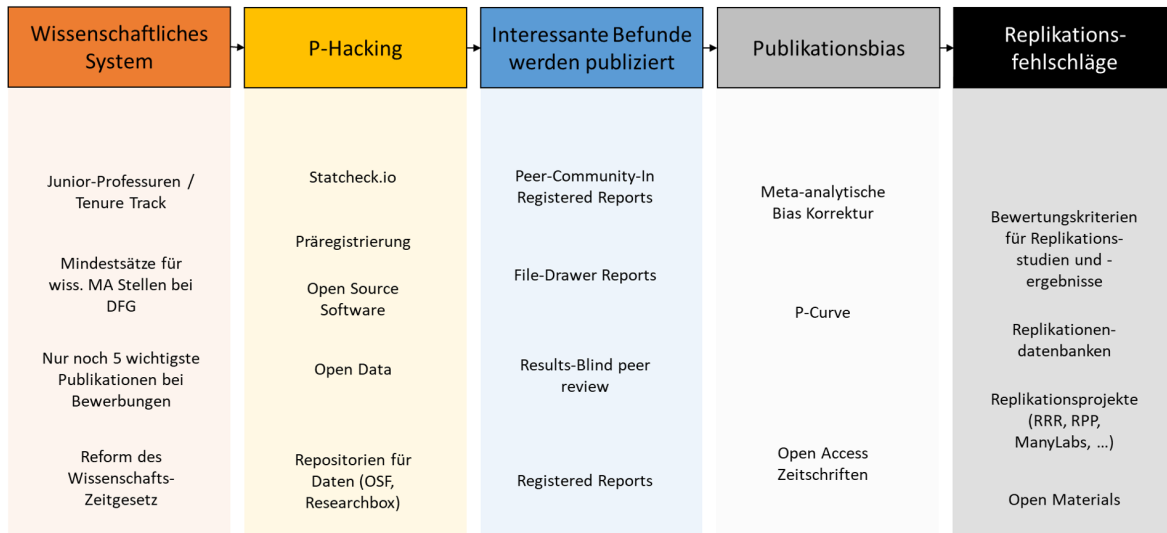


Figure 13.1: Lösungsansätze und welches Problem in der Wissenschaft sie aufgreifen

14 Das System

Fangen wir mit dem System an. Ansätze, die dort greifen bergen am meisten Potenzial, denn um im System Geld verdienen zu können, müssen Forschende sich an die Regeln halten. Und solange Publikationen die Währung sind und Paper mit knackigen Titeln und eindeutigen Ergebnissen als qualitativ hochwertiger befunden werden, sind Forschende darin motiviert, statt der Wahrheit eben nach knackigen Titeln und eindeutigen Ergebnissen zu suchen.

Allgemein ist eine positive Entwicklung sichtbar (Korbmacher et al. 2023) und eine Veränderung der Anreizstruktur wird anvisiert. Sie lässt sich als Angleichung des wissenschaftlichen Systems an die *Mertonschen Normen* (nach Robert Merton) auffassen (Merton 1973): (1) Kommunismus: Das wissenschaftliche Wissen sollte allen Wissenschaftler*innen gleichermaßen gehören, um die Zusammenarbeit zu fördern. (2) Universalismus: Wissenschaftliche Güte ist unabhängig vom soziopolitischen Status und persönlichen Attributen der Teilhabenden. (3) Desinteresse: Wissenschaftliche Institutionen handeln im Interesse der Wissenschaft und nicht für persönlichen Gewinn. (4) Organisierter Skeptizismus: Wissenschaftliche Behauptungen sollten einer kritischen Prüfung unterzogen werden bevor sie akzeptiert werden.

Nosek empfiehlt in einem [Blogpost](#) eine Maßnahmenstruktur, nach welcher die gewünschten Veränderung nacheinander ...

1. möglich (z.B. durch Infrastruktur wie online Repositorien, in denen Forschungsmaterialien öffentlich und gratis hochgeladen werden können),
2. einfach (z.B. durch barrierearme Angebote, mehrsprachige Anleitungen),
3. normativ (z.B. durch Wissenschaftliche Communities, die gemeinsam hinter Forderungen der Verbesserung stehen),
4. belohnend (z.B. durch designierte Preise), und
5. notwendig (z.B. durch Mindeststandards, die von Zeitschriften oder Drittmittelgebern gefordert werden)

gemacht werden sollen. Wie die verschiedenen Ansätze bei den verschiedenen Akteuren, also Politik, Universitäten, oder Zeitschriften konkret aussehen, wird im Folgenden diskutiert.

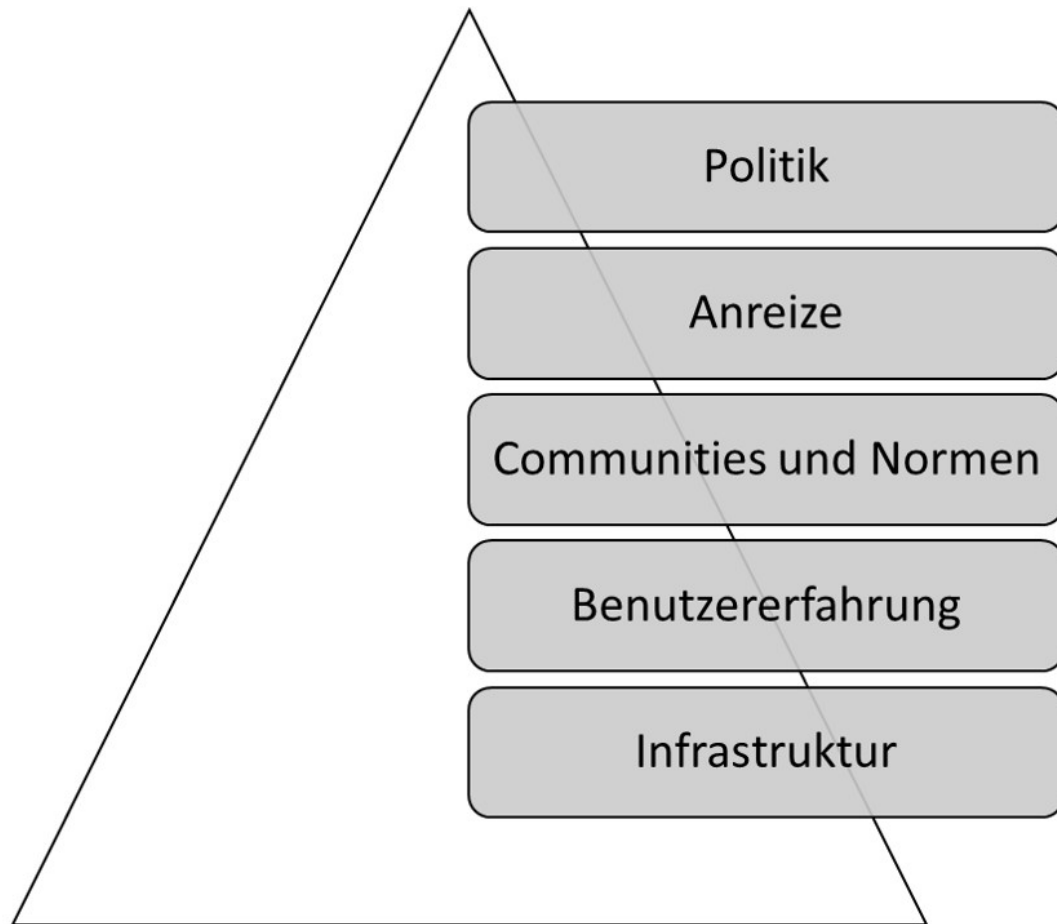


Figure 14.1: Kulturwandel in der Wissenschaft nach Nosek (<https://www.cos.io/blog/strategy-for-culture-change>)

14.0.1 Politik

International stehen politische Parteien und Vereinigungen deutlich hinter Open Science und Open Access. Beispielsweise empfiehlt die UNESCO einen universellen Zugang zu wissenschaftlichen Wissen ungeachtet von Herkunftsland, Geschlechterrolle, politischen Grenzen, Ethnizität, oder ökonomischen oder technologischen Hürden [UNESCO (2020); p. 3]. [Arbeitsgruppen](#) für politische Instrumente, Förderung, und Infrastruktur wurden entsprechend gegründet. Die [G7](#) setzen sich für wissenschaftliche Integrität, akademische Freiheit, und Open Science ein. [Offener Zugang](#) aber auch [Transparenz](#) des wissenschaftlichen Vorgehens wird auch seitens der Europäischen Union gefordert. Infrastruktur (z.B. die [European Open Science Cloud](#)) und diverse Open Science Forschungsprojekte werden gezielt gefördert. Für alle durch die EU geförderten Forschungsprojekte steht zudem die kostenlose Publikations- und Begutachtungs-Plattform [Open Research Europe](#) zur Verfügung.

In Deutschland hat sich die Regierung der Periode 2021-2025 im Rahmen des Koalitionsvertrages vorgenommen, „Open Access ... als gemeinsamen Standard [zu] etablieren.“ Einzelne Bundesländer wie Nordrhein-Westfalen haben in Zusammenschlüssen aus den jeweiligen Universitäten darüber hinaus Open Access Strategien entwickelt (Openness 2023) und arbeiten aktuell an Open Science Strategien. Andere Länder, wie zum Beispiel Schweden, haben bereits [nationale Richtlinien zu Open Science](#) entwickelt. Hinsichtlich der Problematik von Machtmissbrauch wird das Problem beispielsweise in einem Eckpunkte-Papier des Landes NRW anerkannt, doch als [Einzelfall](#)- statt System-Problem verstanden ([Kommentar dazu](#)).

14.0.2 Universitäten[LR9]

i Welche Universitäten tun etwas? (Beispiele)

Das Thema Open Science hat bei vielen Universitäten bereits Anklang gefunden. Während die meisten deutschen Universitäten (zeitlich begrenzte) Mittel für Open Access Publikationen haben, existieren an einigen darüber hinaus Open Science Policies (z.B. [FAU Erlangen](#)), Open Science Centers (z.B. [LMU Open Science Center](#); [Köln Open Science Center](#); [Münster Center for Open Science](#); [Mannheim Open Science Office](#)). Darüber hinaus unterstützen das Leibniz-Informationszentrum Wirtschaft in Kiel und das Leibniz-Institut für Psychologie Replikationsforschung beispielsweise mit einer Replikationszeitschrift (<https://www.jcr-econ.org>) oder im Rahmen einer Juniorprofessur für Psychologische Metawissenschaft. Eines der größten Zentren für Metawissenschaft innerhalb Europas hat sich in den Niederlanden in Tilburg gebildet. Die Berliner Universitäten haben eine gemeinsame Open Access Erklärung entwickelt (<https://openaccess.mpg.de/Berliner-Erklaerung>) und in Frankreich ist die Universität Sorbonne ein Pionier: Gemeinsam mit der Universität Amsterdam und dem Universitätscollege London wurde eine Erklärung über die Veröffentlichung von Forschungsdaten unterzeichnet. Seit 2024 hat die Universität Sorbonne darüber hinaus den Vertrag mit Clarivate für die Nutzung der Forschungsdatenbank „Web of Science“ gekündigt und arbeitet seitdem mit der Open Source Software „OpenAlex“ (Priem, Piowar, and Orr 2022).

Für die langfristige Entwicklung der Wissenschaften haben Universitäten dadurch eine besondere Verantwortung, dass sie Wissenschaftler*innen beschäftigen und an ihnen die Auswahl für die wenigen unbefristeten Arbeitsplätze in der Wissenschaft fallen. Wenn jahrzehntelang Professuren auf Basis subjektiver, nicht-reproduzierbarer, und für gute Wissenschaft nachrangigen Kriterien gewählt werden (z.B. Anzahl an Publikationen in Fachzeitschriften), kann sich das negativ auf die Entwicklung von Wissenschaften auswirken. Diesem Problem entgegenwirkend wurde ein Forschungspreis des Berlin Institute of Health (BIH), der jährlich für Projekte zur Förderung von wissenschaftlicher Integrität verliehen wird, an ein Projekt, das objektive und sinnvolle Auswahlkriterien für Professor*innen entwickelt vergeben (Schönbrodt, Gärtner, et al. 2022; Gärtner, Leising, and Schönbrodt 2022). Innerhalb von Universitäten spie-

len außerdem die Bibliotheken eine aufklärerische Rolle hinsichtlich Forschungsdatenmanagement und Publikationskultur (Schmidt et al. 2024). Über sie kann der Forschungsprozess mit entsprechender Infrastruktur (z.B. zum Lagern und Veröffentlichen von Forschungsmaterialien und -ergebnissen) unterstützt werden (Quan 2021) und verhindert werden, dass sich eine „Abhängigkeit von wenigen kommerziellen Anbietern“ ergibt, die „begrenzen, was [bei] der Forschung an Arbeitsmöglichkeiten und Fragestellungen erreichbar ist“ (Siems 2024). Eine Pflicht, Mitglieder einer Universität zur Einhaltung von Open Science Strategien zu bewegen, gibt es an Universitäten durch den hohen Stellenwert der „Freiheit der Forschung“ kaum. Damit könnten sie Gefahr laufen, für Forschende weniger attraktiv zu werden: Wenn beispielsweise auf namhafte Zeitschriften nicht mehr über die Universität zugegriffen werden kann, weil Verträge mit closed-access Zeitschriften gekündigt wurden, erschwert das den Forschenden die Arbeit. Beispielsweise klagte die juristische Fakultät der Universität Konstanz gegen eine Zweitveröffentlichungspflicht: Forschende in Deutschland haben das [Recht zur Zweitveröffentlichung](#), das heißt, dass sie ihre Forschung, wenn sie in einer Fachzeitschrift veröffentlicht wurde, auch selbst (z.B. über eigene Websites) veröffentlichen dürfen. In Konstanz konnten die Forschenden nicht dazu verpflichtet werden, davon Gebrauch zu machen.

Ein weiteres Stellrad von Universitäten ist die Bezuschussung von Publikationskosten bei Zeitschriften. Wird beispielsweise die wissenschaftliche Qualität einer Zeitschrift angezweifelt, kann eine Universität diese [Bezuschussung stoppen](#).

Eine oft vernachlässigte Rolle kommt außerdem der universitären Lehre hinzu. Durch die Freiheit von Forschung und Lehre und bereits durchgeplanten Studiengängen gestaltet sich die Integration neuer Themen wie Open Science schwierig. Forschende, in deren Lehre die Thematik eine Rolle spielt, teilen proaktiv ihre Materialien, erstellen gemeinsam Curricula, und sind beispielsweise in großen internationalen Initiativen wie dem Framework for Open and Reproducible Research Training ([FORRT.org](#)) vernetzt (konkrete Vorschläge zur Integration von Open Science in die Lehre hat zum Beispiel Pennington and Pownall (2024) veröffentlicht).

14.0.3 Institute und Vereinigungen

Wissenschaftliche Gebiete leben vor allem durch Communities, also alle in dem Bereich forschenden Personen. Sie organisieren sich in Vereinen (z.B. Deutsche Gesellschaft für Psychologie), Interessensverbunden, oder ähnlichen Gemeinschaften. Eine besondere Stellung hat in Deutschland die Deutsche Forschungsgemeinschaft, welche staatlich und über die Bundesländer mit mehreren Milliarden Euro ausgestattet Forschungsgelder vergibt. Als eine der wichtigsten nationalen Institution hat ihre Open Science Positionierung einen hohen Stellenwert (Deutsche Forschungsgemeinschaft 2022), erfahrungsgemäß gehen Veränderungen jedoch nicht von der DFG aus, sondern die DFG wartet auf Anstöße aus den Fächern. In der Psychologie fördert darüber hinaus das [ZPID](#) die Infrastruktur durch Zeitschriften, Pre-Print Server, und weitere Methoden und in den Wirtschaftswissenschaften verwaltet das ZBW wichtige Informationen oder Zeitschriften (REF). Auch interdisziplinäre Vereinigungen

wie das [CERN](#) oder internationale Akteure wie die [American Psychological Association](#) verpflichten sich zu Offenheit und Transparenz.

Im Rahmen der Open Science Reform entstanden außerdem viele neue Vereinigungen. Das interdisziplinäre und besonders von Wissenschaftler*innen in der frühen Karrierephase geleitete FORRT (Azevedo et al. 2019) setzt sich für eine Verankerung von Open Science in der Lehre ein. Der Verbesserung psychologischer Forschung hat sich die Society for the Improvement of Psychological Science (SIPS) verschrieben. Sogenannte „grassroot“-Initiativen (also von jungen Wissenschaftler*innen ausgehende Bewegungen) haben sich an zahlreichen Universitäten herausgebildet und zu Netzwerken wie dem [Netzwerk der Open Science Initiativen \(NOSI\)](#) und „Reproducibility Networks“ wie dem deutschen Netzwerk *GRN* (<https://reproducibilitynetwork.de>), dem im Vereinigten Königreich *UKRN* (<https://www.ukrn.org>) und weiteren zusammengeschlossen. Aber auch Zusammenschlüsse von Professor*innen zur Änderung von Kurzzeitverträgen existieren (z.B. Netzwerk Nachhaltige Wissenschaft, <https://netzwerk-nachhaltige-wissenschaft.de>).

14.0.4 Zeitschriften

Wissenschaftliche Zeitschriften gelten als Bühne des wissenschaftlichen Diskurses und bestimmen maßgeblich, welche Elemente des Forschungsprozesses zum „scientific record“ gehören und damit relevant sind. Sie sind darüber hinaus als Organisatorinnen des Begutachtungsprozesses für die Qualitätssicherung in der Wissenschaft verantwortlich. Dem Mangel an Qualität entgegen existieren bereits ausführliche Empfehlungen zur Gestaltung von Zeitschriften, es bilden sich neue Zeitschriften, und vollständige neue Begutachtungs- und Publikationsmodelle werden vorgeschlagen und vielseitig implementiert. Das Journal of [Open Source Software](#) basiert beispielsweise auf einem öffentlich einsehbaren Programmiercode und seine Infrastruktur lässt sich für weitere Zeitschriften einfach kopieren und anpassen.

14.0.4.1 Empfehlungen

Herausgeber*innen, die sich in Bezug auf die von ihnen verwaltete Zeitschrift mit Open Science Praktiken auseinandersetzen möchten, können inzwischen auf einen umfangreichen Leitfaden zurückgreifen (Silverstein et al. 2023). Über eine Diskussionsplattform (Journal Editors Discussion Interface, JEDI) wurden Vorschläge gesammelt und es wird erklärt, worum es sich bei Dingen wie Registered Reports, Open Peer Review, Diversifizierung, und Open Access handelt und wie diese in eine Zeitschrift implementiert werden können. Eventuelle Sorgen und Ängste werden angesprochen und beantwortet. Das Committee on Publication Ethics (COPE) setzt sich ebenfalls für Aufklärung und Lehre ein, die Herausgeber, Universitäten, und Forschungsinstitute im Umgang mit Problemen im Publikationssystem helfen soll. Es bietet beispielsweise [Richtlinien](#) unter welchen Umständen Publikationen zurückgezogen oder korrigiert werden sollten, oder welche ethischen Standards ein [Begutachtungsprozess](#) erfüllen

sollte. Herausgeber*innen, die Zeitschriften für kommerzielle Verlage verwalten und auf Systeme umsteigen möchten, die vollständig in der Hand der Forschenden liegen, können über Universitätsbibliotheken Hilfe bei der Migration von den kommerziellen zu offenen und kostenfreien Systemen erhalten und Zeitschriften beispielsweise mit dem Open Journal System verwalten (siehe z.B. [OJS Netzwerk](#)). Eine Datenbank mit bereits über 20,000 offen zugänglichen Zeitschriften verwaltet das Directory of Open Access Journals ([DOAJ](#)). Gutachter*innen von Forschungsartikeln können über die Reviewer Zero Initiative (<https://www.reviewerzero.net>) auf Lehrmaterialien und Leitfäden zugreifen (<https://osf.io/e7z5k/wiki/Resources/>).

14.0.4.2 Open Science Praktiken hervorheben

Aus der Psychologie ist lange bekannt, dass Motivation über Belohnung besser funktioniert als über Bestrafung (REF). Im Straßenverkehr, in dem es bis vor einiger Zeit nur Bestrafungen für Verletzen von Regeln gab, äußert sich die Erkenntnis beispielsweise an Geschwindigkeitstafeln, die den Autofahrer*innen beim Einhalten des Tempolimits einen fröhlichen Smiley zurückmelden. Wissenschaftliche Zeitschriften heben Einhaltung von Empfehlungen (z.B. öffentlich verfügbare Datensätze) mit Plaketten (*Badges*) hervor. Die Zeitschrift *Psychological Science* geht dabei seit 2024 so weit, dass Badges schon wieder abgeschafft werden und Open Science dort der Standard für alle Artikel ist. Auf der Website topfactor.org sind Zeitschriften und deren Einhaltung verschiedener Standards aufgelistet und in einer Rangliste abgebildet. Zuletzt besteht in jedem System mit Belohnungen das Problem, dass die Akteure ihr Verhalten auf die Belohnungen hin ausrichten und dabei versuchen, Abkürzungen zu gehen (Klonsky 2024).

14.0.4.3 Review Systeme

Wissenschaft zeichnet sich durch Systematik aus (Hoyningen-Huene 2013). Der wohl systematischste Weg, die wissenschaftliche Qualitätssicherung zu garantieren, wäre eine Studie, die verschiedene Systeme vergleicht. Während aktuelle Forschung ähnliches unternimmt (Soderberg et al. 2021), werden alternative Begutachtungssysteme aktuell ausprobiert. Zur Erinnerung: Wissenschaftler*innen verfassen Artikel, die sie bei Zeitschriften einreichen. Dort ist eine Person (Editor) dafür zuständig, dass der Artikel, sofern er zur Zeitschrift passt, an Gutachtende gesendet wird.

Um zu prüfen, ob die Urteile im Begutachtungsprozess zwischen den Urteilenden übereinstimmen, haben Etzel et al. (2024) verschiedene Gutachtende hinsichtlich klassischer Kriterien befragt. Sie fanden heraus, dass das nicht der Fall ist und schlagen Kriterien vor, die einen klaren Wert haben, und sich gut erfassen lassen (z.B. ob Daten öffentlich verfügbar sind).

14.0.4.3.1 Open Peer Review

Seit der Auseinandersetzung Forschender mit dem Begutachtungssystem wird darüber hinaus diskutiert, was mit Gutachten passiert: Traditionell bleiben sie unter Verschluss. In der Fachzeitschrift erscheint der finale Artikel und alle vorherigen Versionen sind nur Autorinnen, Herausgeberin, und Gutachter*innen bekannt. Diese bleiben zudem meistens anonym, das heißt, wenn sie sich keine Mühe gegeben haben, wird es wahrscheinlich nie auffallen. Außerdem haben Forschende keinen Anreiz, Gutachten zu verfassen - höchstens erhalten sie einen Nachweis, dass sie ein Manuskript bei einer Zeitschrift begutachtet haben. Einige Zeitschriften haben inzwischen das *Open Peer Review* eingeführt. Der Name passt nur halb, denn veröffentlicht werden Gutachten nur dann, wenn der Artikel bei der Zeitschrift akzeptiert wird. Das birgt die Gefahr, dass schwerwiegende Probleme, welche für eine Ablehnung üblicherweise nötig sind, nicht ans Licht kommen. Forschende können den Artikel dann ohne Überarbeitung bei einer anderen Zeitschrift einreichen, wo die Probleme vielleicht nicht entdeckt werden. Dieses Vorgehen führt zu Doppelarbeit und enormen Kosten (Aczel, Szaszi, and Holcombe 2021). Zoltan Kekecs merkte dazu bei einer Diskussion im Rahmen einer Konferenz an, dass selbst Kasinos, die in Konkurrenz zueinander stehen, Listen von Betrügerinnen *miteinander austauschen*. Herausgeberinnen von Zeitschriften tun das noch nicht.

Gutachten bleiben unter Verschluss	Gutachten werden bei Publikation veröffentlicht	Gutachten werden bei Publikation und Ablehnung veröffentlicht
Es ist kein Nachweis der Qualitätskontrolle möglich.	Abgelehnte Artikel können bei anderen Zeitschriften ohne Überarbeitung eingereicht werden und führen zu Mehraufwand.	Qualitätskontrolle ist nachvollziehbar und transparent.

Umstrittener als die Veröffentlichung von Gutachten ist die Anonymität der Gutachterinnen: *Der Standard ist, dass Gutachten anonym sind, aber bei Wunsch unterzeichnet werden können. Promovierende, die einen Artikel eines potenziellen zukünftigen Chefs oder einer zukünftigen Chefin schlecht beurteilen, werden dadurch geschützt. Selbst Professorinnen können bei negativen Beurteilungen riskieren, dass der oder die Kollegin später einen Forschungsgeldantrag von ihnen begutachtet und sich für die Kritik rächt. Auf der anderen Seite kann Anonymität zur Folge haben, dass Kritik gegen die Person gerichtet ist und nicht konstruktiv ist.*

Eine weitere Art, wie Peer-Review offen sein kann, ist, dass sich jede Person daran beteiligen kann. Das ist zum Beispiel bei Meta-Psychology möglich. Problematisch ist dabei jedoch die relativ geringe Beteiligung. Bei den Zeitschriften Zeitschrift Synlett und ASAPbio werden Gruppe dazu koordiniert, wie es sie zur Diskussion spannender Artikel schon in Form von *Journal Clubs* gibt - nur eben als "[Pre-Print Review Club](#)".

14.0.4.3.2 Begutachtung von Pre-Prints

Was ist ein Pre-Print?

Ein Pre-Print nennt man ein Manuskript *in dem* oder *vor dem* Stadium der Einreichung bei einer Zeitschrift. Es wurde möglicherweise noch nicht begutachtet oder nach Begutachtung abgelehnt, auf einer Internetseite veröffentlicht, ist zitierbar, und kostenlos zugänglich. In manchen Fällen mag es sinnvoll sein, ein wissenschaftlichen Beitrag *nicht* der Begutachtung zu unterziehen (z.B. bei Kommentaren, Positions-Artikel, oder öffentlichem Austausch). Typischerweise zählen begutachtete Beiträge für die wissenschaftliche Karriere mehr. Der Zweck von Pre-Prints ist vielfältig: Sie erhöhen die Verfügbarkeit von Wissen, erlauben eine schnellere Veröffentlichung (Beweise in der Mathematik müssen aufwändig im Laufe von bis zu mehreren Jahren nachgeprüft werden), und verhindern, dass einem andere Forschende mit einer innovativen Idee zuvorkommen. In der Medizin und den Sozialwissenschaften wurden sie aufgrund des schnelleren Austausches im Zuge der Corona-Pandemie ausgiebig verwendet (Fraser et al. 2020). In der Epidemiologie konnte nicht nachgewiesen werden, dass Pre-Prints qualitativ schlechter sind (L. Nelson et al. 2022).

Über Plattformen und Gemeinschaften wie *PCI* oder *f1000research* werden Pre-Prints begutachtet. Sie werden dann nicht bei einer Zeitschrift sondern bei dem der jeweiligen Einrichtung eingereicht und dort begutachtet. Die Qualitätssicherung wird also von Forschenden selbst organisiert und ist unabhängig von kommerziellen Verlagen. Das Modell bei PCI ist, dass Artikel, die dort ein positives Gutachten erhalten haben, ohne weiteres Gutachten bei einer der teilnehmenden (*PCI-friendly*) Zeitschriften veröffentlicht werden können. Auf F1000research.com fungiert wie eine Zeitschrift, bei der Artikel direkt zugänglich sind, sich über die Zeit durch Peer Review verändern. Auf ähnliche Weise gibt es bei der Zeitschrift für digitale Geisteswissenschaften eine Peer-Review-Ampel: Nach Einreichung sind Artikel dort direkt verfügbar und eine Ampel gibt an, ob sie unter Begutachtung sind, und falls sie begutachtet wurden, ob es (noch) schwerwiegende Probleme gibt oder nicht.

14.0.4.3.3 Gedächtnis von Gutachten

Eine weitere Möglichkeit, Gutachten fest an Forschungsartikel “dranzuheften”, besteht über Plattformen, die eine schnelle Kommentierung ermöglichen. Zum Beispiel lassen sich über Pubpeer.com oder hypothes.is alle wissenschaftlichen Beiträge (z.B. auch Daten) öffentlich und wenn gewünscht anonym kommentieren. Das ist für Pre-Prints möglich, sodass diese Kommentare dann dauerhaft damit verknüpft sind. Mittels Plug-Ins für Internetbrowser werden dann Artikel, zu denen es bei Pubpeer Diskussionen gibt, markiert. Weitere Plattformen sind alphaxiv.org, Disqus, oder scirev.org.

14.0.4.4 Aufmerksamkeit zum Thema in bestehenden Zeitschriften

Anforderungen an wissenschaftliche Artikel seitens der Zeitschriften sind 2010 maßgeblichen Änderungen untergangen. In den Sozialwissenschaften orientieren sich zahlreiche Zeitschriften an den Richtlinien zur „Transparency and Openness Promotion“ (TOP) und erhalten entsprechende TOP-Faktoren (<https://topfactor.org/summary>). Dabei wird festgehalten, welcher Grad an Offenheit und Transparenz von Forschungsdaten und -materialien gefordert wird und ob Replikationen bei der jeweiligen Zeitschrift veröffentlicht werden. Neue Herausgeber*innen bei existierenden Zeitschriften haben große Änderungen vorgenommen. Beispielsweise haben Hardwicke and Vazire (2023) für die Zeitschrift *Psychological Science* ein standardmäßiges Nachrechnen aller berichteten Ergebnisse ab 2024 angekündigt, indem sie mit dem Institute for Replication zusammenarbeiten (<https://i4replication.org>). Vereinzelt haben Zeitschriften Spezialausgaben herausgegeben, bei denen der Fokus auf Replikationsstudien oder der Reproduzierbarkeit von Ergebnissen lag (Carriquiry, Daniels, and Reid 2023).

14.0.4.5 Zeitschriften für “nicht Innovatives”

Durch die Selektion spannender Ergebnisse gibt es für Forschung, die nicht bahnbrechend und dennoch höchst relevant ist, keine Plattform. Schätzungen zufolge werden bis zu 40% aller durchgeführten Studien innerhalb von 4 Jahren nach Durchführung nicht veröffentlicht (Ensinn and Lakens 2023). Andere Forschende können nicht davon lernen und Ressourcen, die in die Sammlung und Auswertung der Daten geflossen sind, Zeit von Versuchspersonen, und lange Vorbereitungen der Forschung werden schließlich verschwendet. Zur Lösung dieses Problems haben sich neue Zeitschriften und Formate gebildet. In der Ökonomie hat sich auf Forderungen (Zimmermann 2015) hin beispielsweise eine Zeitschrift für Replikationen und Kommentare gebildet. Die Zeitschrift Meta-Psychology bietet ein Format für Replikationen und eines für „Schubladenberichte“ an. Letzteres ist für Studien vorgesehen, die wegen wenig überraschenden Ergebnissen oder Fehlern in der Durchführung anderweitig in der Schublade landen würden aber dennoch wichtige Informationen beinhalten. Ebenfalls zur Abbildung des für den Forschungsprozess typischen Fehlschlagens wurde das *Journal of Trial and Error* gegründet, und bei *ReScience C* und *Rescience X* können Berichte über Reproduzierbarkeit und Replikationen veröffentlicht werden.

14.0.4.6 Publikationsmodelle

Noch radikalere Vorschläge als die Anpassung bisheriger Zeitschriften ist der Vorschlag, das bisherige System durch ein neues zu ersetzen. Dabei handelt es sich um ein soziales Dilemma, bei dem Millionen von Forschenden sich plötzlich anders verhalten müssen und dabei entgegen der Spielregeln des wissenschaftlichen Systems handeln müssen (Brembs et al. 2023). Das Dilemma wurde von kommerziellen Verlagen gestaltet, welche daraus Geld verdienen. Brembs et al. (2023) haben einen präzisen Vorschlag erarbeitet, bei welchem

ein dezentrales System aufgebaut wird, das soziale Netzwerke wie Mastodon als Vorbild hat, von Wissenschaftler*innen organisiert wird, und Forschungsprodukte wie Daten oder Programme ebenso wie die traditionellen Forschungsberichte wertschätzt. Plattformen, die ein solches Mikro-Publishing-System bereits implementieren sind [Research-Equals](#) oder Octopus.ac ([Hsing?](#)). Dem System, bei dem alle Produkte begutachtet werden, der Begutachtungsprozess aber nicht die Qualität sicherstellt, stehen hier Ampel-Systeme und öffentliche Kommentierung entgegen, die signalisieren, was und ob begutachtet wurde, welche Kritikpunkte vorlagen, und wie damit umgegangen wurde.

14.0.5 Forschende

Unabhängig von Nationalität, Wissenschaftsgebiet, und Universität haben viele Forschende ihre Arbeitsweisen im Zuge von Open Science überdacht und angepasst. [Hunderte](#) haben öffentliche Erklärungen zur Forschungstransparenz und der Forderung von Open Science Praktiken in der Rolle von [Gutachter*innen](#) unterzeichnet. Einzelne Forschende führen im Rahmen von Lehre Replikationsstudien durch (Boyce, Mathur, and Frank 2023; Jekel et al. 2020; Korell, Reinecke, and Lott 2023), schließen sich weltweit zusammen um gemeinsam Projekte durchzuführen, die einzelne nicht stämmen könnten (z.B. Psychological Science Accelerator, <https://psysciacc.org>), und entwickeln [Informations-Sammlungen](#) (Open Scholarship Knowledge Base, <https://oercommons.org/hubs/OSKB>), [Leitfäden](#) und Glossare (Parsons et al. 2022), um den Zugang zu Open Science zu erleichtern. Eine noch unerfüllte Forderung ist die, Forschende das System über Zusammenschlüsse im Rahmen von Gewerkschaften zu reformieren (Rahal, Fiedler, Adetula, Berntsson, Dirnagl, Feld, Fiebach, Himi, Horner, Lonsdorf, and others 2023).

14.0.6 Bewertungskriterien

https://www.researchgate.net/publication/351638932_The_Natural_Selection_of_Good_Science

- Systematische Erarbeitung objektiver Kriterien
 - o <https://psycharchives.org/en/item/dca8878a-1f6a-4599-abe0-16134b4b7f64>
 - o <https://osf.io/preprints/psyarxiv/5yexm>
- Peer review guidelines https://www.researchgate.net/publication/375607348_Quantitative_manuscript_peer_review_template
 - <https://psyarxiv.com/rgh5b/>
 - <https://psyarxiv.com/5yexm/>

•

- Alternative Auswahlkriterien für Professor*innen: <https://osf.io/preprints/metaarxiv/juwck>
-
- Multilevel-Selektion: <https://journals.sagepub.com/doi/pdf/10.1177/17456916231182568>
-
- <https://tu-dresden.de/mn/psychologie/ifap/differentielle-psychologie/die-professur/Mitarbeiter/dipl-psych-anne-gaertner> Einstein Award Thema (Auswahlkriterien)
-
- Problem von Kriterien ist, dass sie oft Dinge jenseits des Mainstreams benachteiligen
<https://open.lnu.se/index.php/metapsychology/article/view/3764>
-

14.0.6.1 Alternativen zum Impact Factor

Altmetric für Einfluss jenseits von Publikationen (z.B. in sozialen Medien, Wikipedia, news outlets),

DORA: <https://sfedora.org/read/read-the-declaration-deutsch/>

COARA <https://coara.eu> Alternative Bewertungskriterien (qualitativ, nicht quantitativ; kann man signieren)

TOP guidelines

- https://www.researchgate.net/publication/369670371_Evaluating_Implementation_of_the_Transparency_and_Openness_Promotion_Guidelines_Reliability_of_Instruments_to_Assess_Journal_Policies_Procedures_and_Practices

CiTO <https://sparantologies.github.io/cito/current/cito.html>

Free Lunch Index <https://link.springer.com/article/10.1007/s11192-023-04862-8>

Im Rahmen von XXX wurden Richtlinien zur Förderung von Transparenz und Offenheit in der Wissenschaft erstellt (Transparency and Openness Promotion guidelines; kurz: TOP guidelines). Eine Zeit lang gaben manche Zeitschriften freiwillig an, welche Richtlinien wie eingehalten werden. Darin befindet sich zum Beispiel die Frage, ob mindestens eine Studie eines publizierten Artikels präregistriert sein muss, oder ob Datensätze veröffentlicht werden müssen. Je Aspekt (z.B. Öffentlichkeit der Daten) gibt es dann XX Levels: Das niedrigste Level 0 gibt an, dass es zu dem jeweiligen Aspekt keine Empfehlungen oder Vorschriften gibt. Das höchste Level 3 hingegen fordert, dass alle verwendeten Datensätze online verfügbar sind. [XXX genau anpassen an Vorgaben].

Seit 2022 lassen sich Zeitschriften, die TOP guidelines auf verschiedenen Levels umsetzen, in der topfactor.org Datenbank durchsuchen. Somit haben Forschende eine Alternative zum Impact Factor, die tatsächlich mit der Qualität der Zeitschrift zusammenhängt.

14.0.7 Infrastruktur (Open Infrastructure)

- ermöglichend
- NFDI
- Prinzipien: <https://openscholarlyinfrastructure.org> (übersetzen, einfügen, zitieren)
- Beispiele
 - o Wege, um Materialien, Daten, oder Ergebnisse
- § Literaturdatenbanken (OpenAlex)
- § (Data) Repositories, re3data.org; OSF.io
- § Pre-Print Server (arXiv), Zeitschriften, Zeitschriftensysteme (OJS)
- § Begutachtungssysteme (PCI)
- § Post-Publication Review
- o Meta-Daten, um Verweise zu ermöglichen (ORCID-ID, ROR <https://ror.org>, DOI)
- § ORCID-ID Kritik: https://www.sciencedirect.com/science/article/pii/S0099133324000144?dgcid=rss_sd_all
- Übersicht: <https://kumu.io/access2perspectives/open-science#disciplines/by-os-principle/open-infrastructure?focus=1> von <https://access2perspectives.org/mapping-open-science-resources/>

14.0.8 Open Access Publikationen

Typen von Open Access (siehe zenodo NRW AG Open Science Auflistung)

An manchen Orten gibt es ungeschriebene Gesetze wie “wenn du während deiner Promotion in einem hochrangigen Journal publizierst, wirst du die Bestnote kriegen”. Damit wird den sogenannten prestigereichen Zeitschriften immer mehr Macht zugeschoben. Leuten wird außerordentlich dafür gratuliert, dass sie etwas in Psychological Bulletin oder sogar Nature Human Behavior veröffentlicht haben. Dass niemand, der nicht mit einer Universität affiliert ist (also an einer arbeitet oder studiert), den Artikel im Psychological Bulletin lesen kann, oder dass für die Veröffentlichung in Nature XXX Euro bezahlt wurden, spielt dabei keine Rolle. Leider sind genaue Regeln, wann welche Artikel von diesen Zeitschriften veröffentlicht werden intransparent. Auf den Seiten der Zeitschriften sind zwar üblicherweise Hinweise für

Autor*innen und kurze Texte über die Zeitschrift, ob eine spezifische Studie genommen wird oder nicht, entscheiden zuerst alleine die Herausgeber*innen oder auf Englisch Editors. Zwischen zahlreichen Einreichungen müssen sie diejenigen auswählen, die an die Gutachtenden weitergeleitet werden.

Diesem intransparenten System stehen Open Access Zeitschriften entgegen. Obgleich ihr Kern die Öffentlichkeit aller Artikel ist, geht das Prinzip oft damit einher, dass die Ergebnisse des Gutachtenprozesses ebenfalls veröffentlicht werden.

- Guide

- o <https://psycharchives.org/en/item/4d0f12d8-a542-4cfd-b718-9ee7c9d11f5b>

- Paywall umgehen

- o Sci-hub: Schattenbibliothek, in der wissenschaftliche Zeitschriftenartikel gratis herunterladbar sind (<https://de.wikipedia.org/wiki/Sci-Hub>).

Daten über Sci-Hub: 2018 beinhaltet Sci-Hub fast 70% aller 81.6 Millionen wissenschaftlichen Artikel (<https://doi.org/10.7554/eLife.32822>). Nutzungsstatistiken aus Deutschland von Strecker (2019) aufbereitet.

- o 12ft.io
- o #canihazpaper
- o Autor*innen persönlich anschreiben (z.B. Mail, Researchgate.org, Academia.edu)

- Offene zeitschriften

- o Directory of Open Access Journals: <https://doaj.org>
- o Free Journal Network: <https://freejournals.org/current-member-journals/>
- o Journals finden: <https://service.tib.eu/bison/>
- o TOP Factor

- Predatory Publishers

- o Bealls Liste, <https://beallslist.net>
- o <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5493177/>
- o <https://thinkchecksubmit.org> (oder bei Konferenzen: <https://thinkcheckattend.org>)
- o White list: <https://doaj.org>
- o Bei Bibliotheken nachfragen!

- Monitoring

- o OpenAPC

- o <https://open-access-monitor.de>
- o <https://hal.science/hal-04121339v2>
- o Offener Datensatz mit von AI kodierten Indikatoren von Open Science des Verlages PLOS: https://theplosblog.plos.org/2023/10/open-science-indicators-q2-2023/?utm_id=354&utm_source=internal&utm_medium=email&utm_campaign=354lib1023&utm_content=link
- Open Acces Typen
 - o (moving wall
 - o Green
 - o Gold
 - o diamond), siehe NRW OS Strategie <https://zenodo.org/record/8322048> s. 11-12
 - o “guerilla open access” (sci-hub)
 - o
- Open Journals:
 - o <http://www.theoj.org>
 - o OJS
 - o Directory of open access journals
- Offene Bücherei Badge: <https://badge.openbiblio.eu>
- DEAL Verträge mit Publishern: <https://deal-konsortium.de/publizierende>
- Bücher (Open Textbooks): <https://content.iospress.com/articles/education-for-information/efi190260>
- Editorial Mass Resignations
 - o Sprachwissenschaften: Linguista à Glossa (Johan Rooryck; Vortrag am 10.01.2024)
- § Herausgeber (Wissenschaftler) hat sich um Zeitschrift gekümmert, Elsevier Vertrag um Typesetting
- § Wissenschaftler*innen wollten Open Access Zeitschrift mit APCs, die transparent gemacht werden
- § 2015: 4 weitere Zeitschriften, die durch Community geleitet wurden, haben ebenfalls Diamond OA durchgesetzt
- § Unterstützung durch Bücherei und Open Library of Humanities
- § Eine vollständige Community wurde erfolgreich zu Diamond OA verschoben

§ Zeitschrift als Vehikel für eine wissenschaftliche Community – wie ein Auto, das von einer Familie gekauft wird: man lässt das alte zurück und nimmt das neue, Editors, Reviewers, Readers, sind alle dieselben Personen

§ Übergang zu OA ist schwierig

- Finanzielle Ungewissheit
- Impact Factors benachteiligen neue Journals, sind unabhängig von Content, Community, Mission Statement
- Alle Herausgeber*innen müssen zustimmen
- Montgomery & Neylon 2019: „The value of a journal is the community it creates, not the papers it publishes” <https://blogs.lse.ac.uk/impactofsocialsciences/2019/03/29/the-value-of-a-journal-is-the-community-it-creates-not-the-papers-it-publishes/>
- Was eine Zeitschrift nicht sein sollte: ein Vehikel zur Kommunikation Forschender, ein juristisches Objekt, das einem Verlag gehört, ein Ort, an dem Prestige gesammelt wird, eine willkürliche Selektionsmaschine, eine Geldmaschine → EINE COMMUNITY, Qualitätskontrolle
- Shared Ownership à Zeitschrift kann nicht an Verlag verkauft werden
- Content vs. Service: content controlled by academic community, technical services (publishing, typesetting) can be paid for
 - o Neuroimage (Chris Chambers)

§ Übergang zu OA 2020 mit APC (mehrere tausend Dollar)

§ Pandemie war zeitgleich, Community fand es nicht gut, hat es aber mitgemacht

§ Wegen großem Erfolg wurden APCs erhöht (3150\$?)

§ Community hat viele Jahre hart gearbeitet und wird dann mit höheren Kosten bestraft, hat aber nichts von den Kosten

§ Chief Editor (Steve Smith) wollten Reduktion auf 2000\$ verhandeln, Elsevier wollte das nicht: „Market Forces support the current APC” und fand die Kosten absolut berechtigt. Solche Kosten sind eine große Hürde für globalen Süden.

§ Editorial Board hatte genug und alle haben resigniert. 3 Journals: Neuroimage, Neuroimage reports, Neuroimage clinical à ~40 Leute, haben das koordiniert, um öffentliche Wirkung zu maximieren (mit Medien, sozialen Medien, haben mit non-profit Publisher gesprochen)

§ Kommerzielle Verlage sind Parasiten, fügen der Wissenschaft keinen Wert hinzu, existieren nur, um Profit aus Wissenschaft zu ziehen und geben den Wissenschaftler*innen fast nichts zurück (höchstens Typesetting)

§ Wichtige Message an alle: „from this point onwards, it is socially unacceptable to support these journals” “it’s important that we do this all together”

§ Die Zeitschriften nun so gut wie Tod à nicht einfach ein anderes Journal erstellen, sondern klar machen, dass der Name nichts bedeutet, sondern die Community

§ Und das reicht noch nicht: Mass Resignations sollten eher wöchentlich stattfinden, sind aber ziemlich selten; strong editor in chief ist nötig; wir brauchen auch gar keine Publisher (Peer Community In Initiative zeigt das)

§ Verlag (Elsevier) hat wohl versucht, öffentliche Kommunikation über EMR zu beeinflussen und versucht, Forschende einzuschüchtern

o Critical Public Health (Judith XXX)

§ Editorin hatte Moratorium für 1 Jahr, laut Vertrag durfte sie also 1 Jahr nach Ende des Vertrages Verbot, für eine Zeitschrift zu arbeiten

§ Verlag wollte Verträge ändern: statt „page budget” (5 issues pro Jahr mit vorgegebener Seitenzahl) „minimum number of papers” Modell; Editorial Board hatte dabei kein Mitspracherecht

§ Graduelle Erosion von akademischer Kontrolle

§ 2023: Board unilateral resigned (erstes Mal, dass sich alle einig waren)

§ <https://cphn.net/breaking-news/>

§ Journal of Critical Public Health neu gegründet, über OJS gehostet

§ Schwierigkeit:

· Netzwerk, dem das Journal gehört, musste formalisiert werden (bisher vor allem jiscmail Liste)

· Nachhaltigkeit ist schwierig langfristig mit diamond Open Access

· Keine Indizierung bisher (ISSN, IF, Abstracting)

· Legale Fragen zum Intellectual Property unklar und nicht geprüft: Die Community hat Submission Guidelines usw. geschrieben, Verlag (Taylor and Francis) behauptet nun, dass diese Texte dem Verlag gehören. Wissenschaftler*innen haben keine Kapazitäten für Gerichtsprozess.

o Neuropathologie: XXX à Free Neuropathology

o Verlage sind dagegen ausgestattet

§ Editorial Board wird rotiert, damit sich Editors nicht so stark vereinen können

§ Unwahrscheinliche Treffen in Person

§ Non-compete Klauseln / Moratorien

- o Liste von EMR: <https://retractionwatch.com/the-retraction-watch-mass-resignations-list/>
- o <https://sparceurope.org/the-seed-of-a-global-federation-for-diamond-open-access-has-been-planted/>

Eine besondere Form des öffentlichen Zugangs verbreitet sich aktuell in der biotechnologischen Forschung aus. Damit Wissenschaftler*innen, Ingenieur*innen, aber auch die allgemeine Bevölkerung Zugang zu Konstruktionen hat, wird dort zu Klemmbausteinen wie beispielsweise LEGO® Steinen gegriffen [eigenes Foto von Lego Steinen) <https://doi.org/10.1016/j.tibtech.2022.02.003> . So hat David Aguilar (Tabelle 1) beispielsweise einen prosthetischen Arm mit Klemmbausteinen entworfen.

14.0.9 Pre-Prints

Wie bereits im Kapitel zur Begutachtung von Pre-Prints erläutert, sind Pre-Prints immer öffentlich und kostenlos verfügbar. Forschende können verschiedene Lizenzen vergeben, die beispielsweise eine kommerzielle Verwendung verbietet, sind dabei jedoch fast immer sehr liberal. Neben den bereits diskutierten Vorteilen von Pre-Prints, sind Änderungen bei ihnen um ein vielfaches schneller: Forschende können jederzeit auf Kritik reagieren und Fehler korrigieren. In einem [Artikel zum Coronavirus](#) fiel ein Fehler kurz nach der Veröffentlichung auf und wurde innerhalb von zwei Tagen korrigiert. Bei einem Zeitschriftenartikel kann sich dieser Prozess über viele Jahre ziehen. Gerade bei ernsthaften Problemen, die eine Retraction zur Folge haben, sind Verlage vergleichsweise langsam. Einen Pre-Print können Autor*innen jederzeit vom Netz nehmen - wobei er über Suchmaschinen dabei sicherlich noch auffindbar bleibt.

Pre-Prints können außerdem dazu beitragen, dass Ressourcen effizienter benutzt werden: Durch die beschleunigte Kommunikation zwischen Forschenden fällt schneller auf, wenn verschiedene Gruppen an derselben Fragestellung arbeiten. So können sich Kooperationen bilden oder Gruppen verlagern ihre Schwerpunkte.

i Unberechtigte Sorge vor Ideenklau

In manchen wissenschaftlichen Disziplinen sind Forschende Pre-Prints gegenüber zögerlich. Sie haben Angst, dass jemand ihre Idee klaut und schneller als sie einen Artikel bei einer Fachzeitschrift dazu veröffentlicht. Während diese Angst beim klassischen System z.B. durch böswillige Gutachtende oder Zuhörer*innen bei einer Konferenz ebenfalls besteht und selbst bei veröffentlichten Fachzeitschriftenartikeln passiert, ist der Vorteil von Pre-Prints, dass sie mit einem Datum versehen sind und für alle klar nachvollziehbar ist, wann was veröffentlicht wurde.

14.0.10 Ansätze gegen die Selektion spannender Ergebnisse

Die Ergebnisse einer Untersuchung sind das, was am wenigsten in der Hand der forschenden Person liegt (bzw. liegen sollte - immerhin interessiert uns ja die Wahrheit und nicht die Kompetenz Forschender, Daten möglichst stark zu schönen). Umso frustrierender ist es, dass Zeitschriften das Ergebnis als Kriterium zur Publikation verwenden. Unter der Vielzahl von Einreichungen werden vor allem diejenigen Artikel gewählt, die spannende Ergebnisse erzielt haben oder die ihre anfängliche Vermutung bestätigen konnten (*Confirmation Bias*). Die folgenden Ansätze lösen dieses Problem zum Beispiel dadurch, dass die Ergebnisse aus dem Begutachtungsprozess ausgeschlossen werden.

14.0.10.1 Results-blind peer review

Der einfachste Weg ist dabei, den Ergebnisteil einfach zu schwärzen oder wegzulassen. Verschiedene Zeitschriften bieten das als Option an. Da es sich hierbei zurzeit (2024) eher um eine Ausnahme handelt, ist den meisten Gutachtenden jedoch klar, dass vor allem diejenigen die Option zum results-blind peer review wählen, deren Ergebnisse nicht “hübsch genug” für den klassischen Weg sind.

14.0.10.2 Registered Report

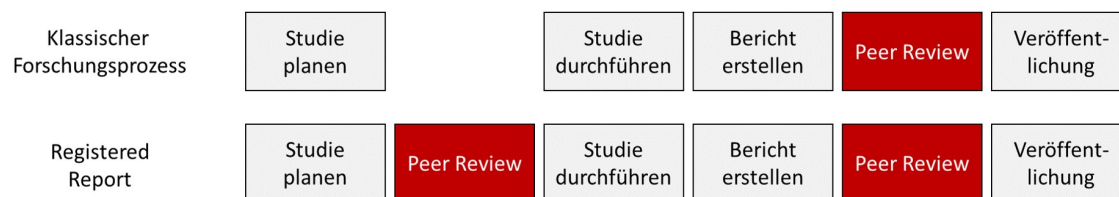


Figure 14.2: *Klassisches Vorgehen beim Forschen und Veröffentlichen im Vergleich mit Registered Reports: Bei letzterem wird bereits der Studienplan (für gewöhnlich in Form eines unvollständigen Berichtes) begutachtet und revidiert. Eine Ablehnung im zweiten Peer Review aufgrund der Ergebnisse ist nicht möglich.*

Ein radikalerer Ansatz als die Begutachtung ohne Ergebnisteil ist die Begutachtung des Artikels, ohne dass Ergebnisse überhaupt existieren. Dieses Format heißt *Registered Report*. Dabei wird das Manuskript mit der zu prüfenden Theorie, Methodik, und dem Analyseplan bei der Zeitschrift eingereicht, ohne dass überhaupt Daten erhoben wurden. Kommt es zur Akzeptanz dieses „halbfertigen“ Artikels (*in principle acceptance*), werden die Daten gesammelt, wie geplant ausgewertet, und es folgt eine weitere Begutachtungsrunde. Hierbei ist

vorgeschrieben, dass die Autor*innen nichts an den bereits verfassten Teilen verändern dürfen und die Gutachter*innen im Nachhinein keine Kritik am bereits geprüften Vorgehen üben dürfen. Es geht nur noch darum, ob der Plan eingehalten wurde und ob die Schlussfolgerungen auf den geplanten Analysen fußen. Damit soll verhindert werden, dass Artikel abgelehnt werden, weil die Ergebnisse nicht spannend genug, innovativ genug, oder den Erwartungen entsprechend sind. Erste Untersuchungen können bereits nachweisen, dass sich damit die Qualität der Forschung gegenüber dem traditionellen Vorgehen verbessert (Soderberg et al. 2021). Eine Übersicht über Zeitschriften, die dieses Format anbieten ist online verfügbar (<https://www.cos.io/initiatives/registered-reports> à Participating Journals; (Chambers?); Chambers and Tzavella (2022)). Ebenfalls wird dadurch deutlich, dass Forschung einem massiven Publikationsbias unterliegt (d.h. es werden vor allem Studien veröffentlicht, die ihre Vermutungen bestätigen konnten und kaum Studien, in denen das nicht geschah): (Scheel2021?) zeigten, dass der Anteil erwartungskonformer Ergebnisse bei Registered Reports mit 44% deutlich unter den in der Psychologie üblichen 96% liegt.

14.0.10.3 Pre-Print basierte Modelle

Durch die immer häufigere Veröffentlichung von Pre-Prints, also noch nicht begutachteten Manuskripten, eröffnen sich für die Begutachtung neue Wege. Sogenannte Overlay Journals (eLife) wählen unter Pre-Prints solche aus, die sie an Gutachtende schicken um deren Meinungen einzuholen. Sofern die Autor*innen des Preprints einverstanden sind, erhalten sie dann Gutachten und ihr Artikel wird schließlich in der Zeitschrift veröffentlicht.

Je nach Fach haben unterschiedlich viele Zeitschriften mit PCI Initiativen Vereinbarungen, dass sie die akzeptierten Artikel ohne eigenes Peer Review veröffentlichen. Mehr und vor allem bekannte teilnehmende Zeitschriften machen PCIs für Forschende attraktiver. Zeitschriften, die Interesse an einem PCI haben, aber die Begutachtung nicht aus der Hand geben möchten, können sich als „PCI interested Journals“ listen lassen (vs. „PCI friendly journals“). Teilnehmende Zeitschriften sparen dadurch Arbeit und bleiben relevant, indem sich Ihre Funktion dahin verschiebt, dass sie thematisch relevante Forschung sammeln und disseminieren. In einem Fall hat bereits eine Zeitschrift, die als „PCI friendly“ eine Vereinbarung mit PCI-RR hatte, ein Manuskript mit einer *Recommendation* zum erneuten Peer Review versendet und wurde sofort von den PCI Partnern entfernt. Forschende, die Gutachtenprozesse für PCIs organisieren möchten – analog zu Herausgeber*innen von klassischen Zeitschriften – können *Recommender* werden und müssen dazu ein Mindestmaß an Wissen haben sowie eine Schulung absolvieren (<https://rr.peercommunityin.org/about/recommenders>).

Tabelle 2

Zusammenfassung der verschiedenen Begutachtungsmodelle und der jeweiligen Art des Umgangs mit den Forschungsergebnissen

Begutachtungsprozedur	Ausblendung der Ergebnisse
-----------------------	----------------------------

Traditionell	Ergebnisse sind sichtbar und fließen in die Beurteilung ein
Results-Blind Peer Review	Ergebnisse liegen vor, werden den Begutachtenden jedoch vorenthalten
Registered Report	Ergebnisse liegen noch nicht vor
Peer-Community-In Registered Report (PCI-RR)	Ergebnisse liegen noch nicht vor

14.0.10.4 Open peer-review

So können Leser*innen von Meta-Psychology oder dem Journal of Open Psychology Data zu jedem Artikel einen Review Report (XXX richtiges Wort? Link einfügen) lesen und nachvollziehen, was die Reviewer genau getan haben und wie sich der Artikel dadurch entwickelt hat. Vereinzelt Studien zufolge (REF, REF) führt das offene Review zu faireren und positiveren Einschätzungen.

XXX: Arten von Open Access einfügen - Gold – Diamond

Weitere Peer-Review Modelle

- <https://reuse-dept.org> Replikation in Computer Science

Review Ampel

Meine Erfahrungen mit Peer Review

Peer Review ist brutal. Das ist meine persönliche Erfahrung mit dem Prozess. Schnell habe ich in meiner wissenschaftlichen Arbeit gelernt, dass es in vielen Fällen ein Glücksspiel ist. Renommierete Wissenschaftler*innen erklärten mir, dass sie Artikel hätten, die sie über Jahre immer wieder bei Zeitschriften immer wieder eingereicht hätten, und die schließlich positiv aufgenommen worden wären, ohne, dass sie sie stark verändert hätten. Jenseits von der Akzeptanz eines Artikels zur Publikation geht es auch um die Gründe für die Ablehnung: Häufig lesen Gutachtende Artikel nicht aufmerksam und Kritiken sind nicht konstruktiv. Hierzu ein paar Beispiele:

Einreichung bei Collabra: Es handelte sich um einen Artikel, der zwischen den Disziplinen steht. Es geht nicht nur um Erwartungen, nicht nur um Produktbewertungen, nicht nur um die Methode, Daten direkt aus dem Internet herunterzuladen. Der “bunte-Vogel-Artikel” war bereits bei drei Zeitschriften abgelehnt worden. Bei keiner wurde er an die Reviewer weitergegeben, weil er nie zur Zeitschrift passte. Die Zeitschrift *Collabra*, bei der wir ihn schließlich einreichten, war zu dem Zeitpunkt noch wenige Jahre alt und war breit aufgestellt. Nach der Einreichung im Mai 2020 haben wir über ein halbes Jahr lang auf das Gutachten gewartet. Länger zu warten ist erstmal ein gutes

Zeichen: Der Artikel wurde wohl an Reviewer rausgeschickt. In dem Fall wurden wir jedoch bitter enttäuscht: Im November 2020 erfuhren wir auf Nachfrage hin, dass 14 Gutachter*innen angefragt wurden, sodass die Herausgeberin auf Basis eines Gutachtens entschied, das Manuskript abzulehnen. Grund dafür war, dass die Methode aufgrund der inzidentellen Daten nicht geeignet für die Fragestellung war. In der Psychologie sowie den Wirtschaftswissenschaften ist das Problem mit “echten Daten” seit mehreren Jahrzehnten bekannt und wir hatten es bereits ausgiebig im Artikel diskutiert.

Einreichung bei Journal of Experimental Social Psychology: Wir hatten eine Replikation einer dort erschienen Studie durchgeführt - der Befund ließ sich nicht replizieren. Meine Überlegung war, der Zeitschrift, die den nicht robusten Befund ursprünglich publizierte, selbst die Chance zur Selbstkorrektur zu geben. Die Reviews waren fair und positiv, es gab ein paar Punkte zu diskutieren, aber uns war klar, dass es sich hier um Verständnisprobleme und keine inhaltlichen Aspekte handelte. Nicht so der Editor: Er erklärte, dass der Befund nicht neu genug wäre und klar wäre, dass es nicht replizierbar ist. Ich erklärte ihm, dass noch niemand den Befund zu replizieren versucht hatte und wir selbst sogar vor der Analyse der Ergebnisse eine Abstimmung darüber gemacht hatten, welches Ergebnis wir erwarteten: Es war sehr ausgewogen 50-50. Darüber hinaus, wurde ein weiterer Artikel publiziert, der entgegengesetzt Ergebnisse hatte. Wir stellten klar, dass wir alle aufgezeigten Probleme einfach lösen können und gerne die Chance zur Revidierung hätten. Der Editor hätte dabei kaum Arbeit außer den Artikel anschließend nochmal an Gutachtende zu schicken. Auf unsere Mail erhielten wir nur eine kurze Antwort: *Hi. I know my decision is disappointing, but I'm going to stick with my decision on this one.* Hier befand ich mich an einem Scheideweg: Warum ist ein Forscher nicht bereit, Gründe für eine wissenschaftliche Entscheidung zu erörtern? Wir entschieden uns, einen anderen Editor direkt zu kontaktieren. Nach kurzer Zeit erhielten wir die Einladung zu einer Revision. Der Artikel wurde schließlich in der revidierten Fassung veröffentlicht.

Einreichung bei European Journal of Personality: Die zentrale Aussage dieses Artikel war, dass verschiedene Messwerte einer angeblichen Eigenschaft nicht miteinander zusammenhängen. Der Befund stellte die Annahme infrage, dass es sich dabei überhaupt um eine Eigenschaft handelte. Die Ablehnungsgründe zweier Gutachtenden und der Herausgeberin machten deutlich: Niemand hatte den Artikel überhaupt gelesen. Ein Gutachter merkte an, dass etwas mit den Werten nicht stimme, weil sie laut einer der Tabellen nicht miteinander zusammenhängen. Genau das war ja unsere Aussage. Wir zeigten, dass es nicht an unseren Daten lag, sondern sich in anderen Datensätzen so verhielt. Hätte er die Überschrift der Tabelle gelesen, den Absatz davor, oder den danach, wäre das klar geworden. Hat er aber nicht...

Das sind nur kurze Auszüge aus dutzenden Einreichungen und Ablehnungen. Darüber hinaus kann fast jede*r Forschende*r von substanzlosen persönliche Beleidigungen berichten. Meiner Erfahrung nach ist anonymes versus öffentliches Peer Review wie ein Vergleich von anonymen Kommentarspalten mit nicht anonymen: Bei Anonymität

14.0.11 Open Science Incentives FORRT NOSI Badges Open Science in Berufungskommission

- N-best evaluation (<https://osf.io/preprints/psyarxiv/tzaqh>): nur die z.B. $N = 5$ für die ausgeschriebene Stelle wichtigsten Publikationen werden berücksichtigt, statt auszuzählen werden diese dann ganz genau an
- StudySwap “However, a potential barrier to independent, prepublication replication attempts is that many researchers have a difficult time finding other labs to conduct such attempts. StudySwap can be used to find an independent research team to conduct a replication attempt of a not-yet-published study” (Chartier et al., 2018, p. 575). Zitat von: <https://osf.io/preprints/psyarxiv/dtvs7/> p. 11
- Reviewer Initiativen und Boycotts
Richard D. Morey et al. (2016) The Peer Reviewers’ Openness Initiative: incentivizing open research practices through peer review, Royal Society Open Science; DOI: 10.1098/rsos.150547. Published 13 January 2016

14.0.12 Replikationsforschung[LR49] : Ansehen und Kompetenz (Normalität) im Umgang mit Replikationsstudien steigern

-
- Unterscheidung Vertrauenskrise und Replikationskrise: <https://philarchive.org/archive/FEEWRI>
 -
 - Replikationsstudien können keine Forschung ersetzen, die auf ein tieferes Verständnis von Theorien abzielt
 -
 - Replizierbarkeit betrifft Robustheit, Mindestanforderungen an Verallgemeinerbarkeit, und Erhöhung des Vertrauens dadurch, dass QRPs und Datenfälschung weniger wahrscheinlich werden [was, wenn original- und replikations-studie gefälscht sind?]
 -
 - Kein Schwarz-weiß-denken: Unsicherheiten sind mit Replikations- und Original-Studien verbunden <https://journals.sagepub.com/doi/abs/10.1177/10755470241239947?journalCode=scxb>

-
-
- Es wird klar, was replizierbar sein sollte und was nicht.
-
- Wahre, statistische Phänomene können nicht replizierbar sein
 - *
 - * Methode könnte sehr ungenau sein und trotz „wahrem“ Befund nur eine sehr geringe Replikationsrate haben (mathematischer Beweis dafür bei <https://royalsocietypublishing.org/doi/10.1098/rsos.200805#d1e454>)
 - *
 - * Mere-Measurement Effect
 - *
 - * Ethnografie, qualitative Methoden, Archäologie („Befund wird bei Ausgrabung aus Kontext gerissen und dadurch zerstört“)
 - *
-
- Sachen, die keine allgemeinen Gesetzmäßigkeiten darstellen können trotzdem replizierbar sein
 - *
 - * Artefakte
 - *
 - * Unpassende Modelle, deren Annahmen inkorrekt sind, lassen sich manchmal reproduzieren und mit neuen Daten auch replizieren
 - *
 - * Heisenbug (?)
 - *
-
- Replizierbarkeit ist nicht hinreichend für Wahrheit
-

- Belief updating nach Replikationen <https://www.nature.com/articles/s41562-021-01220-7>
 - <https://royalsocietypublishing.org/doi/10.1098/rsos.200805#d1e454>
 - listen notwendige [LR51] Bedingungen dafür auf, dass „wahre“ Befunde replizierbar sein können
 -
 - Ziele von Replikationen
 - Fletcher (2020) describes the function of replication is to confirm one of the following claims regarding the findings of the original research: (1) they are not due to mistakes in data analysis, (2) they are not due to sampling error, (3) they do not depend on contextual factors, (4) they do not arise from fraud or questionable research practices, (5) they generalize to a larger population than that sampled in the original, (6) their aspects pertaining to the theoretical hypothesis of interest hold even when that hypothesis is operationalized or tested in completely different ways
 - <https://cornerstone.lib.mnsu.edu/cgi/viewcontent.cgi?article=2424&context=etds>
 -
- 💡 Notwendige und hinreichende Bedingung

erklären
- Niemand repliziert
 -
 - <https://www.econstor.eu/bitstream/10419/267931/1/I4R-DP013.pdf>
 -
 - <https://osf.io/preprints/psyarxiv/fzngs>
 -
 - <https://journals.sagepub.com/doi/10.1177/1745691620979806>
 -
 - <https://peerj.com/articles/7654/>
 -
 - <https://doi.org/10.1016/j.respol.2018.07.019>

-
- <https://doi.org/10.1111/lang.12286>
-
- <https://journals.sagepub.com/doi/10.1177/0741932516646083>
-
- <https://journals.sagepub.com/doi/10.1177/1477370815578197>
-
- <https://journals.sagepub.com/doi/10.3102/0013189X14545513>
-
- <https://journals.sagepub.com/doi/10.1177/1745691612460688>
-
- <https://www.journals.uchicago.edu/doi/10.1086/506236>
-
- <https://osf.io/preprints/psyarxiv/sa6rc>
-
- <https://www.pnas.org/doi/abs/10.1073/pnas.2208863120>
-
- tweet von gilad: <https://twitter.com/giladfeldman/status/1735950123291779119>
-
-
- Geringes Ansehen verändert sich
 -
 - Vor 10 Jahren: <https://osf.io/preprints/psyarxiv/dtvs7/> p. 9 replications of Bem’s pre-cognition study were desk-rejected by the editor of JPSP, Eliot Smith, who stated “This journal does not publish replication studies, whether successful or unsuccessful” and “We don’t want to be the Journal of Bem Replication” (Aldhous, 2011).
 -

- Jetzt immer mehr Replikation als innovativ: “Defining replication as a confrontation of current theoretical expectations clarifies its important, exciting, and generative role in scientific progress.” <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000691>
-
- Mehr Replikationen, aber zB in Soc Psych noch nicht Mainstream <https://psycharchives.org/en/item/74463c10-7347-4bf2-8156-5618d42c4e93>
-
- Replikationen Initiative: <https://i4replication.org/reports.html>
-
- https://openresearch.amsterdam/image/2018/1/15/20180115_replication_studies_web.pdf
-
- Ruf nach Zeitschrift (<https://s3.amazonaws.com/real.stlouisfed.org/wp/2015/2015-016.pdf>) erhört (JCRE)
-
- Diskussionen sind zum Teil noch hart und Replikationsstudien werden falsch dargestellt <https://osf.io/96pnj>
-
-
- Arten von Replikationsprojekten
 -
 - RRR: https://www.researchgate.net/publication/366862422_A_Roadmap_to_Large-Scale_Multi-Country_Replications_in_Psychology
 -
 - Registered Repots /PCI-RR (steht dazu schon woanders etwas? Ist ja nicht explizit für replikationen)
 -
 - Abschlussarbeiten für Replikationen nutzen:
 - *
 - * <https://www.nature.com/articles/s41562-021-01192-8>

*

* Studierende als Replizierende: <https://link.springer.com/article/10.1007/s12286-023-00578-4>

*

* Hagen Cumulative Science Project:
<https://journals.sagepub.com/doi/10.1177/1475725719868149>

*

* Gilad Feldmans Projekte CORE <https://osf.io/5z4a8/>

*

—

•

§ Leipziger Replikationsstudien: <https://home.uni-leipzig.de/lerep/>

§ CREP <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00247/full>

§ Laufende Großprojekte

§ I4R und Replication Games: <https://www.nature.com/articles/s41562-023-01807-2>

§ ZWB Lab²

§ Beteiligung von Original-Autor*innen

§ <https://osf.io/preprints/psyarxiv/dwg9v>

•

- Veröffentlichung von Replikationen

—

- Vorschlag: jede Zeitschrift ist zuständig für Replizierbarkeit eigener Studien und MUSS Replikationsversuche veröffentlichen; pottery barn rule <https://thehardestscience.com/2012/09/27/a-pottery-barn-rule-for-scientific-journals/>

—

- Neue Journals

*

* Replikationen Journal: <https://www.jcr-econ.org>

*

- * Reproducibility in Neuroscience: <https://jrn.epistemehealth.com/about/editorialteam/>
- *
- * Journal of Trial and Error <https://trialanderror.org>
- *
- * <http://rescience.org/x> Replication Journals
- *
-
- Datenbanken für Replikationen
 - *
 - * Göttingen Replication Wiki https://replication.uni-goettingen.de/wiki/index.php/Special:FormEdit/New_Replication or <https://blog.repec.org/2020/08/04/a-replication-database-for-economics-and-social-sciences-the-replicationwiki/> (Jan H. Höffler) <https://doi.org/10.1045/march2017-hoeffler>
 - *
 - * CurateScience.org
 - *
 - * FORRT Replications and Reversals
 - *
 - * Replication Database (ReD)
 - *
-
- Auffinden von Replikationen: <https://arxiv.org/pdf/2311.15055.pdf>
-
-
- Ressourcen zum Lernen über Replikationen
 -
 - Review über Replikationen in quantitativer Soziologie: <https://www.annualreviews.org/doi/10.1146/annurev-soc-060116-053450>
 -

- Wiki für Studierende / Webinar in Göttingen: https://replication.uni-goettingen.de/wiki/index.php/Webinar_series:_Replicating_empirical_studies_in_economics_-_an_opportunity_for_students Replication Wiki
-
-
- Replikationsmethodologie weiterentwickelt
 -
 - Terminologie wie am Anfang vorgestellt (Hüffmeier, Paper zu Generalizability von dem akademischen Rat aus Bamberg [evolutionsbiologie])
 -
 - Small telescopes approach
 -
 - Bayesian sample size planning https://www.researchgate.net/publication/373047996_Bayesian_approaches_to_designing_replication_studies
<https://link.springer.com/article/10.1007/s11749-023-00916-4>
 -
 - Selection of replication studies <https://www.sciencedirect.com/science/article/pii/S0010945223002691>
 -
 - https://www.researchgate.net/publication/365181293_Bayesian_approaches_to_designing_replication_studies
 -
 - Empfehlungen, wie Autor*innen von Replikationsstudien mit den Originalautor*innen kommunizieren sollten: <https://doi.org/10.1017/S1049096520000943>
 -
 - Kriterien für Bewertung: <https://docs.google.com/document/d/1p7GeOpwzQyTuzAWsD1w3zql2dS/edit#heading=h.oti7bgwflneo>
 -
 - *
 - * Problem beim Updaten von Materialien: wenn original, dann veraltet und unpassend – wenn neu, dann anders und unpassend; siehe auch Kommentar von John Protzko <https://rr.peercommunityin.org/PCIRegisteredReports/articles/rec?id=750>

*

—

- Identifikation problematischer Forschungsliteratur: Original Replication of Meta-Analyses or ORMA https://www.researchgate.net/publication/365057816_ORMA_A_strategy_to_

—

•

§ Qualitative Forschung profitiert evtl. nicht von Replikationen <https://open.lnu.se/index.php/metapsychology/article/view/3764>

KASTEN / Leitfaden: Eine Replikationsstudie durchführen

1. Wahl der Studie

- Wichtigkeit der Originalstudie (Grundlegend, evtl. erkennbar durch viele Zitationen)
- Ungewissheit der Originalstudie (Replikationswahrscheinlichkeit möglichst nicht 0 oder 1)
- Bisher noch keine Replikationen
- Diskussionen dazu beachten (z.B. via hypothes.is oder pubpeer.com → auf pubpeer.com Paper suchen und schauen, ob es dazu Kommentare gibt)
- Machbarkeit (online möglich? Querschnitt? Instrumente Verfügbar bzw. bezahlbar?)
- Etablierte Replikationsstandards für den Bereich? (bei FMRI Daten oder Längsschnittanalysen schwieriger als bei einfachen F- oder t-Tests)
- Zusatz: Meta-analytische Indikatoren: auffällige P-Curve/Z-Curve, klare Hinweise auf Publikationsbias oder QRP

2. Durchführung

- Bestandsaufnahme der veröffentlichten Materialien und Daten
- Daten berechnen
- Bei verfügbaren Daten: computationale Reproduzierbarkeit prüfen
- Bei fehlenden Teststatistiken diese ausrechnen
- Bei fehlenden Werten zu allem möglichen, evtl. andere Studien hinzuziehen
- Kontaktaufnahme zu Originalautor*innen, wenn möglich
- Stichprobenplanung: nicht den Originalstichprobenumfang nehmen, sondern Poweranalyse-Ansätze nutzen
- Small-Telescopes-Approach

- Untere KI-Grenze
- Äquivalenztests
- Anpassung der Methode
- Ggf. Update veralteter Materialien
- Übersetzung in andere Sprache
- Notwendigkeit einer besonderen Stichprobe?
- Gibt es etablierte Standards bezüglich des statistischen Tests (v.a. bei standardisierten Effektstärken $[r, d, f]$, nicht bei komplexen Verfahren wie RSA, Clustering, Decision Trees, FMRI Daten)
- Gibt es ähnliche Replikationsstudien in dem Bereich, die für die Methode herangezogen werden können (z.B. via Replication Database suchen)
- Anpassungen
- Qualitätssicherung: Welche Veränderungen sind nötig oder sinnvoll? Was wissen wir neues, was wir berücksichtigen sollten?
- Extensions: Lassen sich zusätzliche Items o.ä. „hinten“ an die Studie anhängen, womit sich die externe Validität erhöhen könnte? Gibt es zusätzliche einfache Ergänzungen, bei denen eine Replikation nicht zu erwarten ist?
- Präregistrierung mittels Template (Empfehlung: Replication Recipe, Brandt et al., 2014 <http://dx.doi.org/10.1016/j.jesp.2013.10.005>)
- Ggf. als PCI-RR submitten
- 3. Analyse
 - Berechnung der Ergebnisse
 - Vergleich Original vs. Replikation
 - Vergleich naher vs. „ferner“ Ergebnisse (z.B. bei denselben Items vs. Bei Extension [falls vorhanden])
- 4. Diskussion
 - Bewertung des Ergebnis
 - LeBel Kriterien (signal/no signal + size)
 - Umgangssprache (success, failure, practical failure, inconclusive)
 - Abweichungen von Präregistrierung (jeweils Beschreibung, Begründung, und Auswirkung auf Ergebnisse)

- Abweichung von Originalstudie (jeweils Beschreibung, Begründung, und ob es die Ähnlichkeit der Ergebnisse hätte beeinflussen können)

- Bei anderen Ergebnissen Originalautor*innen um Kommentar bitten

5. Bericht

- Manuskript: Orientierung an state-of-the Art von Gilad Feldman (LINK/REF) und an I4R Replication Report Template (<https://osf.io/j2qrx>)

- Eintragung in Replication Database (via Fragebogen oder StaRT)

- Registrierung der Ergebnisse über Replication Recipe („post completion“)

- Ggf. Verweis auf Replikationsstudie unter der Originalstudie bei pubpeer.com

- Preprint: auf passendem Server hochladen

- Zeitschrift: Originalzeitschrift (meistens kein Interesse), Collabra: Psychology, PCI (wenn es eins gibt), Meta-Psychology (designierte Section für Replikationen); Hinweise zu Zeitschriften hier: <https://i4replication.org/publishing.html>

14.0.13 Modellierung von Replizierbarkeit

Alle Studien in einem Wissenschaftsbereich zu replizieren ist aktuell unrealistisch und würde enorme Ressourcen benötigen. Daher befinden sich verschiedene Methoden in Entwicklung, um auf Basis von Eigenschaften einer Studie vorherzusagen, welche Replikationsergebnis zu erwarten ist. Im [repliCATS](#) Projekt, das Teil des SCORE Projektes zur Messung wissenschaftlicher Qualität ist, werden Einschätzungen von Forschenden, Reproduktionsversuche, und Replikationsversuche kombiniert. Ähnliche Projekte verwenden komplexe statistische Modelle (z.B. Large-Language-Models) oder meta-analytische Methoden wie p-curve oder Z-curve zur Vorhersage von Replikationsraten. Solche Modelle benötigen üblicherweise sehr viele Beobachtungen und Vorhersagen sind nur auf der Ebene von hunderten Studien sinnvoll. Beispielsweise rechneten Boyce et al. 2023 und Hagen Cumulative Science project Moderationsanalysen, prüften also welche Eigenschaften von Studien sich auf das Replikationsergebnis auswirken (z.B. Wechsel von Studie im Labor zu Online-Studie). Mittels umfassender Datenbanken wie der FORRT Replication Database können in Zukunft präzisere Modelle erstellt werden. Aktuelle Ergebnisse sind als vorläufig und mit Vorsicht zu interpretieren, da sie auf nicht-zufälligen Stichproben basieren, die Studieneigenschaften nicht systematisch und zufällig variiert wurden, und Replikationen solcher meta-analytischen Befunde schwierig sind.

14.0.14 Veröffentlichung aller Ergebnisse

- <https://opentrials.net/about/>
- Unveröffentlichte Medizinische Studien: Quest Dashboard: <https://quest-cttd.bihealth.org/#tabStart>
- File drawer reports
 - o Meta psychology
 - o psychfiledrawer.org meta-psychology
- Datenbanken, die auch unveröffentlichte Ergebnisse aufnehmen (auch Lösung für cumulative science)
 - o Bodypositions, OpAQ Community Augmented: [metalab.stanford](https://metalab.stanford.edu), [metabus](https://metabus.org), <https://camarades.shinyapps.io/AD-SOLES/> , ...
 - o Replication database
 - o Metapsy <https://www.metapsy.org>
 - o Metalab stanford <https://langcog.github.io/metalab/>

Tabelle aktualisieren (siehe PPT + Metspy.org)

Name[LR54]	Link	Topics
MetaLab	langcog.github.io/metalab	Early Language, Cognitive Development
metaBUS	metabus.org	All of social sciences
SOLES	camarades.shinyapps.io/AD-SOLES/	Animal models of Alzheimer's disease
OpAQ	t1p.de/openanchoring	Anchoring Effects

FReD	t1p.de/ReD	Replications (all fields)
Power Posing	metaanalyses.shinyapps.io/BodyPositions/	Body Positions Effects
Metadataset	metadataset.com	Agriculture
METAPSY	https://www.metapsy.org	Psychotherapy

Anleitungen:

metaUI: <https://github.com/lukaswallrich/metaui>

Dynameta: <https://github.com/gls21/Dynameta>

Materialien: <https://cama.psychopen.eu>

- Journal of Articles in Support of the Null Hypothesis
- Journal of Negative Results <https://www.jnr-eeb.org/index.php/jnr/about>

14.0.15 Umgang mit Fehlern

- Fehlerkultur
- Retraction
 - o Standardisierte Retraction Notes: https://www.researchgate.net/publication/374082265_A_tale_of_three_retractions_a_call_for_standardized_categorization_and_criteria_in_retraction_statements
 - o Wissenschaftliche Fehler können extreme Folgen haben; Lancet MMR Autism Fraud <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2831678/>
 - o Einige Wissenschaftler*innen prüfen große Mengen an Artikeln, Elisabeth Bik z.B. hat zu Retractions von fast 600 und Korrekturen von fast 500 Artikeln beigetragen (<https://www.buzzfeednews.com/article/stephaniemlee/elisabeth-bik-didier-raoult-hydroxychloroquine-study>)

- o Aktuell Studie zur Sichtbarkeit von Retractions mit Möglichkeit, sich über zurückgezogene Artikel, die man zitiert, informieren zu lassen (RetractoBot, <https://www.retracted.net>) vom Bennett Institute for Applied Data Science at the University of Oxford.
- Pubpeer.com
- o Beispiel einer Diskussion eines Papers vom Nobelpreisträger Thomas Südhof <https://pubpeer.com/publications/DAF32F6DB6C166337E5381F769AE52#1>
- o Hohe Aufmerksamkeit führt zu Korrektur vieler Fehler, zum Beispiel von 12 Fehlern (Stand: Dezember 2023) im Südhof Lab (<https://med.stanford.edu/sudhoflab/science-resources/integrity---pubpeer.html>)
- Error project (malte elson)

14.0.16 Offene Lehre / Open Teaching / Open Educational Resources

- [https://zpid.cloud.panopto.eu/Panopto/Pages/Sessions/List.aspx?utm_campaign=ptos_videos#folderID=384a-4cf1-a8ea-ae500adc653"¬ificationBannerShown=true](https://zpid.cloud.panopto.eu/Panopto/Pages/Sessions/List.aspx?utm_campaign=ptos_videos#folderID=384a-4cf1-a8ea-ae500adc653)
 - Repositorien für Bildungs- und Lehrmaterialien (Beispiele u.a. von <https://leibniz-psychology.org/practices-and-tools-of-open-science/open-science-in-der-lehre>)
 - o <https://www.oerbw.de>
 - o <https://www.twillo.de/oer/web/>
 - o <https://portal.hoou.de>
 - o <https://zenodo.org>
 - o Osf.io
 - o <https://forrt.org/syllabus/> (spezifisch open science)
 - Offene Webinare, zB zu Replikationsforschung https://acm-rep.github.io/2024/posters/ACM_REP24_paper_30.pdf
 - Anleitungen und Glossare (Parsons et al., 2022) zum Beispiel wundervoll gesammelt bei FORRT.org
 - Open Scholarship Knowledge Base (OSKB)
 - ROSiE Knowledge Hub <https://rosie-project.eu/knowledge-hub/#!training/health-and-life-sciences>
 - Open Economics Guide <https://openeconomics.zbw.eu/en/>
 - ARIADNE
- [LR1]REF

- [LR2]<https://opusproject.eu/openscience-news/open-science-horizon-unescos-recommendation/>
- [LR3]https://www8.cao.go.jp/cstp/kokusaiteki/g7_2023/2023_bestpracticepaper.pdf
- [LR4]<https://www.consilium.europa.eu/en/press/press-releases/2023/05/23/council-calls-for-transparent-equitable-and-open-access-to-scholarly-publications/>
- [LR5]https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/triggers-audits-good-laboratory-practice-glp-studies_en.pdf
- [LR6]<https://zenodo.org/record/8322048>
- [LR7]<https://opal.landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMV18-2422.pdf>
- [LR8]<https://www.jmwiarda.de/https-www.jmwiarda.de-2024-05-28-bitte-nachschaerfen/>
- [LR9]Erhöhung der Offenheit an finnischen Universitäten <https://www.zbw-mediatalk.eu/de/2024/02/open-science-monitoring-offenheit-als-wichtiger-teil-guter-forschungspraxis-in-finnland/>
- Netzwerk Nachhaltige Wissenschaft
- <https://zenodo.org/records/11204833>
- [LR10]Das zu Universitäten?
- [LR11]Worum geht es dabei genau?
- [LR12]<https://arxiv.org/abs/2205.01833>
- [LR13]schönbrodt research practices I und II
- [LR14]<https://www.cambridge.org/core/books/abs/data-science-in-the-library/toward-reproducibility-academic-libraries-and-open-science/6FB18D7BCD8B82E597ED0D4B390C72EE>
- [LR15]<https://www.degruyter.com/document/doi/10.1515/bfp-2024-0008/html>
- [LR16]<https://de.wikipedia.org/wiki/Zweitveröffentlichung>
- [LR17]<https://doi.org/10.31234/osf.io/prx8w>
- [LR18]<https://zenodo.org/records/7193838>
- [LR19]ergänzen, Gesellschaft für Wirtschaftswissenschaften?
- o [LR20]Public Open Science Institut für die Psychologie (ZPID)
- [LR21]<https://openscience.cern>
- o [LR22]https://www.apa.org/pubs/journals/resources/publishing-tips/transparency-openness-promotion-guidelines?utm_campaign=apa_publishing&utm_medium=direct_

[email&utm_source=businessdevelopment&utm_content=openscience_promo_11302023&utm_term=text_middle_learnmore](#)

[LR23]<https://osf.io/bnh7p>

[LR24]<https://joss.theoj.org>

[LR25]erklären

[LR26]https://www.researchgate.net/publication/372509504_A_Guide_for_Social_Science_Journal_Editors_on_Easing_into_Open_Science

[LR27]Idee: Review als Kommentar bei Pubpeer posten bei Ablehnung

[LR30]<https://osf.io/preprints/psyarxiv/5hkjz>

[LR31]<https://s3.amazonaws.com/real.stlouisfed.org/wp/2015/2015-016.pdf>

[LR32]link

[LR33]<https://royalsocietypublishing.org/doi/10.1098/rsos.230206>

[LR34]<https://royalsocietypublishing.org/doi/10.1098/rsos.230206>

[LR35]boyce 2023 eleven years

[LR36]hagen cumulative science project

[LR37]forrt glossary

[LR38]<https://osf.io/preprints/psyarxiv/bkxfq>

[LR39]<https://open.lnu.se/index.php/metapsychology/article/view/3764>

[LR40]<https://blogs.lse.ac.uk/impactofsocialsciences/2019/03/29/the-value-of-a-journal-is-the-community-it-creates-not-the-papers-it-publishes/>

[LR41]

[LR42]Hier Bild von Lego-Steinen einfügen (selbst schießen)

[LR43]<https://doi.org/10.1016/j.tibtech.2022.02.003>

[LR44]<https://osf.io/preprints/metaarxiv/7x9vy>

[LR45]ergänzen:

· [10.1038/s41562-021-01193-7](https://doi.org/10.1038/s41562-021-01193-7)

[LR46]<https://journals.sagepub.com/doi/10.1177/25152459211007467>

[LR47]Irgendwo diskutieren, warum negative gutachten nicht veröffentlicht werden

[LR48]<https://osf.io/preprints/psyarxiv/tzaqh>

[LR49]hier nochmal durchschauen: https://lakens.github.io/statistical_inferences/17-replication.html#analyzing-replication-studies

und

<https://pubmed.ncbi.nlm.nih.gov/26214497/>

[LR50]<https://royalsocietypublishing.org/doi/10.1098/rsos.200805#d1e454>

[LR51]Kasten: Begriffserklärung für *Notwendige und hinreichende Bedingungen*

[LR52]Fletcher (2020) describes the function of replication is to confirm one of the following claims regarding the findings of the original research: (1) they are not due to mistakes in data analysis, (2) they are not due to sampling error, (3) they do not depend on contextual factors, (4) they do not arise from fraud or questionable research practices, (5) they generalize to a larger population than that sampled in the original, (6) their aspects pertaining to the theoretical hypothesis of interest hold even when that hypothesis is operationalized or tested in completely different ways

<https://cornerstone.lib.mnsu.edu/cgi/viewcontent.cgi?article=2424&context=etds>

[LR53]https://www.youtube.com/watch?v=qa2_9NrEmeA

[LR54]aktualisieren (siehe ppt)

[LR55]<https://rosie-project.eu/knowledge-hub/#!training/health-and-life-sciences>

14.0.16.1 Weiterführende Informationen

- (**Henderson2022-ci?**) beschreibt zehn einfache Regeln zum Schreiben eines Registered Reports
- Ein Erklärvideo zu PCIs ist online verfügbar (<https://www.youtube.com/watch?v=4PZhpnC8wwo>).

14.0.17 Literatur

15 Methoden

Große Menge an Lösungsvorschlägen, für einzelne Forschende oft überwältigend

Wegweise:

- Ariadne
- Kumu Access2perspectives <https://kumu.io/access2perspectives/open-science#disciplines/>

15.0.0.1 Meta-Analyses und Reviews

- use meta-analysis for replication studies https://www.researchgate.net/publication/360848749_Replication_Is_for_Meta-Analysis
- Richtlinien im Hinblick auf neue Standards: <https://osf.io/preprints/psyarxiv/59vsw>

15.0.0.1.1 Assessing publication bias

- https://www.researchgate.net/publication/363583595_Assessing_Publication_Bias_a_7-Step_User's_Guide_with_Best-Practice_Recommendations
- Toolbox: <https://www.sciencedirect.com/science/article/abs/pii/S0148296323005489>
- z-curve und z-curve validation preprint
- Test for excess significance (Ioannidis & Trikalinos, 2007)
- Incredibility Index (Schimmack, 2012)
- Funnel plot (Light & Pillemer, 1984)
- P-curve (Simonsohn, Nelson, & Simmons, 2014)
- Z-Curve (Bartos & Schimmack, 2020)
- Sensitivitätsbatterie (bushman kepes röseler körner)
- RoBMA bayesian <https://onlinelibrary.wiley.com/doi/full/10.1002/jrsm.1594>
- Cumulative science

- o Machine readable manuscripts
- o Reporting templates (e.g., Brandt et al., 2014, replication recipe post completion registration on OSF, REF successes and failures)
- o Dynamic meta analyses
- § Bodypositions,
- § OpAQ Community Augmented:
- § metalab.stanford,
- § metabus,
- § soles, ...
- § Replication database
- § ReproSci <https://reprosci.epfl.ch>
- § <https://portlandpress.com/clinsci/article/137/10/773/233083/Systematic-online-living-evidence-summaries>
- § metaUI r Paket
- o Multiverse analyses
- o

15.0.1 Reproducibility Checks / Reproduzierbarkeit

- Meta-Psychology
- o Societies: Besteht auf Reproducibility: <https://doi.org/10.15626/MP.2023.4020>
- Statcheck
- o <https://www.mdpi.com/2306-5729/1/3/14>
- o M nuijten paper
- <https://osf.io/mydzv/>
- Guidelines: <https://docs.google.com/document/d/1p7GeOpwzQyTuzAWsD1w3zql2dS6mFgEhdf38Lc3uX>
- Container / Capsules
- Journal für Computational Reproducibility: <http://rescience.github.io>
- Übergreifendes Gebiet: **Research Software Engineering** (RSE)
- o <https://arxiv.org/pdf/1609.00037.pdf>

- Großangelegte Reproduktionsprojekte
 - o Vor allem zwei Projekte, beteiligte Personen überschneiden sich, sind aber unabhängig
 - o Verwenden ordinalen Reproducibility Score
- § Erklärung <https://bitss.github.io/ACRE/assessment.html#levels-of-computational-reproducibility-for-a-specific-display-item>
- o Social Science Reproduction Platform, <https://www.socialsciencereproduction.org/metrics>
- § Leitfaden für Durchführung von Reproduktionen: <https://www.socialsciencereproduction.org/metrics>
- o Institute for Replication (I4R)
- § Eigene Datenbank
- § Internationale Events „Replication Games”
- § Vor allem Ökonomie und Politikwissenschaften <https://journals.sagepub.com/doi/10.1177/20531680241233439>
- § Ergebnisse von 110 Reproduktionen „Mass reproducibility and replicability: a new hope” <https://econpapers.repec.org/paper/zbwi4rdps/>
- o Open Code wird seitens Zeitschriften und zukünftig eventuell auch für den Abschluss von Datennutzungsverträgen mit Panels nötig (SOEP hat zB nur 20% open code, https://www.wifa.uni-leipzig.de/fileadmin/Fakultät_Wifa/Institut_für_Theoretische_Volkswirtschaftslehre/Professur_Makroökonomik/Economics_Research_Seminar/ERS-Paper_Marcus.pdf)

15.0.2 Robustness checks

- Multiverse analysis https://www.researchgate.net/publication/367204012_EEGManyPipelines_A_large-scale_grass-root_multi-analyst_study_of_EEG_analysis_practices_in_the_wild

15.0.2.1 Markdown Documents

-
- Computational Reproducibility Pilot Project: <https://psyarxiv.com/k8d4u/> | HIER AUCH DISKUTIERT ALS SERVICE (könnte ULB machen)
-

interactive multiverse / multiverse analysis

Statistik

- Bayesianische Statistik
 - Verbannung von NHST bei Basic and Applied Psychology, bisher mit negativen Auswirkungen, die die Probleme eventuell noch verschärfen (<https://www.tandfonline.com/doi/full/10.1080/00031305.2018.1537892#:~:text=Part%20of%20the%20BASP%20editors,their%20expository%20text%20with%20a>)
- #### Transparency Checklist

15.0.3 Open Data

- Abbruch einer Promotion, Wechsel der Institution, usw. können dazu führen, dass wertvolle Zeit & Arbeit verloren gehen; im Extremfall wurden Tiere aufgezogen, operiert, und Forschung kann nicht weitergeführt werden wegen schlechter Dokumentierung
 - Vor Durchführung der Studie sollte klar sein, wie die Daten aussehen werden (Variablen, Bedeutung, Umfang, usw.)
 - <https://www.dataad.org>
- falsifizierte Daten können nur durch Sichtung identifiziert werden <https://associationofanaesthetists-publications.onlinelibrary.wiley.com/doi/full/10.1111/anae.15263>
- Myths: https://www.sortee.org/blog/2024/04/12/2024_open_data_myths/
 - <https://leibniz-psychology.org/das-institut/drittmittelprojekte/datawiz-ii>
 - originalstudie hierzu,
 - Vanpaemel et al., 2015,
 - und das: https://www.researchgate.net/publication/370001752_Data_sharing_upon_request_and_statistical_consistency_errors_in_psychology_A_replication_of_Wicherts_Bakker_and_Molenaar_2011
 - Repositorien für daten
 - o Osf.io
 - o Psych archives zpid
 - o Re3
 - o Universitätsspezifische (z.B. datastore der Uni Münster)
 - o Fachspezifische (info auch hier: <https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/repositorien/>)

- § Economics & social sciences: data.gesis.org
- § Earth science and life science: Pangaea.de
- § Humanities, cultural studies: de.dariah.eu
- § Linguistics: www.clarin.eu
- § Biology: gfbio.org
- § Materialwissenschaften: nomad-lab.eu
- § Psychologie: Researchbox.org
- o Coscine <https://about.coscine.de>
- o Zenodo (interdisziplinär)
- o Git
- o Open Source Lösungen statt proprietärer Software, <https://www.mkw.nrw/hochschule-und-forschung/digitalisierung-hochschule-und-wissenschaft/forschungsdatenmanagement-fdm>
<https://www.nfdi.de>
- o <https://www.researchobject.org/ro-crate/>
- o <https://www.protocols.io>
- o <https://www.icpsr.umich.edu/web/pages/ICPSR/index.html>
- o Für Qualitative Daten: <https://qdr.syr.edu/deposit/data>
- Kriterien für Daten teilen: FAIR
- o Pyramid of requirements <https://content.iospress.com/articles/information-services-and-use/isu805>
- o Halbautomatische Extraktion von Meta-Daten <https://www.nfdi4chem.de/de/lister-halbautomatische-metadatenextraktion-aus-kommentierter-experimentdokumentation-in-elabftw/>
- o Datenstruktur: <https://psych-ds.github.io>
- o Fair Aware Tool zum Testen des Wissens zu FAIR <https://fairaware.dans.knaw.nl>
- o Ohne Fair kostet es Europäische Wirtschaft jedes Jahr 10.2 Milliarden Euro
https://publications.europa.eu/resource/cellar/d375368c-1a0a-11e9-8d04-01aa75ed71a1.0001.01/DOC_1
- o Feld-spezifische Templates für FAIRe daten: https://www.cos.io/blog/cedar-embeddable-editor?utm_source=linkedin&utm_medium=social
- CARE

- o Collective benefit: Ergebnisse sind auch von Gesellschaft verwendbar, Bürgerbeteiligung
- o Authority to control:
- o Responsibility: verschiedene Sprachenu und Weltanschauungen werden berücksichtigt
- o Ethics: Wertschätzung
 - Zeitschriften
- o JOPD (psychologie)
- o Ing.grid (ingenieurwissenschaften) <https://www.inggrid.org>
- o Weitere wichtige?
 - Beispiele für Datenbanken
- o <https://db.mocoda2.de/c/home> Chatverläufe
- o OpAQ Ankereffekte
- o Cooperation database
- o Stanford meta analysis aggregator
- o <http://metabus.org#>
- o Listen
- § Aaron Charlton Mktgo.org
- § <https://replications.clearerthinking.org/replications/>
- Teilen von qualitativen Daten, dort ist Anonymisierung relative schwierig: <https://doi.org/10.1177/25152459231205832>
- Data parasites / data police / data terrorists
- Nachfragen: https://www.whatdotheyknow.com/request/trial_protocols_behavioural_insi_2
- Mehr Zitate? <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0230416>
- Workshop-Unterlagen: <https://j-5chneider.github.io/PTOS-open-data/>
- Infosammlung zu Forschungsdaten: <https://github.com/UB-Mannheim/awesome-RDM>
- Daten auch wertvoll als Gedächtnis der Gesellschaft, soetwas sollte archiviert werden, z.B. <https://liebesbriefarchiv.de>
- o Daten von Personen aus allen Feldern der Gesellschaft (crowdsourced) und öffentlich verfügbar

- o Andere Beispiele für Citizen Science: <https://www.buergerschaftenwissen.de> und <https://wissenschaft-im-dialog.de/mitmachen/>
 - Langzeitarchivierung vs. Archivierung
 - o 1000 Jahre <https://archiveprogram.github.com>
- Daten automatisch anonymisieren lassen mit [AMNESIA](#)

15.0.4 Replicability

Assuring Replicability in Primary Research

1. Replikationsstudien in Sozialpsychologie schwierig wegen Anreizstruktur und Ressourcen
<https://doi.org/10.1027/1864-9335/a000548>
2. Andere Felder, Soziologie, Präregistrierung: https://www.researchgate.net/publication/368454140_Preregistration_and_Registered_Reports_in_Sociology_Strengths_Weaknesses_and_Other_Considerations
- 3.
4. Irise project (EU-gefördertes Projekt): <https://irise-project.eu/research-outputs>
- 5.
6. Messungenauigkeit: <https://osf.io/2me7t>
- 7.

15.0.4.1 Preregistration / Präregistrierung

- Erklärung
- Gütekriterien: Struktur + Vollständigkeit (simmons nelson vs. Pham consumer psych)
- Analyseplan wichtig, um wirklich P-Hacking vorzubeugen: <https://www.journals.uchicago.edu/doi/abs/10.1086/730455>
- https://www.researchgate.net/publication/375575020_Preregistration_in_practice_A_comparison_of_preregistered_and_non-preregistered_studies_in_psychology
- High replicability neues project <https://www.nature.com/articles/s41562-023-01749-9>
- Prävalenz: <https://datacolada.org/115>

- Es sollte nicht nur über Bereitschaft von Forschenden gehen, sondern auch im Peer Review verankert werden, sauber zu präregistrieren: <https://journal.trialanderror.org/pub/reflections-on-preregistration/release/2>
 - Deviations:
 - o Alle weichen ab: <https://osf.io/preprints/psyarxiv/nj4es> , <https://osf.io/f2z7y> slide 12/21
 - o niemand prüft auf Abweichungen: <https://osf.io/preprints/psyarxiv/nh7qw>
 - o Transparent changes: <https://osf.io/6fk87>
 - o How to deviate: <https://osf.io/preprints/psyarxiv/ha29k>
 - Repositorien
 - o <https://meta-meta-resources.org/running-studies/preparation/pre-reg-repos/#!>
 - o <https://www.alltrials.net>
 - o <https://clinicaltrials.gov>
 - o <http://www.crd.york.ac.uk/PROSPERO/>
 - Auswirkung auf Ergebnisse
 - o Wahrscheinlichkeit positiven Befunden statt >90% dadurch nur 40-50%, Scheel, Schijen, & Lakens(2021) [siehe auch <https://osf.io/f2z7y>, slide 17/21]
 - o <https://drive.google.com/file/d/1gcyBE78tb9zerl4M35uS3npVGvMp-MPZ/view> The effectiveness of preregistration in psychology: Assessing preregistration strictness and preregistration-study consistency olmo van den akker[LR2]
- ##### Pre-Registration templates what is a good prereg? (see Simmons et al) What to preregister? Analysis script
- OSF arbeitet neue ein <https://www.cos.io/blog/call-for-submission-for-preregistration-templates>
- - Diskrepanz zwischen Präregistrierung/Prereg und Publikation: <https://bmjopen.bmj.com/content/13/10/e076264.long>
 -
 - Kritik: <https://journals.sagepub.com/doi/10.1037/gpr0000135>
 -

- Prereg als Mittel zur Transparenz in der Forschungsplanung auch für Exploration sinnvoll, bei Hypothesentesten v.a. Eliminierung von Freiheitsgraden und Erhöhung von Vertrauen
-

15.0.4.2 Internal Replications

<https://perception.yale.edu/papers/17-Scholl-APSObsServer.pdf>

15.0.4.3 Power Analysis

- Power
- Small-telescopes approach
- Bayesianischer Ansatz
 - o So lange erheben, bis Bayes-Faktor konvergiert; kann Stichprobeumfang reduzieren, wurde so zB für Verhaltensforschung bei Tieren empfohlen (<https://www.nature.com/articles/s41684-023-01308-9>)

15.0.4.4 New statistics

- Frequentistische Statistik wird benutzt, ist oft missverstanden und Inhalte sind mit anderen Perspektiven vermischt (<https://journals.sagepub.com/doi/10.1177/0959354314546157?icid=int.sj-full-text.similar-articles.3>)
 - o Gigerenzer P-Value unwissen
- Bayesian, Kritik: <https://journals.sagepub.com/doi/10.1177/25152459231213371>
- Effektstärken statt p-Werte
 - o Equivalence testing
 - o Rölle von kleinen Effekten: <https://journals.sagepub.com/doi/full/10.1177/17456916221100420>
 - o practical relevance of small effect sizes https://www.researchgate.net/publication/352412241_Evaluating_the_practical_relevance_of_observed_effect_sizes_in_psychological_research
 - o effect sizes guide
 - o
- Alternative lakens paper “alternative to p-value is correctly used p-value”

- o Philosophische Diskussion über Nutzen von P-Werten: <https://journals.sagepub.com/doi/full/10.1177/09593543231160112> und hier <https://journals.sagepub.com/doi/10.1177/0959354312465483?icid=int.sj-full-text.similar-articles.1>

15.0.4.5 21 Word Solution

15.0.4.6 Open Materials

15.0.4.7 Transparency

- Bewertungskriterien: <https://osf.io/preprints/psyarxiv/djmcq>
- Transparency Checklist

15.0.5 Qualitative Forschung und Open Science

- Pseudonymisierung von Daten ist hier sehr besonders, denn...: <https://doi.org/10.1177/160940692110346>
- Präregistrierungs-Template: <https://docs.google.com/document/d/16vLoRAs6RmXKy0v1RCx6osSaYk4KtWbeM/edit#heading=h.fwbi14d4b65g>
- Sharing sensitive qualitative data: <https://www.youtube.com/watch?v=-Pha3EINF3s>
- o Berichte von Opfern von Vergewaltigungen à müssen anonymisiert werden, extrem aufwändig, kann stellvertretene Traumata auslösen

15.0.6 Open Source Software

- Research Software Engineering (RSE) [siehe auch oben bei reproducible code, aber hier sollte das meiste dazu stehen]
- Non proprietär à Gefahr vor Lock-in
- GNU Lizenz
- Open Source Bedeutung
- Bibliotheken, die Wissen verwalten, positionieren sich neu, um nun Infrastruktur zu stellen (z.B. <https://www.cambridge.org/core/books/abs/data-science-in-the-library/toward-reproducibility-academic-libraries-and-open-science/6FB18D7BCD8B82E597ED0D4B390C72EE#>)
- Von Chan-Zuckerberg-Initiative Unternehmen: <https://chanzuckerberg.com/rfa/essential-open-source-software-for-science/>
- Alternative zu Clarivate Analytics Web of Science Literaturdatenbank: <https://openalex.org> <https://arxiv.org/abs/2205.01833>

Infrastruktur	Kommerzielle Software	Open Source Software
Literaturdatenbank		https://openalex.org
Literaturverwaltung	Citavi, Mendeley	Zotero
Datenerhebung	Unipark, Millisecond Inquisit	PsychoPy
Datenanalyse	IBM SPSS, Stata	GNU R, Python, Jupyter Notebook
Begutachtung und Veröf- fentlichung	Editorial Manager	Open Journal System

15.0.7 Slow Science

Viele der Entwicklungen lassen sich als „langsame Wissenschaft“ (oder engl. Slow Science) zusammenfassen. Der traditionellen Produktion von Forschungsartikeln steht die achtsame Auseinandersetzung entgegen. Dass viele der vorgeschlagenen Lösungen im Wettbewerb um veröffentlichte Forschungsartikel nachteilig sind, kritisiert beispielsweise Hyman (2024 https://www.researchgate.net/publication/377965519_Freeing_social_and_medical_scientists_from_the_replication_crisis). Er schlägt stattdessen vor, die Probleme mittels künstlicher Intelligenz zu lösen. Verlage wie Elsevier benutzen diese beispielsweise bereits für Peer Review (<https://www.elsevier.com/about/policies-and-standards/the-use-of-generative-ai-and-ai-assisted-technologies-in-the-review-process>), wenngleich Modelle wie ChatGPT 4.0 nicht dafür geeignet sind <https://arxiv.org/abs/2402.05519#:~:text=Practical%20implications%3A%20Overall%2C%20Cha>).

15.0.8 Big Team Science

- CRediT und Übersetzungen <https://github.com/contributorshipcollaboration/credit-translation>

[LR1]<https://journals.sagepub.com/doi/10.1177/20531680241233439>

[LR2]viele paper von denen dazu

[LR3]<https://arxiv.org/abs/2205.01833>

[LR4]https://www.researchgate.net/publication/377965519_Freeing_social_and_medical_scientists_from_the_replication_crisis

[LR5][https://arxiv.org/abs/2402.05519#:~:text=Practical%20implications%3A%20Overall%2C%20ChatGPT%](https://arxiv.org/abs/2402.05519#:~:text=Practical%20implications%3A%20Overall%2C%20ChatGPT%20)

15.0.9 Literatur

16 Theorien

Theorien übernehmen in der Wissenschaft oft die Funktion eines Wegweisers. Durch sie lässt sich einschätzen, wie die Ergebnisse einer Untersuchung aussehen, bevor sie durchgeführt werden. Falls die Ergebnisse also nicht wie erwartet sind, zum Beispiel wenn eine Replikation fehlschlägt, besteht der Verdacht, die Theorie könnte falsch oder unvollständig sein. In der Psychologie, in welcher Theorien häufig in Alltagssprache und nicht formalisiert (also als mathematische Formeln) notiert werden, ist der Sinn von formalen Varianten schon lange diskutiert (Meehl 2004; Platt 1964; Wilcox and Rousselet 2024) und Replikationsfehlschläge werden darauf zurückgeführt, dass die Theorien zu unpräzise, missverständlich, und nicht gut ausgearbeitet sind. Es herrscht das Motto, Theorien seien wie Zahnbürsten und man solle niemals die von jemand anderem nehmen. Daran, dass das Problem seit Jahrzehnten beschrieben wird, hat sich wenig geändert (Lakens 2023; Muthukrishna and Henrich 2019). Inzwischen gibt es jedoch Bemühungen, das Problem auch zu lösen (Sion and Earp 2024; Erkan and Devezer 2023), beispielsweise durch das herausstellen der Vorteile von bekannten formalen Theorien (Robinaugh et al. 2021), Entwicklung von Werkzeugen zur Formalisierung von Theorien (<https://www.theorymaps.org>), oder durch einzelne Forschende. Der Wert von Replikationsstudien wird im Rahmen der „Theorie-Krise“ darin gesehen, dass nur durch wiederholte Untersuchung festgestellt werden kann, unter welchen Bedingungen ein bestimmtes Phänomen nachweisbar ist (Erkan and Devezer 2023).

16.0.1 Spezifizierung und Formalisierung

Das Problem einer Theorie in Alltagssprache lässt sich wie folgt veranschaulichen: Wagenmakers et al. (2016) versuchten, eine Studie zur Facial Feedback Hypothese zu replizieren. Die Hypothese besagt, dass Personen, wenn sie einen Stift so im Mund halten, dass dabei diejenigen Muskeln beansprucht werden, die auch beim Grinsen genutzt werden, Dinge lustiger finden, als wenn sie den Stift anders im Mund halten. Damit geprüft werden konnte, ob Versuchspersonen ihre Stifte richtig in den Mund nahmen, wurden sie in der Replikationsstudie gefilmt. Das war in der Originalstudie nicht der Fall. Die Replikationsstudien schlugen fehl und der Unterschied der Kamera wurde von den Autoren der Originalstudie als Grund genannt, weshalb die Replikation fehlschlug (Strack 2016). Ob und wie die Präsenz einer Kamera in dem Versuchsaufbau relevant ist, war nicht Teil der Theorie beziehungsweise Hypothese. Durch die Formulierung der Theorie in gewöhnlicher Sprache können unendlich viele solcher „Post Hoc“ Erklärungen in die Diskussion eingebracht werden, die jegliche Replikationsfehlschläge relativieren. Bei einer formalen Theorie sind solche Missverständnisse seltener.

Aus dem Problem, dass Theorien schwer verständlich und wenig informativ sind, folgt die Tatsache, dass die meisten Theorien nie falsifiziert werden, aber auch nicht weiter untersucht werden („Zombie-Theorien“, Ferguson and Heene (2012)). Ähnlich verhält es sich mit psychologischen Skalen, bei denen die meisten nur einmal verwendet werden (Elson et al. 2023). Flexible Theorien vereinfachen außerdem das p-Hacking: In der Theorie ist keine Vorgabe zum Aufbereiten der Daten, also kann diejenige Variante gewählt werden, die die Theorie bestätigt (Van den Akker et al. 2022).

16.0.2 Gegnerische Kollaborationen

Ein Ansatz, bei dem verhindert werden soll, dass uninformative Studien durchgeführt werden und sich Diskussionen auf Verstehensweisen von Theorien fokussieren, sind gegnerische Kollaborationen (adversarial collaborations, Cleeremans (2022)). Dabei arbeiten Vertreter*innen von zwei inkompatiblen Standpunkten miteinander und erstellen beispielsweise eine Studie, die den Konflikt beilegen soll (Cowan et al. 2020; Mellers, Hertwig, and Kahneman 2001).

16.0.2.1 Weiterführende Informationen

- Trafimow and Earp (2016) beschreiben, wie Replikationen auch ohne Theorie möglich sind.

16.0.3 Literatur

17 Die Welt

Nachdem nun systemische, methodische, und theoretische Gründe für die Replikationskrise diskutiert wurden, endet die Diskussion mit einem allgemeinen und ontologischen Punkt: Die Erwartung, dass sich Befunde wiederholt feststellen lassen müssen beruht auf der Annahme oder Theorie, dass die Welt – oder zumindest die jeweils in der Wissenschaft untersuchte Gesetzmäßigkeit – stabil ist. Bei fluktuierenden Mustern in der Welt oder im menschlichen Verhalten sind Replikationsfehlschläge wenig überraschend. Für die Psychologie hat beispielsweise Smedslund (2015) kritisiert, dass diese Annahme falsch ist und hat diese *Historizität* als einen der Gründe angeführt, weshalb Psychologie keine Naturwissenschaft sein kann. Ähnlich argumentierte ebenfalls Gergen (1973). Beide Sichtweisen haben ihren Ursprung möglicherweise in der Grundlagendiskussion der Psychologie: Anfang des 20. Jahrhunderts, als die Psychologie sich langsam von der Philosophie distanzierte und vermehrt empirisch (also durch Beobachtungen statt nur Überlegungen) arbeitete, unterteilte Dilthey and Riedel (1970) Naturwissenschaften und Geisteswissenschaften dahingehend, dass in es in den Naturwissenschaften nicht um das Nacherleben und Begreifen nachrangig sind und es um abstrakte und allgemeine (*nomothetische*) Gesetzmäßigkeiten geht. Geisteswissenschaften hingegen betrachten Vorgeänge in einem konkreten und einzigartigen (*idiographischen*) Zusammenhang mit allen dazugehörigen Besonderheiten. Es etablierte sich langfristig der naturwissenschaftliche Ansatz, der bereits 50 Jahre zuvor fruchtbare Erkenntnisse lieferte (Fechner 1860).

Eine besondere Rolle beim Verallgemeinern von einzelnen Beobachtungen auf allgemeingültige Gesetzmäßigkeiten spielt auch der Zufall. Das Besondere bei statistischen Methoden ist beispielsweise, dass die Unsicherheit als Zahl ausgedrückt wird und damit *quantifiziert* wird. Die wichtige Frage im Rahmen von Replikationsproblemen ist, ob die Unterschiede zwischen Befunden zufällig oder aufgrund irgendeiner Unterschiede entstanden sind. Diese zwei Arten des Zufalls werden als ontologischer und epistemischer Zufall bezeichnet. Ontologischer Zufall bedeutet, dass ein einfach keinen Grund gibt, während wir ihn beim epistemologischen Zufall bloß nicht kennen. Der Ausgang bei einem Münzwurf, also ob die Münze auf Kopf oder Zahl landet, ist beispielsweise ein Fall epistemischen Zufalls: Würden wir alle wichtigen Parameter kennen (Höhe der Hand, Stärke des Wurfes, welche Seite beim Werfen oben liegt, usw.), könnten wir eindeutig vorhersagen, wie die Münze landet. Übrigens landen faire Münzen meistens auf der Seite, mit der sie gestartet sind (Bartoš et al. 2023).

17.0.1 Literatur

Part VIII

Fazit und Ausblick

- AUFLISTEN / TABELLE: Gründe für „nicht repliziert“
- Neue Wissenschaft: Meta-Wissenschaft, die mit sozialwissenschaftlichen Methoden arbeitet
- Höherer Grad an Systematizität (Hoyningen-Huene)
- Selbstkorrektur von Wissenschaft als aktive Arbeit statt Invisible Hand

Was hat sich verbessert?

- Review der Veränderungen https://www.researchgate.net/publication/355421853_Replicability_Robustness_and_Reproducibility_in_Psychological_Science
- journal guidelines und reporting: https://www.researchgate.net/publication/364374923_The_influence_o
- <https://www.nature.com/articles/s41562-023-01749-9> extreme Erhöhung von Replikationsrate durch neue Methoden in kleiner Stichprobe von Phänomenen
- <https://www.nature.com/articles/s44271-023-00003-2> positive structural, procedural, and community changes
- Neue Meinung zu OS Praktiken: https://www.researchgate.net/publication/373685253_Survey_of_open_science_practices_and_attitudes_in_the_social_sciences
- Meta-Wissenschaftliche Evaluation von neuen Methoden
 - o Bisherige Vorschläge sind zum Teil intuitiv und nicht ausführlich geprüft <https://journals.sagepub.com/doi/10.1177/10892680241235120>; manche der Intuitionen sind auch falsch (z.B. wichtig ist nur, was replizierbar ist)
 - o Initial evidence registered reports
 - o Nosek 2023 paper (protzko?)
 - o Hilft Preregistration gegen p-hacking? <https://ideas.repec.org/p/zbw/i4rdps/101.html>
- Strikte Normen könnten Flexibilität verringern und Offenheit und Innovationspotenziale einschränken
 - o Anfänglich unplausible Hypothesen sollten nicht direkt verworfen werden, manchmal führt auch eine Beharrungstendenz von Wissenschaftler*innen dazu, dass sich neue, anfänglich unplausible Theorien, durchsetzen (<https://journals.sagepub.com/doi/10.1177/10892680241235120>)

Was ist noch zu tun?

Wissenschaft zeigt, dass Selbstkorrekturprozesse existieren, Replikationskrise ist allerdings noch nicht annähernd überwunden

- Lösungen für Strukturelle Probleme erst am Anfang der Umsetzung
- Open Access Problematik aktuell mit Momentum, allerdings unklar, ob weitere Jahrzehnte zur Lösung benötigt werden
- Änderungsvorschläge müssen wissenschaftlich geprüft werden
 - o Interventionen oft sinnvoll, manchmal aufwändig, aber oft ist unklar, wie viel sie bringen (z.B. Präregistrierungen)
 - o Selbst Artikel, die sich mit Open Science befassen, halten sich selbst manchmal nicht an die Standards (z.B. REF[LR1])
- Replikationsraten (Replikationen pro Publikationen) maximal 5%, eher <1%
 - o <https://twitter.com/giladfeldman/status/1735950123291779119>
 - o <https://osf.io/preprints/psyarxiv/sa6rc> in Psychologie „54 out of 88 journals did not publish any direct replication articles in the eleven years surveyed” (Abstract)
- Interdisziplinäres Problem erfordert Zusammenarbeit
 - o Replikationsdefinitionen wurden zB in Psychologie (Hüffmeier et al.), Digital Humanities (Repetitive Research) und Marketing (<https://doi.org/10.1093/jcr/ucae007>, Table 1) erstellt, sind aber völlig redundant

Policy pyramide[LR2] ?

- Offene thematische Punkte
 - o Wie sieht Replizierbarkeit jenseits von Sozialwissenschaften aus? (cognitive, behavioral, social)
 - o Welche alternativen Anforderungen bestehen in logisch-formalen und qualitativ/hermeneutischen Disziplinen?
 - o Credibility beyond replicability https://www.researchgate.net/publication/359766772_Credibility_Beyond_Replicability_Improving_the_Four_Validities_in_Psychological_Science
 - o Welche Rolle spielt Open Science bei sich verschließenden Nationen und im Krieg

§ Positivistisch: Abtrennung

§ Relationistisch: Die anderen haben eh ein anderes Bild, ihnen hilft das nicht

§ Altruistisch/Idealistisch: Wissenschaft soll der Menschheit helfen und nicht diskriminieren

- § Nationalistisch: Gesellschaft bezahlt Wissenschaft, sie muss auch primär davon profitieren
- Infrastructure (make it possible)
 - o Offenheit des gesamten Wissenschaftsprozesses fördern
<https://doi.org/10.1371/journal.pbio.3002362>
 - o Deal-Gruppe ist damit, Ausbeutung von Wissenschaft durch Großverlage zu stoppen, gescheitert (<https://www.faz.net/aktuell/wissen/forschung-politik/wissenschaftliche-grossverlage-wissenschaftliche-grossverlage-19344181.html>)
 - o
 - User interface / experience (make it easy)
 - o Viele OS Lösungen sehr einfach
 - o Open Source Software z.T. sehr schwierig (Literaturverwaltung, Statistik, Publikationssysteme erfordern Eingabe via klobigem Interface)
 - Communities (make it normative)
 - o Wissenschaftler*innen, die angesehene Kolleg*innen kritisieren werden oft verklagt (z.B. Datacolada oder Elisabeth Bik, <https://www.buzzfeednews.com/article/stephaniemlee/elisabeth-bik-didier-raoult-hydroxychloroquine-study>)
 - o
 - Incentives (make it rewarding)
 - o Forschende und Bibliotheken stehen immernoch in einer Art Gefangenendilemma und wer sich zuerst bewegt, hat einen Nachteil
 - o Impact Factors und Berufungskommissionen immernoch verrückt nach Zitationen: <https://elsevier.digitalcommonsdata.com/datasets/btchxktzyw/6>
 - o Index jacking (<https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/asi.24855>)
 - o
 - Policy (make it required)
 - o TOP guidelines als Norm vielerorts noch nicht durchgesetzt
 - o Open Science ist noch nicht die Norm, wie jeder dazu beitragen kann, wird hier zusammengefasst <https://doi.org/10.7554/eLife.89736>
 - o Neue Probleme wie Open Washing (<https://epdos.nl>) oder vermeintliche Transformationsverträge von Verlagen mit Universitätsbibliotheken, bei denen das Ziel ist, sich langfristig von Verlagen unabhängig zu machen, gleichzeitig aber keine Transformation stattfindet sondern Folgeverträge vereinbart werden
 - o

Was geschieht jetzt gerade?

This is not final. Netherlands reports of scientific misconduct Stay up to date

- pschyh methods discussion group
- open science initiatives (Germany: NOSI)
- Retractionwatch blog
- FORRT
- Institutionalisierung von Open Science im Rahmen von dafür zuständigen Personen in Universitäten (MüCOS!)

Literaturempfehlungen

- Chris Chambers Buch <https://doi.org/10.2307/j.ctvc779w5>
- Charlotte Pennington Buch, hier ein Kapitel: <https://osf.io/preprints/psyarxiv/2tqep>
- Matters of Significance <https://discovery.ucl.ac.uk/id/eprint/10186829/1/Matters-of-Significance.pdf>

[LR1]https://ideas.repec.org/p/stg/wpaper/2020_01.html

[LR2]

- Abduh, Akira J. 2023. “A Critical Analysis of the World’s Top 2% Most Influential Scientists: Examining the Limitations and Biases of Highly Cited Researchers Lists.”
- Aczel, Balazs, Barnabas Szaszi, and Alex O Holcombe. 2021. “A Billion-Dollar Donation: Estimating the Cost of Researchers’ Time Spent on Peer Review.” *Research Integrity and Peer Review* 6: 1–8.
- Albert, Hans. 2010. *Traktat über Kritische Vernunft*. Nachdr. d. 5., verb. und erw. Aufl. Vol. 1609. UTB. Tübingen: Mohr Siebeck.
- Alogna, V. K., M. K. Attaya, P. Aucoin, Štěpán Bahník, S. Birch, A. R. Birt, B. H. Bornstein, et al. 2014. “Registered Replication Report: Schooler and Engstler-Schooler (1990).” *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 9 (5): 556–78. <https://doi.org/10.1177/1745691614545653>.
- Azevedo, Flavio, Sam Parsons, Leticia Micheli, Julia Feld Strand, Eike Mark Rinke, Samuel Guay, Mahmoud Medhat Elsherif, et al. 2019. “Introducing a Framework for Open and Reproducible Research Training (FORRT).”
- Bakker, Marjan, Annette van Dijk, and Jelte M. Wicherts. 2012. “The Rules of the Game Called Psychological Science.” *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 7 (6): 543–54. <https://doi.org/10.1177/1745691612459060>.

- Banker, Sachin, Sarah E. Ainsworth, Roy F. Baumeister, Dan Ariely, and Kathleen D. Vohs. 2017. "The Sticky Anchor Hypothesis: Ego Depletion Increases Susceptibility to Situational Cues." *Journal of Behavioral Decision Making* 87 (1): 23. <https://doi.org/10.1002/bdm.2022>.
- Bargh, John A, Mark Chen, and Lara Burrows. 1996. "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action." *Journal of Personality and Social Psychology* 71 (2): 230.
- Bartoš, František, Alexandra Sarafoglou, Henrik R Godmann, Amir Sahrani, David Klein Leunk, Pierre Y Gui, David Voss, et al. 2023. "Fair Coins Tend to Land on the Same Side They Started: Evidence from 350,757 Flips."
- Baumeister, Roy F., and Kathleen D. Vohs. 2016. "Misguided Effort with Elusive Implications." *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 11 (4): 574–75. <https://doi.org/10.1177/1745691616652878>.
- Bem, Daryl J. 1967. "Self-Perception: An Alternative Interpretation of Cognitive Dissonance Phenomena." *Psychological Review* 74 (3): 183–200. <https://doi.org/10.1037/h0024835>.
- . 2011. "Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect." *Journal of Personality and Social Psychology* 100 (3): 407–25. <https://doi.org/10.1037/a0021524>.
- Bik, Elisabeth M., Arturo Casadevall, and Ferric C. Fang. 2016. "The Prevalence of Inappropriate Image Duplication in Biomedical Research Publications." *mBio* 7 (3). <https://doi.org/10.1128/mbio.00809-16>.
- Bol, Thijs. 2023. "Gender Inequality in Cum Laude Distinctions for PhD Students." *Sci. Rep.* 13 (1): 20267.
- Borgstede, Matthias, and Marcel Scholz. 2021. "Quantitative and Qualitative Approaches to Generalization and Replication—a Representationalist View." *Frontiers in Psychology* 12: 605191. <https://doi.org/10.3389/fpsyg.2021.605191>.
- Bouwmeester, S., P. P. J. L. Verkoeijen, B. Aczel, F. Barbosa, L. Bègue, P. Brañas-Garza, T. G. H. Chmura, et al. 2017. "Registered Replication Report: Rand, Greene, and Nowak (2012)." *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 12 (3): 527–42. <https://doi.org/10.1177/1745691617693624>.
- Boyce, Veronica, Maya Mathur, and Michael C. Frank. 2023. "Eleven Years of Student Replication Projects Provide Evidence on the Correlates of Replicability in Psychology." *Royal Society Open Science* 10 (11): 231240. <https://doi.org/10.1098/rsos.231240>.
- Brembs, Björn. 2018. "Prestigious Science Journals Struggle to Reach Even Average Reliability." *Frontiers in Human Neuroscience* 12: 37. <https://doi.org/10.3389/fnhum.2018.00037>.
- Brembs, Björn, Philippe Huneman, Felix Schönbrodt, Gustav Nilsson, Toma Susi, Renke Siems, Pandelis Perakakis, Varvara Trachana, Lai Ma, and Sara Rodriguez-Cuadrado. 2023. "Replacing Academic Journals." *Royal Society Open Science* 10 (7). <https://doi.org/10.1098/rsos.230206>.
- Breznau, Nate, Eike Mark Rinke, Alexander Wuttke, Hung H. V. Nguyen, Muna Adem, Jule Adriaans, Amalia Alvarez-Benjumea, et al. 2022. "Observing Many Researchers Using the Same Data and Hypothesis Reveals a Hidden Universe of Uncertainty." *Proceedings of the National Academy of Sciences of the United States of America* 119 (44): e2203150119.

- <https://doi.org/10.1073/pnas.2203150119>.
- Buttlere, Brett. 2024. “Was This Registered Report Pilot Tested? Examination of Vaidis, Slegers, van Leeuwen, DeMarree, ... & Priolo, d. (2024).” *PsyArXiv*.
- Buzbas, Erkan O, and Berna Devezer. 2023. “Tension Between Theory and Practice of Replication.” *Journal of Trial & Error* 4 (1).
- Byers-Heinlein, Krista, Christina Bergmann, Catherine Davies, Michael C Frank, J Kiley Hamlin, Melissa Kline, Jonathan F Kominsky, et al. 2020. “Building a Collaborative Psychological Science: Lessons Learned from ManyBabies 1.” *Canadian Psychology/Psychologie Canadienne* 61 (4): 349.
- Byrne, Jennifer A., and Jana Christopher. 2020. “Digital Magic, or the Dark Arts of the 21st Century-How Can Journals and Peer Reviewers Detect Manuscripts and Publications from Paper Mills?” *FEBS Letters* 594 (4): 583–89. <https://doi.org/10.1002/1873-3468.13747>.
- Cabanac, Guillaume, Cyril Labbé, and Alexander Magazinov. 2021. “Tortured Phrases: A Dubious Writing Style Emerging in Science. Evidence of Critical Issues Affecting Established Journals.” <https://doi.org/10.48550/arXiv.2107.06751>.
- Calder, Bobby J., Lynn W. Phillips, and Alice M. Tybout. 1981. “Designing Research for Application.” *Journal of Consumer Research* 8 (2): 197. <https://doi.org/10.1086/208856>.
- Camerer, Colin F., Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al. 2018. “Evaluating the Replicability of Social Science Experiments in Nature and Science Between 2010 and 2015.” *Nature Human Behaviour* 2 (9): 637–44. <https://doi.org/10.1038/s41562-018-0399-z>.
- Carlsson, Rickard, Henrik Danielsson, Moritz Heene, Åse Innes-Ker, Daniël Lakens, Ulrich Schimmack, Felix D. Schönbrodt, Marcel van Assen, and Yana Weinstein. 2017. “Inaugural Editorial of Meta-Psychology.” *Meta-Psychology* 1: a1001. <https://doi.org/10.15626/MP2017.1001>.
- Carriquiry, Alicia L, Michael J Daniels, and Nancy Reid. 2023. “Editorial: Special Issue on Reproducibility and Replicability.” *Stat. Sci.* 38 (4).
- Carter, Evan C., Felix D. Schönbrodt, Will M. Gervais, and Joseph Hilgard. 2019. “Correcting for Bias in Psychology: A Comparison of Meta-Analytic Methods.” *Advances in Methods and Practices in Psychological Science* 2 (2): 115–44. <https://doi.org/10.1177/2515245919847196>.
- Cesario, Joseph. 2014. “Priming, Replication, and the Hardest Science.” *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 9 (1): 40–48. <https://doi.org/10.1177/1745691613513470>.
- Chambers, Christopher D, and Loukia Tzavella. 2022. “The Past, Present and Future of Registered Reports.” *Nature Human Behaviour* 6 (1): 29–42.
- Charlton, Aaron. 2022. “Replications of Marketing Studies.” OpenMKT.org. <https://openmkt.org/research/replications-of-marketing-studies/>.
- Cheung, Irene, Lorne Campbell, Etienne P. LeBel, Robert A. Ackerman, Bülent Aykutog˘lu, Štěpán Bahník, Jeffrey D. Bowen, et al. 2016. “Registered Replication Report: Study 1 from Finkel, Rusbult, Kumashiro, & Hannon (2002).” *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 11 (5): 750–64. <https://doi.org/10.1177/1745691616664694>.

- Choi, Hong Hui. 2023. "In Defense of the Resampling Account of Replication." *Journal of Theoretical and Philosophical Psychology* 43 (4): 249–51. <https://doi.org/10.1037/teo0000224>.
- Cleeremans, Axel. 2022. "Theory as Adversarial Collaboration." *Nature Human Behaviour* 6 (4): 485–86.
- Coles, Nicholas A., David S. March, Fernando Marmolejo-Ramos, Jeff T. Larsen, Nwadiogo C. Arinze, Izuchukwu L. G. Ndukaihe, Megan L. Willis, et al. 2022. "A Multi-Lab Test of the Facial Feedback Hypothesis by the Many Smiles Collaboration." *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-022-01458-9>.
- Coretta, Stefano, Joseph V. Casillas, Simon Roessig, Michael Franke, Byron Ahn, Ali H. Al-Hoorie, Jalal Al-Tamimi, et al. 2023. "Multidimensional Signals and Analytic Flexibility: Estimating Degrees of Freedom in Human-Speech Analyses." *Advances in Methods and Practices in Psychological Science* 6 (3). <https://doi.org/10.1177/25152459231162567>.
- Cowan, Nelson, Clément Belletier, Jason M. Doherty, Agnieszka J. Jaroslawska, Stephen Rhodes, Alicia Forsberg, Moshe Naveh-Benjamin, Pierre Barrouillet, Valérie Camos, and Robert H. Logie. 2020. "How Do Scientific Views Change? Notes from an Extended Adversarial Collaboration." *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 15 (4): 1011–25. <https://doi.org/10.1177/1745691620906415>.
- Cronbach, Lee J., Richard E. Snow, and David E. Wiley. 1991. *Improving Inquiry in Social Science: A Volume in Honor of Lee J. Cronbach*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Deutsche Forschungsgemeinschaft. 2022. "Open Science Als Teil Der Wissenschaftskultur. Positionierung Der Deutschen Forschungsgemeinschaft." Zenodo.
- Dijksterhuis, Ap. 2014. "Welcome Back Theory!" *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 9 (1): 72–75. <https://doi.org/10.1177/1745691613513472>.
- Dilthey, Wilhelm, and Manfred Riedel. 1970. *Der Aufbau Der Geschichtlichen Welt in Den Geisteswissenschaften*. Suhrkamp Frankfurt am Main.
- Doyen, Stéphane, Olivier Klein, Cora-Lise Pichon, and Axel Cleeremans. 2012. "Behavioral Priming: It's All in the Mind, but Whose Mind?" *PloS One* 7 (1): e29081. <https://doi.org/10.1371/journal.pone.0029081>.
- Easley, Richard W, and Charles S Madden. 2013. "Replication Revisited: Introduction to the Special Section on Replication in Business Research." *J. Bus. Res.* 66 (9): 1375–76.
- Eerland, Anita, Andrew M. Sherrill, Joseph P. Magliano, Rolf A. Zwaan, Jack D. Arnal, Philip Aucoin, Stephanie A. Berger, et al. 2016. "Registered Replication Report: Hart & Albarracín (2011)." *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 11 (1): 158–71. <https://doi.org/10.1177/1745691615605826>.
- Elsherif, Mahmoud Medhat, Sara Lil Middleton, Jenny Mai Phan, Flavio Azevedo, Bethan Joan Iley, Magdalena Grose-Hodge, Samantha Lily Tyler, et al. 2022. "Bridging Neurodiversity and Open Scholarship: How Shared Values Can Guide Best Practices for Research Integrity, Social Justice, and Principled Education." <https://doi.org/10.31222/osf.io/k7a9p>.
- Elson, Malte, Ian Hussey, Taym Alsalti, and Ruben C. Arslan. 2023. "Psychological Measures Aren't Toothbrushes." *Communications Psychology* 1 (1). <https://doi.org/10.1038/s44271->

023-00026-9.

- Ensinnck, Eline Noëlle Fleur, and Daniel Lakens. 2023. "An Inception Cohort Study Quantifying How Many Registered Studies Are Published." *PsyArXiv*.
- Erkan, Buzbas, and Berna Devezer. 2023. "Tension Between Theory and Practice of Replication." *J. Trial Error* 4 (1): 73–81.
- Etzel, Franka, Anna Seyffert-Müller, Felix D Schönbrodt, Lucie Kreuzer, Anne Gärtner, Paula Knischewski, and Daniel Leising. 2024. "Inter-Rater Reliability in Assessing the Methodological Quality of Research Papers in Psychology."
- Fanelli, Daniele. 2009. "How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data." *PloS One* 4 (5): e5738. <https://doi.org/10.1371/journal.pone.0005738>.
- Fanelli, Daniele, Matteo Schleicher, Ferric C Fang, Arturo Casadevall, and Elisabeth M Bik. 2022. "Do Individual and Institutional Predictors of Misconduct Vary by Country? Results of a Matched-Control Analysis of Problematic Image Duplications." *PLoS One* 17 (3): e0255334.
- Fanelli, Daniele, Pedro Batista Tan, Olavo Bohrer Amaral, and Kleber Neves. 2022. "A Metric of Knowledge as Information Compression Reflects Reproducibility Predictions in Biomedical Experiments."
- Fecher, Benedikt, and Sascha Friesike. 2014. *Open Science: One Term, Five Schools of Thought*. Springer International Publishing.
- Fechner, Gustav Theodor. 1860. *Elemente Der Psychophysik*. Leipzig: Breitkopf und Härtel.
- Feest, Uljana. 2019. "Why Replication Is Overrated." *Philosophy of Science* 86 (5): 895–905.
- Ferguson, Christopher J., and Moritz Heene. 2012. "A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null." *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 7 (6): 555–61. <https://doi.org/10.1177/1745691612459059>.
- Feyerabend, Paul K. 1975/2002. *Against Method*. Reprinted der 3. ed. 1993. London: Verso.
- Fiorentino, Carlos, and Carlos Montana Hoyos. 2014. "The Emerging Discipline of Biomimicry as a Design Paradigm Shift." *Int. J. Des. Objects* 8 (1): 1–15.
- Fisher, Zachary, Elizabeth Tipton, and Hou Zhipeng. 2017. "Robumeta: Robust Variance Meta-Regression." <https://CRAN.R-project.org/package=robumeta>.
- Fleck, Ludwik. 1935/2015. *Entstehung Und Entwicklung Einer Wissenschaftlichen Tatsache [Formation and Development of a Scientific Fact]: Einführung in Die Lehre Vom Denkstil Und Denkkollektiv [Introduction to Thinking Style and Thinking Collective]*. 10. Auflage. Vol. 312. Suhrkamp-Taschenbuch Wissenschaft. Frankfurt am Main: Suhrkamp.
- Forster, Nick, and Daniel W Lund. 2018. "Identifying and Dealing with Functional Psychopathic Behavior in Higher Education." *Glob. Bus. Organ. Excel.* 38 (1): 22–31.
- Francis, Zoë, Marina Milyavskaya, Hause Lin, and Michael Inzlicht. 2018. "Development of a Within-Subject, Repeated-Measures Ego-Depletion Paradigm." *Social Psychology* 49 (5): 271–86. <https://doi.org/10.1027/1864-9335/a000348>.
- Franzen, Delwen L., Benjamin Gregory Carlisle, Maia Salholz-Hillel, Nico Riedel, and Daniel Strech. 2023. "Institutional Dashboards on Clinical Trial Transparency for University Medical Centers: A Case Study." *PLoS Medicine* 20 (3): e1004175. <https://doi.org/10.1371/journal.pmed.1004175>.

[1371/journal.pmed.1004175](https://doi.org/10.1371/journal.pmed.1004175).

- Fraser, Nicholas, Liam Brierley, Gautam Dey, Jessica K Polka, Máté Pálffy, Federico Nanni, and Jonathon Alexis Coates. 2020. "Preprinting the COVID-19 Pandemic." *bioRxiv*. bioRxiv.
- Gärtner, Anne, Daniel Leising, and Felix Schönbrodt. 2022. "Responsible Research Assessment II: A Specific Proposal for Hiring and Promotion in Psychology."
- Gergen, Kenneth J. 1973. "Social Psychology as History." *Journal of Personality and Social Psychology* 26 (2): 309.
- Giner-Sorolla, Roger. 2012. "Science or Art? How Aesthetic Standards Grease the Way Through the Publication Bottleneck but Undermine Science." *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 7 (6): 562–71. <https://doi.org/10.1177/1745691612457576>.
- Glöckner, Andreas, and Tilmann Betsch. 2011. "The Empirical Content of Theories in Judgment and Decision Making: Shortcomings and Remedies." *Judgment and Decision Making* 6 (8): 711–21.
- Gopalakrishna, Gowri, Gerben ter Riet, Maarten J. L. F. Cruyff, Gerko Vink, Ineke Stoop, Jelte M. Wicherts, and Lex Bouter. 2021. "Prevalence of Questionable Research Practices, Research Misconduct and Their Potential Explanatory Factors: A Survey Among Academic Researchers in the Netherlands." <https://doi.org/10.31222/osf.io/vk9yt>.
- Gopalakrishna, Gowri, Jelte M. Wicherts, Gerko Vink, Ineke Stoop, Olmo van den Akker, Gerben ter Riet, and Lex Bouter. 2021. "Prevalence of Responsible Research Practices and Their Potential Explanatory Factors: A Survey Among Academic Researchers in the Netherlands." <https://doi.org/10.31222/osf.io/xsn94>.
- Gould, Elliot, Hannah Fraser, Timothy Parker, Shinichi Nakagawa, Simon Griffith, Peter Vesk, Fiona Fidler, et al. 2023. "Same Data, Different Analysts: Variation in Effect Sizes Due to Analytical Decisions in Ecology and Evolutionary Biology." <https://doi.org/10.32942/X2GG62>.
- Greenwald, Anthony G. 1976. "An Editorial." *Journal of Personality and Social Psychology* 33 (1): 1–7. <https://doi.org/10.1037/h0078635>.
- Hagger, Martin S., Nikos L. D. Chatzisarantis, Hugo Alberts, Calvin Octavianus Anggono, Cédric Batailler, Angela R. Birt, Ralf Brand, et al. 2016. "A Multilab Preregistered Replication of the Ego-Depletion Effect." *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 11 (4): 546–73. <https://doi.org/10.1177/1745691616652873>.
- Hanson, Mark A., Pablo Gómez Barreiro, Paolo Crosetto, and Dan Brockington. 2023. "The Strain on Scientific Publishing." <https://doi.org/10.48550/arXiv.2309.15884>.
- Hardwicke, Tom E, and Simine Vazire. 2023. "Transparency Is Now the Default at Psychological Science." *Psychol. Sci.*, December, 9567976231221573.
- Hedges, Larry V., and Jack L. Vevea. 1996. "Estimating Effect Size Under Publication Bias: Small Sample Properties and Robustness of a Random Effects Selection Model." *Journal of Educational and Behavioral Statistics* 21 (4): 299–332. <https://doi.org/10.3102/10769986021004299>.
- Hicks, Diana, Paul Wouters, Ludo Waltman, Sarah de Rijcke, and Ismael Rafols. 2015. "Bibliometrics: The Leiden Manifesto for Research Metrics." *Nature* 520 (7548): 429–31.

<https://doi.org/10.1038/520429a>.

- Hoebel, Merle, Ana Durglishvili, Johanna Reinold, and Daniel Leising. 2022. "Sexual Harassment and Coercion in German Academia: A Large-Scale Survey Study." *Sexual Offending: Theory, Research, and Prevention* 17. <https://doi.org/10.5964/sotrap.9349>.
- Hoyningen-Huene, Paul. 2013. *Systematicity: The Nature of Science*. Oxford Studies in Philosophy of Science. Oxford: Oxford Univ. Press.
- Hoyningen-Huene, Paul, and Harold Kincaid. 2023. "What Makes Economics Special: Orientational Paradigms." *J. Econ. Methodol.*, March, 1–15.
- Hüffmeier, Joachim, Jens Mazei, and Thomas Schultze. 2016. "Reconceptualizing Replication as a Sequence of Different Studies: A Replication Typology." *Journal of Experimental Social Psychology* 66: 81–92. <https://doi.org/10.1016/j.jesp.2015.09.009>.
- Hume, David. 1748/2011. *Eine Untersuchung über Den Menschlichen Verstand [an Enquiry Concerning Human Understanding]*. Vol. 5489. Reclams Universal-Bibliothek. Stuttgart: Reclam.
- Hussey, Ian, and Sean Hughes. 2018. *Hidden Invalidity Among Fifteen Commonly Used Measures in Social and Personality Psychology*. <https://doi.org/10.31234/osf.io/7rbfp>.
- Jaremka, Lisa M., Joshua M. Ackerman, Bertram Gawronski, Nicholas O. Rule, Kate Sweeny, Linda R. Tropp, Molly A. Metz, Ludwin Molina, William S. Ryan, and S. Brooke Vick. 2020. "Common Academic Experiences No One Talks about: Repeated Rejection, Impostor Syndrome, and Burnout." *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 15 (3): 519–43. <https://doi.org/10.1177/1745691619898848>.
- Jekel, Marc, Susann Fiedler, Ramona Allstadt Torras, Dorothee Mischkowski, Angela Rachael Dorrough, and Andreas Glöckner. 2020. "How to Teach Open Science Principles in the Undergraduate Curriculum—the Hagen Cumulative Science Project." *Psychology Learning & Teaching* 19 (1): 91–106. <https://doi.org/10.1177/1475725719868149>.
- John, Leslie K., George Loewenstein, and Drazen Prelec. 2012. "Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling." *Psychological Science* 23 (5): 524–32. <https://doi.org/10.1177/0956797611430953>.
- Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47 (2): 263. <https://doi.org/10.2307/1914185>.
- Kakinohana, Regis K., Ronaldo Pilati, and Richard A. Klein. 2022. "Does Anchoring Vary Across Cultures? Expanding the Many Labs Analysis." *European Journal of Social Psychology*. <https://doi.org/10.1002/ejsp.2924>.
- Kepes, Sven, Sheila K. Keener, Michael A. McDaniel, and Nathan S. Hartman. 2022. "Questionable Research Practices Among Researchers in the Most Research-Productive Management Programs." *Journal of Organizational Behavior*. <https://doi.org/10.1002/job.2623>.
- Klein, Richard Anthony, Kate A. Ratliff, Michelangelo Vianello, Reginald B. Adams, Štěpán Bahník, Michael J. Bernstein, Konrad Bocian, et al. 2014. "Investigating Variation in Replicability." *Social Psychology* 45 (3): 142–52. <https://doi.org/10.1027/1864-9335/a000178>.
- Klein, Richard Anthony, Michelangelo Vianello, Fred Hasselman, Byron Gregory Adams, Reginald B. Adams, Sinan Alper, Mark Aveyard, et al. 2018. "Many Labs 2: Investigating Variation in Replicability Across Sample and Setting." <https://doi.org/10.31234/osf.io/9654g>.
- Klonsky, E David. 2024. "Campbell's Law Explains the Replication Crisis: Pre-Registration

- Badges Are History Repeating.” *Assessment*, May, 10731911241253430.
- Korbmacher, Max, Flavio Azevedo, Charlotte R. Pennington, Helena Hartmann, Madeleine Pownall, Kathleen Schmidt, Mahmoud Elsherif, et al. 2023. “The Replication Crisis Has Led to Positive Structural, Procedural, and Community Changes.” *Communications Psychology* 1 (1). <https://doi.org/10.1038/s44271-023-00003-2>.
- Korell, Daniel, Niklas Reinecke, and Lars Lott. 2023. “Student-Led Replication Studies in Comparative Politics: New Findings by Fresh Eyes?” *Zeitschrift für Vergleichende Politikwissenschaft* 17 (3): 261–73. <https://doi.org/10.1007/s12286-023-00578-4>.
- Kuhn, Thomas S. 1970/1996. *The Structure of Scientific Revolutions*. 3rd ed. Chicago: Univ. of Chicago Press.
- Kvarven, Amanda, Eirik Strømmland, and Magnus Johannesson. 2020. “Comparing Meta-Analyses and Preregistered Multiple-Laboratory Replication Projects.” *Nature Human Behaviour* 4 (4): 423–34. <https://doi.org/10.1038/s41562-019-0787-z>.
- Lakens, Daniel. 2023. “Concerns about Replicability, Theorizing, Applicability, Generalizability, and Methodology Across Two Crises in Social Psychology.” <https://doi.org/10.31234/osf.io/dtvs7>.
- Landy, Justin F., Miaolei Liam Jia, Isabel L. Ding, Domenico Viganola, Warren Tierney, Anna Dreber, Magnus Johannesson, et al. 2020. “Crowdsourcing Hypothesis Tests: Making Transparent How Design Choices Shape Research Results.” *Psychological Bulletin* 146 (5): 451–79. <https://doi.org/10.1037/bul0000220>.
- LeBel, Etienne P., Wolf Vanpaemel, Irene Cheung, and Lorne Campbell. 2019. “A Brief Guide to Evaluate Replications.” *Meta-Psychology* 3. <https://doi.org/10.15626/MP.2018.843>.
- Leonelli, Sabina. 2023. *Philosophy of Open Science*. Cambridge University Press. <https://doi.org/10.1017/9781009416368>.
- Lewin, Kurt. 1930. “Der übergang von Der Aristotelischen Zur Galileischen Denkweise in Biologie Und Psychologie.” *Erkenntnis* 1 (1): 421–66. <https://doi.org/10.1007/BF00208633>.
- Liu, Chunli, Lie Wang, Ruiqun Qi, Weiqiu Wang, Shanshan Jia, Deshu Shang, Yangguang Shao, et al. 2019. “Prevalence and Associated Factors of Depression and Anxiety Among Doctoral Students: The Mediating Effect of Mentoring Relationships on the Association Between Research Self-Efficacy and Depression/Anxiety.” *Psychology Research and Behavior Management* 12: 195–208. <https://doi.org/10.2147/PRBM.S195131>.
- Lopez-Nicolas, Ruben, Daniel Lakens, José A López-López, María Rubio Aparicio, Alejandro Sandoval-Lentisco, Carmen López-Ibáñez, Desirée Blázquez-Rincón, and Julio Sánchez-Meca. 2022. “Analytical Reproducibility and Data Reusability of Published Meta-Analyses on Clinical Psychological Interventions.” *PsyArXiv*.
- Martin, G. N., and Richard M. Clarke. 2017. “Are Psychology Journals Anti-Replication? A Snapshot of Editorial Practices.” *Frontiers in Psychology* 8: 523. <https://doi.org/10.3389/fpsyg.2017.00523>.
- McShane, Blakeley B, Ulf Böckenholt, and Karsten T Hansen. 2022. “Variation and Covariation in Large-Scale Replication Projects: An Evaluation of Replicability.” *J. Am. Stat. Assoc.* 117 (540): 1605–21.
- Mede, Niels G., Mike S. Schäfer, Ricarda Ziegler, and Markus Weißkopf. 2020. “The “Replication Crisis” in the Public Eye: Germans’ Awareness and Perceptions of the

- (Ir)reproducibility of Scientific Research.” *Public Understanding of Science* (Bristol, England), 963662520954370. <https://doi.org/10.1177/0963662520954370>.
- Meehl, Paul E. 2004. “Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology.” *Applied and Preventive Psychology* 11 (1): 1. <https://doi.org/10.1016/j.appsy.2004.02.001>.
- Mellers, Barbara, Ralph Hertwig, and Daniel Kahneman. 2001. “Do Frequency Representations Eliminate Conjunction Effects? An Exercise in Adversarial Collaboration.” *Psychological Science* 12 (4): 269–75.
- Merton, Robert K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago press.
- Muhmenthaler, Michèle C, Mirela Dubravac, and Beat Meier. 2022. “The Future Failed: No Evidence for Precognition in a Large Scale Replication Attempt of Bem (2011).” *Psychology of Consciousness: Theory, Research, and Practice*.
- Muthukrishna, Michael, and Joseph Henrich. 2019. “A Problem in Theory.” *Nature Human Behaviour* 349 (Suppl 1): aac4716. <https://doi.org/10.1038/s41562-018-0522-1>.
- Mynatt, Clifford R., Michael E. Doherty, and Ryan D. Tweney. 1977. “Confirmation Bias in a Simulated Research Environment: An Experimental Study of Scientific Inference.” *Quarterly Journal of Experimental Psychology* 29 (1): 85–95. <https://doi.org/10.1080/00335557743000053>.
- Naddaf, Miryam. 2023. “ChatGPT Generates Fake Data Set to Support Scientific Hypothesis.” *Nature* 623 (7989): 895–96. <https://doi.org/10.1038/d41586-023-03635-w>.
- Nelson, Leif D., Joseph Simmons, and Uri Simonsohn. 2018. “Psychology’s Renaissance.” *Annual Review of Psychology* 69: 511–34. <https://doi.org/10.1146/annurev-psych-122216-011836>.
- Nelson, Lindsay, Honghan Ye, Anna Schwenn, Shinhyo Lee, Salsabil Arabi, and B Ian Hutchins. 2022. “Robustness of Evidence Reported in Preprints During Peer Review.” *Lancet Glob. Health* 10 (11): e1684–87.
- Neoh, Michelle Jin Yee, Alessandro Carollo, Albert Lee, and Gianluca Esposito. 2023. “Fifty Years of Research on Questionable Research Practises in Science: Quantitative Analysis of Co-Citation Patterns.” *Royal Society Open Science* 10 (10): 230677. <https://doi.org/10.1098/rsos.230677>.
- Nickerson, Raymond S. 1998. “Confirmation Bias: A Ubiquitous Phenomenon in Many Guises.” *Review of General Psychology* 2 (2): 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>.
- Nosek, Brian A, Charles R Ebersole, Alexander C DeHaven, and David T Mellor. 2018. “The Preregistration Revolution.” *Proceedings of the National Academy of Sciences* 115 (11): 2600–2606.
- Nuijten, Michele B., Marcel A. L. M. van Assen, Chris Hubertus Joseph Hartgerink, Sacha Epskamp, and Jelte M. Wicherts. 2017. “The Validity of the Tool ‘Statcheck’ in Discovering Statistical Reporting Inconsistencies.” <https://doi.org/10.31234/osf.io/tcxaj>.
- Nuijten, Michèle B, Chris HJ Hartgerink, Marcel ALM Van Assen, Sacha Epskamp, and Jelte M Wicherts. 2016. “The Prevalence of Statistical Reporting Errors in Psychology (1985–2013).” *Behavior Research Methods* 48: 1205–26.

- O'Donnell, Michael, Leif D. Nelson, Evi Ackermann, Balazs Aczel, Athfah Akhtar, Silvio Al-drovandi, Nasseem Alshaif, et al. 2018. "Registered Replication Report: Dijksterhuis and van Knippenberg (1998)." *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 13 (2): 268–94. <https://doi.org/10.1177/1745691618755704>.
- O'Grady, Cathleen. 2021. "Honesty Study Was Based on Fabricated Data." *Science* 373 (65-58): 950–51.
- Open Science Collaboration. 2015. "Psychology: Estimating the Reproducibility of Psychological Science." *Science (New York, N.Y.)* 349 (6251): aac4716. <https://doi.org/10.1126/science.aac4716>.
- Openness, D H Nrwl. 2023. "Open-Access-Strategie Der Hochschulen Des Landes NRW." Zenodo.
- Oswald, and Grosjean. 2004. "Oswald, m. E., & Grosjean, s. (2004). Confirmation Bias. In r. F. Pohl (Ed.). Cognitive Illusions. A Handbook on Fallacies and Biases in Thinking, Judgement and Memory. Hove and n.y. Psychology Press." Unpublished. <https://doi.org/10.13140/2.1.2068.0641>.
- Parsons, Sam, Flávio Azevedo, Mahmoud M. Elsherif, Samuel Guay, Owen N. Shahim, Gisela H. Govaart, Emma Norris, et al. 2022. "A Community-Sourced Glossary of Open Scholarship Terms." *Nature Human Behaviour* 6 (3): 312–18. <https://doi.org/10.1038/s41562-021-01269-4>.
- Pennington, Charlotte Rebecca, and Madeleine Pownall. 2024. "What Have We Learned from the Replication Crisis? Integrating Open Research into Social Psychology Teaching." *PsyArXiv*.
- Platt, John R. 1964. "Strong Inference: Certain Systematic Methods of Scientific Thinking May Produce Much More Rapid Progress Than Others." *Science* 146 (3642): 347–53.
- Plesser, Hans E. 2018. "Reproducibility Vs. Replicability: A Brief History of a Confused Terminology." *Frontiers in Neuroinformatics* 11 (January). <https://doi.org/10.3389/fninf.2017.00076>.
- Popper, Karl R. 1959/2008. *The Logic of Scientific Discovery*. Repr. 2008. Routledge Classics. London: Routledge Classics; Routledge.
- Priem, Jason, Heather Piwowar, and Richard Orr. 2022. "OpenAlex: A Fully-Open Index of Scholarly Works, Authors, Venues, Institutions, and Concepts."
- Quan, Joshua. 2021. "Toward Reproducibility: Academic Libraries and Open Science." Cambridge University Press.
- Rahal, Rima-Maria, Susann Fiedler, Adeyemi Adetula, Ronnie P-A Berntsson, Ulrich Dirnagl, Gordon B. Feld, Christian J. Fiebach, Samsad Afrin Himi, Aidan J. Horner, Tina B. Lonsdorf, Felix Schönbrodt, et al. 2023. "Quality Research Needs Good Working Conditions." *Nature Human Behaviour* 7 (2): 164–67. <https://doi.org/10.1038/s41562-022-01508-2>.
- Rahal, Rima-Maria, Susann Fiedler, Adeyemi Adetula, Ronnie P-A Berntsson, Ulrich Dirnagl, Gordon B Feld, Christian J Fiebach, Samsad Afrin Himi, Aidan J Horner, Tina B Lonsdorf, and others. 2023. "Quality Research Needs Good Working Conditions." *Nature Human Behaviour* 7 (2): 164–67.
- Redden, Joseph P., and Stephen J. Hoch. 2009. "The Presence of Variety Reduces Perceived Quantity." *Journal of Consumer Research* 36 (3): 406–17. <https://doi.org/10.1086/598971>.

- Riedel, Nico, Susanne Wieschowski, Till Bruckner, Martin R. Holst, Hannes Kahrass, Edris Nury, Joerg J. Meerpohl, Maia Salholz-Hillel, and Daniel Strech. 2022. "Results Dissemination from Completed Clinical Trials Conducted at German University Medical Centers Remained Delayed and Incomplete. The 2014 -2017 Cohort." *Journal of Clinical Epidemiology* 144: 1–7. <https://doi.org/10.1016/j.jclinepi.2021.12.012>.
- Rife, Sean, Quinn Lambert, Robert Calin-Jageman, Adamkovic Matus, Gabriel Banik, Itxaso Barberia, Jennifer Beaudry, et al. 2024. "Registered Replication Report: Study 3 from Trafimow and Hughes (2012)." *Advances in Methods and Practices in Psychological Science*.
- Robinaugh, Donald J., Jonas M. B. Haslbeck, Oisín Ryan, Eiko I. Fried, and Lourens J. Waldorp. 2021. "Invisible Hands and Fine Calipers: A Call to Use Formal Theory as a Toolkit for Theory Construction." *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 16 (4): 725–43. <https://doi.org/10.1177/1745691620974697>.
- Robinson, Eric. 2011. "Not Feeling the Future: A Failed Replication of Retroactive Facilitation of Memory Recall." *Journal of the Society for Psychical Research* 75 (904).
- Roe, Chris A, Sarah Grierson, and Aaron Lomas. 2012. "Feeling the Future: Two Independent Replication Attempts."
- Röseler, Lukas, Georg Felser, Jana Asberger, and Astrid Schütz. 2020. "No Effect of Variety on Perceived Quantity: Evidence from Six Studies." <https://doi.org/10.31234/osf.io/v643q>.
- Röseler, Lukas, Taisia Gendlina, Josefine Krapp, Noemi Labusch, and Astrid Schütz. 2022. "Successes and Failures of Replications: A Meta-Analysis of Independent Replication Studies Based on the OSF Registries." <https://doi.org/10.31222/osf.io/8psw2>.
- Röseler, Lukas, and Astrid Schütz. 2022. "Open Science." In *Psychologie*, edited by Astrid Schütz, Matthias Brand, Sabine Steins-Loeber, Martin Baumann, Jan Born, Veronika Brandstätter, Claus-Christian Carbon, et al., 187–98. Stuttgart: Kohlhammer.
- Röseler, Lukas, Astrid Schütz, Roy F. Baumeister, and Ulrike Starker. 2020. "Does Ego Depletion Reduce Judgment Adjustment for Both Internally and Externally Generated Anchors?" *Journal of Experimental Social Psychology* 87: 103942. <https://doi.org/10.1016/j.jesp.2019.103942>.
- Rosenthal, Robert. 1979. "The File Drawer Problem and Tolerance for Null Results." *Psychological Bulletin* 86 (3): 638–41. <https://doi.org/10.1037/0033-2909.86.3.638>.
- Rubin, Mark. 2023. "The Replication Crisis Is Less of a 'Crisis' in the Lakatosian Approach Than It Is in the Popperian and Naïve Methodological Falsificationism Approaches." *BITSS*.
- Samuelson, William, and Richard Zeckhauser. 1988. "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty* 1 (1): 7–59. <https://doi.org/10.1007/BF00055564>.
- Sarstedt, Marko, and Susanne J Adler. 2023. "An Advanced Method to Streamline p-Hacking." *Journal of Business Research* 163: 113942.
- Sarstedt, Marko, Susanne J Adler, Christian M Ringle, Gyeongcheol Cho, Adamantios Diamantopoulos, Heungsun Hwang, and Benjamin D Liengaard. 2024. "Same Model, Same Data, but Different Outcomes: Evaluating the Impact of Method Choices in Structural Equation Modeling." *J. Prod. Innov. Manage.*, April.
- Satinsky, Emily N., Tomoki Kimura, Mathew V. Kiang, Rediet Abebe, Scott Cunningham, Hedwig Lee, Xiaofei Lin, et al. 2021. "Systematic Review and Meta-Analysis of Depression,

- Anxiety, and Suicidal Ideation Among Ph.d. Students.” *Scientific Reports* 11 (1): 14370. <https://doi.org/10.1038/s41598-021-93687-7>.
- Schimmack, Ulrich. 2019. “The Implicit Association Test: A Method in Search of a Construct.” *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, 1745691619863798. <https://doi.org/10.1177/1745691619863798>.
- . 2020. “A Meta-Psychological Perspective on the Decade of Replication Failures in Social Psychology.” *Canadian Psychology/Psychologie Canadienne*. <https://doi.org/10.1037/cap0000246>.
- Schmidt, Birgit, Andrea Chiarelli, Lucia Loffreda, and Jeroen Sondervan. 2024. “Emerging Roles and Responsibilities of Libraries in Support of Reproducible Research.” *LIBER Q.* 33 (1): 1–21.
- Schönbrodt, Felix. 2016. “P-Hacker: Train Your p-Hacking Skills!” <http://shinyapps.org/apps/p-hacker/>.
- Schönbrodt, Felix, Anna Baumert, Andreas Glöckner, Mitja Back, Ruben C Arslan, Martin Voracek, Kai T Horstmann, et al. 2022. “Netzwerk Der Open-Science-Initiativen (NOSI).” Open Science Framework.
- Schönbrodt, Felix, Anne Gärtner, Maximilian Frank, Mario Gollwitzer, Malika Ihle, Dorothee Mischkowski, Le Vy Phan, et al. 2022. “Responsible Research Assessment i: Implementing DORA for Hiring and Promotion in Psychology.”
- Serra-Garcia, Marta, and Uri Gneezy. 2021. “Nonreplicable Publications Are Cited More Than Replicable Ones.” *Science Advances* 7 (21). <https://doi.org/10.1126/sciadv.abd1705>.
- Shiffrin, Richard M, Katy Börner, and Stephen M Stigler. 2018. “Scientific Progress Despite Irreproducibility: A Seeming Paradox.” *Proceedings of the National Academy of Sciences* 115 (11): 2632–39.
- Siems, Renke. 2024. “Subprime Impact Crisis. Bibliotheken, Politik Und Digitale Souveränität.” *BIBL. Forsch. Prax.* 0 (0).
- Sikorski, Michał, and Mattia Andreoletti. 2023. “Epistemic Functions of Replicability in Experimental Sciences: Defending the Orthodox View.” *Found. Sci.*, February.
- Silverstein, Priya, Colin Elman, Amanda Kay Montoya, Barbara McGillivray, Charlotte Rebecca Pennington, Chase H Harrison, Crystal Nicole Steltenpohl, et al. 2023. “A Guide for Social Science Journal Editors on Easing into Open Science.”
- Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. “False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant.” *Psychological Science* 22 (11): 1359–66. <https://doi.org/10.1177/0956797611417632>.
- Simonsohn, Uri, Leif D. Nelson, and Joseph P. Simmons. 2014. “P-Curve: A Key to the File-Drawer.” *Journal of Experimental Psychology. General* 143 (2): 534–47. <https://doi.org/10.1037/a0033242>.
- Sion, Tomos G ap, and Brian D Earp. 2024. “Replication and Theory Development in the Social and Psychological Sciences.”
- Smaldino, Paul. 2019. “Better Methods Can’t Make up for Mediocre Theory.” *Nature* 575 (7781): 9. <https://doi.org/10.1038/d41586-019-03350-5>.
- Smaldino, Paul E. 2017. “Models Are Stupid, and We Need More of Them.” In *Computational*

- Social Psychology*, edited by Robin R. Vallacher, 311–31. New York : Routledge, 2017. | Series: Frontiers of social psychology: Routledge. <https://doi.org/10.4324/9781315173726-14>.
- Smedslund, Jan. 2015. “Why Psychology Cannot Be an Empirical Science.” *Integrative Psychological & Behavioral Science*. <https://doi.org/10.1007/s12124-015-9339-x>.
- Smith, Neal, Jr, and Aaron Cumberledge. 2020. “Quotation Errors in General Science Journals.” *Proc. Math. Phys. Eng. Sci.* 476 (2242): 20200538.
- Soderberg, Courtney K., Timothy M. Errington, Sarah R. Schiavone, Julia Bottesini, Felix Singleton Thorn, Simine Vazire, Kevin M. Esterling, and Brian A. Nosek. 2021. “Initial Evidence of Research Quality of Registered Reports Compared with the Standard Publishing Model.” *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-021-01142-4>.
- Sönning, Lukas, and Valentin Werner. 2021. “The Replication Crisis, Scientific Revolutions, and Linguistics.” *Linguistics* 59 (5): 1179–1206. <https://doi.org/10.1515/ling-2019-0045>.
- Soto, Christopher J. 2019. “How Replicable Are Links Between Personality Traits and Consequential Life Outcomes? The Life Outcomes of Personality Replication Project.” *Psychological Science* 30 (5): 711–27. <https://doi.org/10.1177/0956797619831612>.
- 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 (285): 30–34. <http://www.jstor.com/stable/2282137>.
- Sterling, Theodore 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 (285): 30–34.
- Strack, Fritz. 2016. “Reflection on the Smiling Registered Replication Report.” *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 11 (6): 929–30. <https://doi.org/10.1177/1745691616674460>.
- Strack, Fritz, Leonard L. Martin, and Sabine Stepper. 1988. “Inhibiting and Facilitating Conditions of the Human Smile: A Nonobtrusive Test of the Facial Feedback Hypothesis.” *Journal of Personality and Social Psychology* 54 (5): 768–77. <https://doi.org/10.1037/0022-3514.54.5.768>.
- Strecker, Dorothea. 2019. “Nutzung Der Schattenbibliothek Sci-Hub in Deutschland.”
- Stroebe, Wolfgang, Tom Postmes, and Russell Spears. 2012. “Scientific Misconduct and the Myth of Self-Correction in Science.” *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 7 (6): 670–88. <https://doi.org/10.1177/1745691612460687>.
- The Turing Way Community, and Scriberia. 2024. “Illustrations from the Turing Way: Shared Under CC-BY 4.0 for Reuse.” Zenodo. <https://doi.org/10.5281/ZENODO.3332807>.
- Tierney, Warren, Jay H Hardy III, Charles R Ebersole, Keith Leavitt, Domenico Viganola, Elena Giulia Clemente, Michael Gordon, et al. 2020. “Creative Destruction in Science.” *Organizational Behavior and Human Decision Processes* 161: 291–309.
- Trafimow, David, and Brian D Earp. 2016. “Badly Specified Theories Are Not Responsible for the Replication Crisis in Social Psychology: Comment on Klein.” *Theory Psychol.* 26 (4): 540–48.
- UNESCO. 2020. “First Draft of the UNESCO Recommendation on Open Science.”

- Vaidis, David C, Willem WA Slegers, Florian Van Leeuwen, Kenneth G DeMarree, Bjørn Sætrevik, Robert M Ross, Kathleen Schmidt, et al. 2024. "A Multilab Replication of the Induced-Compliance Paradigm of Cognitive Dissonance." *Advances in Methods and Practices in Psychological Science* 7 (1): 25152459231213375.
- van Aert, Robbie Cornelis Maria, and Marcel A. L. M. van Assen. 2018. "P-Uniform*." <https://doi.org/10.31222/osf.io/zqjr9>.
- Van den Akker, Olmo, Marcel A L M van Assen, Manon Enting, M de Jonge, How Hwee Ong, Franziska Frauke Rüffer, Martijn Schoenmakers, Andrea Helena Stoevenbelt, Jelte M Wicherts, and Marjan Bakker. 2022. "Selective Hypothesis Reporting in Psychology: Comparing Preregistrations and Corresponding Publications." *BITSS*.
- van Noorden, Richard. 2023. "How Big Is Science's Fake-Paper Problem?" *Nature* 623 (7987): 466–67. <https://doi.org/10.1038/d41586-023-03464-x>.
- Vohs, Kathleen D, Brandon J Schmeichel, Sophie Lohmann, Quentin F Gronau, Anna J Finley, Sarah E Ainsworth, Jessica L Alquist, et al. 2021. "A Multisite Preregistered Paradigmatic Test of the Ego-Depletion Effect." *Psychological Science* 32 (10): 1566–81.
- Wagenmakers, Eric-Jan, Titia Beek, Laura Dijkhoff, Quentin F. Gronau, Alberto Acosta, Reginald B. Adams, Daniel N. Albohn, et al. 2016. "Registered Replication Report: Strack, Martin, & Stepper (1988)." *Perspectives on Psychological Science : A Journal of the Association for Psychological Science* 11 (6): 917–28. <https://doi.org/10.1177/1745691616674458>.
- Wagenmakers, Eric-Jan, Ruud Wetzels, Denny Borsboom, and Han L. J. van der Maas. 2011. "Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi: Comment on Bem (2011)." *Journal of Personality and Social Psychology* 100 (3): 426–32. <https://doi.org/10.1037/a0022790>.
- Wicherts, Jelte M, Coosje L S Veldkamp, Hilde E M Augusteijn, Marjan Bakker, Robbie C M van Aert, and Marcel A L M van Assen. 2016. "Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking." *Front. Psychol.* 7 (November): 1832.
- Wilcox, Rand R, and Guillaume A Rousselet. 2024. "More Reasons Why Replication Is a Difficult Issue."
- Williams, Lawrence E, and John A Bargh. 2008. "Experiencing Physical Warmth Promotes Interpersonal Warmth." *Science* 322 (5901): 606–7.
- Wittgenstein, Ludwig. 1968. *Philosophical Investigations*. Oxford: Basil Blackwell.
- Xiao, Qinyu, Choi Shan Lam, Muhrajan Piara, and Gilad Feldman. 2021. "Revisiting Status Quo Bias." *Meta-Psychology* 5. <https://doi.org/10.15626/MP.2020.2470>.
- Yamashita, Taichi, and Reza Neiriz. 2024. "Why Replicate? Systematic Review of Calls for Replication in Language Teaching." *Research Methods in Applied Linguistics* 3 (1): 100091.
- Yarkoni, Tal. 2019. "The Generalizability Crisis." <https://doi.org/10.31234/osf.io/jqw35>.
- Yeung, Andy W. K. 2017. "Do Neuroscience Journals Accept Replications? A Survey of Literature." *Frontiers in Human Neuroscience* 11: 468. <https://doi.org/10.3389/fnhum.2017.00468>.
- Yu, Erica C., Amber M. Sprenger, Rick P. Thomas, and Michael R. Dougherty. 2014. "When Decision Heuristics and Science Collide." *Psychonomic Bulletin & Review* 21 (2): 268–82. <https://doi.org/10.3758/s13423-013-0495-z>.

- Ziemann, Mark, Yotam Eren, and Assam El-Osta. 2016. “Gene Name Errors Are Widespread in the Scientific Literature.” *Genome Biology* 17: 1–3.
- Zimmermann, Christian. 2015. “On the Need for a Replication Journal.” *Wp* 2015 (016).