

# **Open Science: Wie sich die Wissenschaft öffnet**

Lukas Röseler

## **Table of contents**

# 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.2020], einer Revolution [Nosek.] oder einer Renaissance [Nelson.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.

## Literatur

# 1 Über das Buch

Wissenschaft wird zu einem maßgeblichen Teil von Steuergeldern finanziert. Das bedeutet, eine Gesellschaft fördert einzelne Personen, um ein tieferes Verständnis von 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 Gütekriterien 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.

## i 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 in Bamberg meinen Dokortitel in Psychologie erhalten. Seit 2023 bin ich Geschäftsführer des Münster Center for Open Science, 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 von außerhalb zumindest teilweise verstehen können, was dort vor sich geht. Angesichts dieses Zieles ist das Buch kostenlos und online verfügbar.



Figure 1.1: Blick über Wernigerode

# **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](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

## 2 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 [OpenScienceCollaboration.2015]. Fecher und Friesike ([fecher2014open], 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.2021], oder psychische Probleme bei Jungwissenschaftler\*innen [Satinsky.2021].

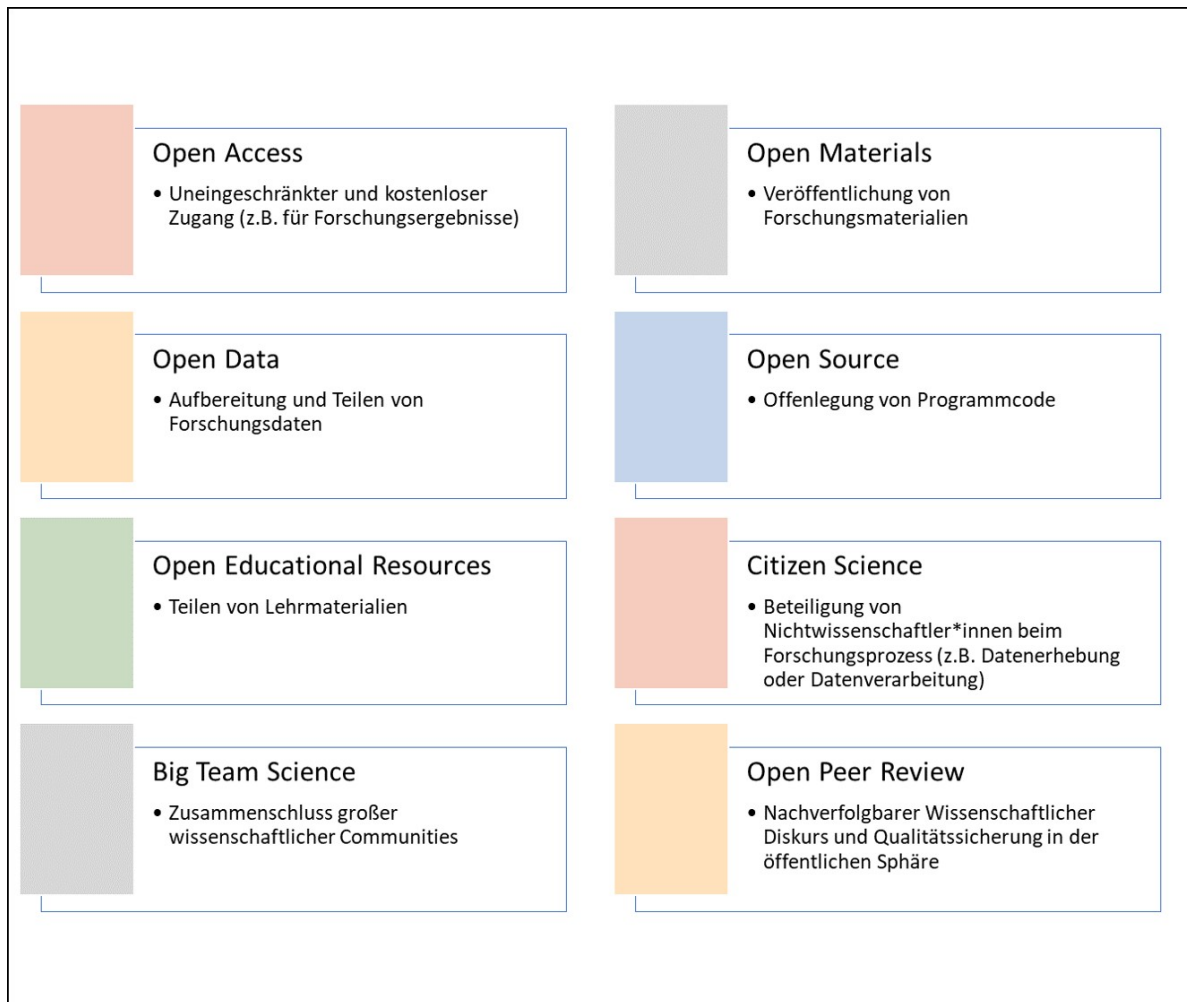


Figure 2.1: Facetten von Open Science

# 3 Wie ist mit Open Science umzugehen?

## 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 3.1: 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**

## Geschichte

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.c].

## Literatur

## 4 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.

### 4.0.1 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.<sup>1</sup> 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 [o2021honesty] 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.2021b]. Dabei ist zu beachten, dass Studien durch gefälschte Daten besonders innovativ, überraschend, oder klar werden - Eigenschaften, die die Veröffentlichung in einer Fachzeitschrift wahrscheinlicher machen.

### 4.0.2 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

---

<sup>1</sup>Mittels 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.2011], Replikationen durch unabhängige Forschende schlugen jedoch fehl [robinson2011not, muhmenthaler2022future, roe2012feeling].

#### 4.0.3 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 [williams2008experiencing]. 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 [bargh1996automaticity]. 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.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.

#### 4.0.4 Literatur

## 5 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 [OpenScienceCollaboration.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. In dem 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. Roseler.2022d). Einige gleichermaßen ambitionierte Vorhaben wurden veröffentlicht, wie zum Beispiel die ManyLabs Studien (z.B. Klein.2014; Klein.2018) oder Versuche, bei denen unabhängige Gruppen dieselben Hypothesen testeten und replizierten [Landy.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.1988] zu replizieren [Wagenmakers.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.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 [byers2020building] oder auf ganze Zeitschriften [Camerer.2018].

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

Projekt	Thema	Link
Reproducibility Project: Psychology CORE Data Replicada	Psychologie	<a href="https://osf.io/ezcu/">https://osf.io/ezcu/</a>
Many Labs 1	Entscheidungsforschung	<a href="https://osf.io/5z4a8/">https://osf.io/5z4a8/</a>
Many Labs 2	Konsumentenverhalten	<a href="https://datacolada.org/archives/category/replication">https://datacolada.org/archives/category/replication</a>
Many Labs 3	und Entscheidungs-	
Many Labs 4	forschung	
Many Labs 5	Psychologie	<a href="https://osf.io/wx7ck/">https://osf.io/wx7ck/</a>
Soto	Psychologie	<a href="https://osf.io/8cd4r/">https://osf.io/8cd4r/</a>
Social Sciences Replication Project	Psychologie	<a href="https://osf.io/ct89g/">https://osf.io/ct89g/</a>
Registered Replication Reports	Psychologie	<a href="https://osf.io/8ccnw/">https://osf.io/8ccnw/</a>
Many Babies 1	Psychologie	<a href="https://osf.io/7a6rd/">https://osf.io/7a6rd/</a>
Sports Sciences Replications	Persönlichkeitspsychologie	<a href="https://doi.org/10.1177/0956797619831612">https://doi.org/10.1177/0956797619831612</a>
Hagen	Verhaltensforschung	<a href="http://www.socialsciencesreplicationproject.com">http://www.socialsciencesreplicationproject.com</a>
Cumulative Science Project		
I4R Replications	Verschiedene	–
Experimental Philosophy		
Reproducibility Project: Cancer SCORE	Entwicklungspsychologie	<a href="https://manybabies.org">https://manybabies.org</a>
REPEAT	Sportwissenschaften	<a href="https://ssreplicationcentre.com">https://ssreplicationcentre.com</a>
CREP	Psychologie	<a href="https://osf.io/d7za8/">https://osf.io/d7za8/</a>
Boyce et al., 2023	Politikwissenschaften	<a href="https://i4replication.org/reports.html">https://i4replication.org/reports.html</a>
ReproSci	Experimentelle	<a href="https://doi.org/10.1007/s13164-018-0400-9">https://doi.org/10.1007/s13164-018-0400-9</a>
Boyce et al. 2024	Philosophie	
	Krebsforschung	<a href="https://www.cos.io/rpcb">https://www.cos.io/rpcb</a>
	(Medizin)	
	Sozialwissenschaften	<a href="https://www.cos.io/score">https://www.cos.io/score</a>
	Gesundheitssystem	<a href="https://www.repeatinitiative.org">https://www.repeatinitiative.org</a>
	Psychologie	<a href="https://www.crep-psych.org">https://www.crep-psych.org</a>
	Psychologie	<a href="https://doi.org/10.1098/rsos.231240">https://doi.org/10.1098/rsos.231240</a>
	Biologie	<a href="https://reprosci.epfl.ch">https://reprosci.epfl.ch</a>
	Psychologie	<a href="https://doi.org/10.31234/osf.io/an3yb">https://doi.org/10.31234/osf.io/an3yb</a>

### 5.0.1 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 5.2: Replikations-Taxonomie nach Turing Way [turingwaycommunity2024illustrations]

		Daten	
Analyse	gleich	gleich reproduzierbar	unterschiedlich replizierbar
	unterschiedlich	robust	verallgemeinerbar

### 5.0.2 Weiterführende Literatur

Für eine systematischere, in den Informationswissenschaften verankerte Taxonomie zur Art der Replikation siehe [Plessner2018]. Eine an den statistischen Methoden angelehnte Taxonomie für die Ergebnisse von Replikationsstudien haben LeBel et al. [LeBel.2019] vorgeschlagen. Philosophisch diskutiert wird Replikationsnähe zum Beispiel von [Choi2023] und [Leonelli2023].

Table 5.3: 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)	<b>Direkte Replikation</b> (selbe Versuchsleiter*innen, selbe Versuchsmaterialien, neue Versuchspersonen) <b>Nahe Replikation</b> (andere Versuchsleiter*innen, möglichst ähnliche Versuchsmaterialien, neue Versuchspersonen) <b>Konzeptuelle oder konstruktive Replikation</b> (andere Versuchsleiter*innen, andere Versuchsmaterialien, neue Versuchspersonen)
Ziel der Replikation	<b>Reproduktion</b> Mit selben Daten und selbem Programmiercode zu denselben Ergebnissen gelangen <b>Replikation</b> Mit anderen Daten zu denselben Ergebnissen gelangen

### 5.0.3 „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, [Sikorski2023-qp], können sie je nach Forschungsgebiet stark variieren [McShane2022-yd, Landy.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 Abbildung X.

ABBILDUNG X: Forest plot mit simulierten Ergebnissen von Original + vielen Einzelstudien; in den Notes dann ausführliche Erklärung.

```
title: "Wald Diagramm (Forest Plot)"
format:
  html:
```

```

code-fold: true
---

{r}
library(ggplot2)
library(metafor)

set.seed(10)
k <- 15
cors <- data.frame("Stichprobenumfang" = round(rchisq(n = k, df = 2, ncp = 0)*5+10, digits =
      , "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 =
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))

```

### 5.0.4 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.2020], waren sie zudem beim Großteil der Studien null, die Phänomene waren also „nicht sichtbar“ [Alogna.2014; Eerland.2016; Bouwmeester.2017; ODonnell.2018; Wagenmakers.2016.; Cheung.2016; vaidis2024multilab; rife2024registered]. So konnte beispielsweise mit einer enormen Präzision gezeigt werden, dass eine Geschichte über einen Professor Versuchspersonen in einem anschließenden Leistungstest *\*nicht\** schlauer macht [ODonnell.2018].