

METHODENKRITIK Seit rund zehn Jahren befindet sich die Psychologie im Umbruch. Vor allem Nachwuchsforscher fordern neue Strategien, um Studien verlässlicher zu machen.

Revolution von unten

VON MICHAEL WENZLER

Ich war am Boden zerstört.« So beschreibt Jürgen Schneider seine Frustration, nachdem er den Workshop »Karriere in der Wissenschaft« absolviert hatte. Während seines Studiums der Pädagogik und Psychologie an der Ludwig-Maximilians-Universität (LMU) in München hatte er sich für Studien und Statistik begeistert. Schnell war ihm klar, dass er nach dem Abschluss weiter forschen wollte. Voller Elan begann er eine Promotion an der Universität Tübingen und besuchte den besagten Workshop. Doch was er dort hörte, desillusionierte ihn.

Statt Tipps für gute, nachhaltige Wissenschaft vermittelten die Seminarleiter vor allem eines: »publish or perish!« Zu Deutsch: Publiziere, oder du verschwindest von der Bildfläche. Um in der Forschung Fuß zu fassen, müsse man nun einmal so viel wie möglich veröffentlichen und Fördergelder einwerben. »Qualität ist zweitrangig, Hauptsache, du produzierst laufend neue Fachartikel«, hieß es.

Das steht dem Zweck von Wissenschaft, nämlich verlässliche Erkenntnisse zu liefern, allerdings häufig ent-

gegen. Nach dem Karriere-Workshop dämmerte es Schneider, dass er die Art von Forschung, die ihm eigentlich vorschwebte, unter solchen Bedingungen kaum würde betreiben können.

Möglichst viele hochrangige Publikationen zu sammeln, ist gerade für den wissenschaftlichen Nachwuchs wichtig, weil die Konkurrenz um die wenigen festen Stellen so groß ist. Laut Deutschem Hochschulverband kamen 2015 auf einen altersbedingt ausscheidenden Emeritus mindestens sieben qualifizierte Jungforscher. Nicht selten bewerben sich hunderte Kandidaten auf eine Ausschreibung. Jenseits von Professuren sind die meisten Posten an Hochschulen und Forschungseinrichtungen zeitlich befristet. Kritisch ist dies vor allem deshalb, weil gemäß dem Wissenschaftszeitvertragsgesetz forschende Mitarbeiter an Universitäten nur maximal zwölf Jahre befristet angestellt sein dürfen. Danach bleibt nur die Entfristung, was eher selten passiert, oder der Abschied vom Labor.

Doch woran bemisst sich akademische Leistung? Traditionell macht man sie vor allem an der Zahl der



UNSER EXPERTE

Michael Wenzler ist Sozialpsychologe und arbeitet derzeit an der FernUniversität Hagen. Wie leicht wir uns selbst und anderen etwas vormachen, weiß er auch aus eigener Forschung zur Moralpsychologie.



EASTON GORDON / GETTY IMAGES / ISTOCK

TÜCKE DES OBJEKTS | Unsere Psyche unterliegt zahllosen Einflüssen. Das macht ihre Erforschung anfällig für Zufallsbefunde.

Auf einen Blick: Neue Standards für bessere Wissenschaft

1 Die Psychologie hat ein Problem: Falsche Anreize fördern die Neigung, allein schillernde Ergebnisse zu publizieren. Viele davon halten näherer Prüfung jedoch nicht stand.

2 Nicht nur statistisch »signifikante« Resultate, sondern auch deren Fehlen in Experimenten ist aufschlussreich. Erst in der Zusammenschau kann man die wahren Effekte ermessen.

3 Zu den Reformen der Forschung zählen das Vorabregistrieren von Studien, Transparenz von Daten oder Methoden sowie die Veröffentlichung aller (auch negativer) Befunde.

Veröffentlichungen in Fachzeitschriften fest. Entscheidend ist hierbei der so genannte Impact-Faktor (siehe »Kurz erklärt«): Je mehr Publikationen in renommierten, häufig zitierten Journalen jemand vorweisen kann, desto besser.

Diesen Druck bekam Jürgen Schneider schon zu Beginn seiner Promotion zu spüren. An vielen Universitäten geben mittlerweile veröffentlichungs-basierte Dissertationen gegenüber der klassischen Monografie den

Ton an. Die Universität Tübingen beispielsweise verlangt zwei akzeptierte »Paper«, also Manuskripte, die ein wissenschaftliches Journal zur Publikation angenommen hat.

Auch bei Berufungsverfahren für Professuren ist die Zahl der »High Impact«-Veröffentlichungen ausschlaggebend. Wenn sich hunderte Personen auf eine Stelle bewerben, bleibt der Berufungskommission oft nichts anderes übrig, als die Güte der Kandidaten anhand der Zahl ihrer Publikationen abzuschätzen. Um diese genau zu lesen und ihre Qualität zu prüfen, fehlt den Mitgliedern dabei meist die Zeit.

Nun ist das Veröffentlichen neuer Ergebnisse natürlich eine wichtige Säule von Wissenschaft. Im Rahmen des so genannten Peer-Review-Verfahrens (siehe »Kurz erklärt«) prüfen andere Forscher die Arbeit ihrer Kollegen, bevor sie gedruckt wird, sprechen Empfehlungen aus oder verlangen inhaltliche Nachbesserungen. Problematisch wird es, wenn diese Kontrolle versagt, weil zum Beispiel entscheidende Details zur Planung, Durchführung oder Auswertung einer Studie nicht offengelegt werden. Je mehr es allein um das Publizieren und Zitiertwerden durch Kollegen geht, desto eher sehen Forscher über Schwächen oder Einschränkungen der eigenen Arbeit hinweg oder kehren sie mitunter sogar aktiv unter den Teppich. Wissenschaftliche Transparenz und Exaktheit stehen hier dem persönlichen Fortkommen im Weg.

Sind Fehlschläge uninteressant?

Ein weiteres Problem für die Verlässlichkeit von Forschung ist der so genannte Publikationsbias (siehe »Kurz erklärt«). Lange Zeit galt nur eine Art von Ergebnissen als publizierbar: statistisch signifikante Bestätigungen für die jeweils untersuchte Hypothese. Sei kein solcher »Effekt« nachweisbar – man spricht hierbei von Nullbefunden –, trage dies nichts zum Verständnis der Fragestellung bei.

Doch das ist falsch. »Wissenschaft ist mehr als sexy Ergebnisse, die statistisch signifikant sind«, erklärt etwa Felix Schönbrodt. Er ist Professor am Lehrstuhl für Psychologische Methodenlehre und Diagnostik an der LMU München. Die Verachtung für Nullbefunde schaffe einen Anreiz, das auszublenden, was nicht ins Bild passt. Wenn ein Experiment nicht das erhoffte Resultat erbringt, stellt sich die Frage, wie Forscher damit umge-



UNIVERSITY OF VIRGINIA, SCHOOL OF MEDICINE

»Nur wenn Studienverfahren und Daten transparent sind, können wir Fehlern und Voreingenommenheit vorbeugen«

Brian Nosek, University of Virginia

hen. Machen sie es kenntlich, riskieren sie, auf Grund »inkonsistenter Befunde« vom Fachjournal abgewiesen zu werden. Folglich werden Fehlschläge eher verschwiegen oder kaschiert.

Bereits 1959 beklagte der US-amerikanische Statistiker Theodore Sterling, dass in einigen Disziplinen, vor allem den Sozial- und Verhaltenswissenschaften, signifikante Befunde mit höherer Wahrscheinlichkeit veröffentlicht werden als Nullbefunde. Das führe zu Verzerrungen und falschen Schlussfolgerungen. So altbekannt das Phänomen ist, die Gemeinde der forschenden Psychologen drückte sich lange darum, die nötigen Konsequenzen zu ziehen.

Wie leicht dies dazu führen kann, dass sich oberflächlicher Erfolg von wissenschaftlicher Qualität abkoppelt, wurde der wissenschaftlichen Community 2011 schlagartig bewusst. Damals veröffentlichte der Sozialpsychologe Daryl Bem im renommierten »Journal of Personality and Social Psychology« einen Aufsatz mit dem Titel »Feeling the future«. Bem demonstrierte anhand von neun Experimenten, dass Probanden in die Zukunft blicken können.

In einer dieser Studien sahen die Teilnehmer mehrfach hintereinander zwei Vorhänge auf einem Computerbildschirm. Sie sollten erraten, hinter welchem ein Bild erscheinen würde oder nicht. Wie sich herausstellte, sagten die Teilnehmer überzufällig oft vorher, wo das Motiv auftauchte – vor allem dann, wenn es sich bei dem später auftauchenden Bild um ein Nacktfoto handelte. Laut diesem und den acht weiteren Versuchen waren die Teilnehmer offenbar dazu in der Lage, kommende Ereignisse vorherzusehen.

Bems Experimente gehorchten allen Regeln der Kunst. Jedes wies in die gleiche Richtung, und die verwendeten Stichproben waren größer als in vielen anerkannten Veröffentlichungen. Felix Schönbrodts erster Gedanke bei Lesen des Artikels war: »Hier läuft etwas grundlegend falsch.« Wenn jemand die Kriterien der Wissenschaft voll erfüllt und dennoch absurde Ergebnisse produziert, müssen die Kriterien eben fehlerhaft sein. Ganz zu schweigen von all den Arbeiten, die nicht so hohe Qualitätsstandards erfüllen wie Bems Hellscher Studien.

Der Forscher selbst wollte mit seiner Veröffentlichung übrigens keineswegs dem eigenen Fach einen Spiegel vorhalten. Er war von der Richtigkeit seiner Befunde überzeugt und verteidigte sie als wissenschaftlich valide. Dies zeigt, wie wichtig harte Methodenstandards sind, damit persönliche Glaubenssätze oder Weltanschauungen nicht als »bewiesen« gelten.

Viele glaubten, sie würden der Wahrheit nur ein wenig »nachhelfen«

Eine weitere Veröffentlichung von 2011 verschärfte die Zweifel. Die Methodenforscher Joseph Simmons, Leif Nelson und Uri Simonsohn wiesen darin nach, wie einfach sich signifikante Ergebnisse aus einem Datensatz gewinnen lassen, obwohl er in Wahrheit keine enthält. Dafür muss man lediglich ein paar statistische Tricks einsetzen, die in Fachartikeln selten erwähnt werden. Das Pikante: Viele dieser Tricks sind unter Forschern gängige Praxis. Experten bezeichnen das Hervorkitzeln von Signifikanz in gegebenen Daten als p-Hacking (siehe »Kurz erklärt«).

»Manche Kritiker unterstellen Kollegen öffentlich Unehrllichkeit und Trickseriei, ohne dass sie Belege dafür haben«

Susan Fiske, Princeton University



SAMUEL KAUS, MIT FRIEDL. GEN. VON SUSAN FISKE

ALEXANDER KURTHING, MITTEL
BILD: VON FELIX SCHÖNBRODT

»Wissenschaft ist mehr als sexy Ergebnisse, die statistisch signifikant sind«

Felix Schönbrodt, Ludwig-Maximilians-Universität München

Vielen sei dabei nicht klar, wie viele falsch positive Ergebnisse dieses Vorgehen produziert, meint die Psychologin Anne Scheel von der Technischen Universität Eindhoven. Stattdessen habe man bisher angenommen, die wahren Effekte würden dadurch im »Rauschen der Daten« nur leichter auffindbar. Doch dabei machten sich forschende Psychologen wohl selbst etwas vor.

Welches grundlegende Problem das Fach hat, zeigte der Sozialpsychologe Brian Nosek von der University of Virginia 2015. Sein Team prüfte, ob sich 100 zufällig ausgewählte psychologische Studien aus drei angesehenen Fachmagazinen mit den gleichen Resultaten wiederholen (»replizieren«) ließen. Siehe da: Lediglich für 36 davon fand man dasselbe Ergebnis – gerade einmal ein

Drittel. Offenbar waren die Befunde also weit weniger robust als angenommen. Dabei ist die Wiederholbarkeit einer Studie eine essenzielle Anforderung an verlässliche Wissenschaft.

Diese Replikationskrise sorgte für Verunsicherung. »Niemand wusste mehr, worauf man aufbauen konnte«, erinnert sich Julia Rohrer, Persönlichkeitspsychologin an der Universität Leipzig. In welche Richtung sollte man überhaupt noch forschen? Welche Befunde waren solide, welche Theorien wirklich durch Daten gestützt? Auf einmal verflüchtigten sich viele anerkannte Modelle und Resultate im Nebel. »Es gibt eine Scheinwelt in unserer wissenschaftlichen Literatur, die mit der Realität nichts zu tun hat«, erklärt Rima-Maria Rahal vom Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern in Bonn. Die Reaktion darauf war Misstrauen und Skepsis.

Dann begannen erste Versuche, die Dinge zum Besseren zu wenden. 2013 gründete Brian Nosek zusammen mit Kollegen das Center for Open Science (COS). Es soll die psychologische Forschung transparenter machen, um fehlerhafte Praktiken in der Forschung aufzudecken und zu unterbinden. Die Reformer schlugen eine Reihe von Maßnahmen und strukturellen Änderungen vor, die nachhaltige Veränderungen anstießen. So registrieren heute viele Forscher ihre Studien vorab, legen also Hypothesen, Vorgehen und Auswertungsverfahren vorher fest. Das hat den Vorteil, dass man genau weiß, von welchen Annahmen sie ausgingen, und verhindert das so genannte HARKing (siehe »Kurz erklärt«) – den Versuch, auf Grundlage bekannter Resultate besser passende Hypothesen zu formulieren.

Ein Wendepunkt in der Karriere

Als Jürgen Schneider von der Open-Science-Initiative hörte, war dies ein Wendepunkt in seiner Karriere. »Ich stand vor der Wahl, mich von der Wissenschaft zu verabschieden oder mich dafür einzusetzen, dass sich etwas ändert.« Er entschied sich für Letzteres.

Die Open-Science-Richtlinien setzt Schneider heute in seiner Forschungsarbeit konsequent um: Er macht Versuchspläne, Analysen und Materialien öffentlich zugänglich. Jeder kann darauf zugreifen und sie nutzen. Zudem hat er eine entsprechende Initiative in Tübingen

gegründet und arbeitet mit Kollegen an einem Positionspapier dazu, was Nachwuchswissenschaftler brauchen, um erfolgreich offene Wissenschaft betreiben zu können.

Doch das Engagement hat seinen Preis. Die Zeit, die Schneider aufwendet, um Studien zu registrieren und zu dokumentieren, fehlt ihm im Labor und fürs Schreiben. Beim ersten Mal habe er zwei Wochen für die vollständige Dokumentation eines Versuchs benötigt. Zwar gewinne er zunehmend Routine, aber das Engagement schlage sich nicht direkt in messbaren Vorteilen für seine akademische Zukunft nieder. Denn High-Impact-Publikationen bleiben weiterhin das Hauptkriterium für die Vergabe von Posten und Forschungsgeldern. Der Druck, nach den Maßstäben des »alten Systems« erfolgreich zu sein, hat somit kaum abgenommen. »Saubere Forschung macht mehr Arbeit«, bestätigt Schönbrodt. Dies müsse stärker honoriert werden.

So manche ältere, etablierte Forscher stehen zudem im Verdacht, ihre akademischen Lorbeeren verteidigen zu wollen. So warf Susan Fiske, eine international renommierte Sozialpsychologin von der Princeton University, einigen Vertretern der Open-Science-Bewegung »methodologischen Terrorismus« vor. Sie kritisierte, viele würden Fehler öffentlichkeitswirksam auf Social-Media-Kanälen wie Twitter anprangern, statt mit den Betreffenden auf Konferenzen direkt in Kontakt zu treten. Das grenze an Mobbing, da hier oft so getan werde, als gebe es anständige Forscher auf der einen Seite und nur Betonköpfe auf der anderen.

Von einem Extrem ins andere

Während solche Vorwürfe vielen Beobachtern als die Besitzstandswahrung einer »alten Garde« erscheinen, sorgen sich andere um das Klima innerhalb der wissenschaftlichen Gemeinschaft sowie um die nötige Offenheit für mutige Ideen. So warnt der Psychologe Mario Gollwitzer vor der Gefahr, Herausgeber und Gutachter könnten angesichts der strengeren Regeln zur »Überkompensation« neigen. Gewagte Forschungsansätze würden häufig zu schnell abgelehnt aus Sorge, der Befund sei vielleicht nicht replizierbar. Man dürfe nicht



MEHR WISSEN AUF
»SPEKTRUM.DE«

Mehr zu Forschungsmethoden und Signifikanz lesen Sie in unserem digitalen Spektrum Kompakt »Statistik – Zahlenspiele mit Mehrwert«:

www.spektrum.de/shop

vom einen Extrem – nur staunenswerte, bahnbrechende Befunde zu suchen – ins andere umschlagen, indem man stets auf Nummer sicher geht.

Wie sehr das Thema die Szene aktuell umtreibt, zeigt eine Stellungnahme der Jungmitglieder der Deutschen Gesellschaft für Psychologie (DGPs) vom Juli 2020. Sie beklagen, dass Kosten und Stellenwert offener Wissenschaftspraktiken in einem schlechten Verhältnis stünden. Sich in dieser Hinsicht zu engagieren, sei sogar häufig karriereschädlich.

Die Autoren fordern, man müsse andere Kriterien heranziehen, um wissenschaftliche Leistung zu beurteilen. So sollten Bewerber für eine Professur explizit ihre Erfahrungen in Sachen offene Wissenschaft und Pläne für zukünftige Projekte benennen. Auch gelte es verstärkt auf überzeugende Konzepte und gute wissenschaftliche Praxis zu achten und weniger auf Ergebnisse.

Hier sind vor allem die Forschungsinstitutionen gefragt, glaubt Schneider. Einerseits brauche es mehr praktische Unterstützung, beispielsweise »Data Stewards«, die Nachwuchswissenschaftlern bei der Aufbereitung und Veröffentlichung ihrer Datensätze helfen. Andererseits müsse sich der Einsatz für gute Wissenschaft stärker als bislang auszahlen.

Zwar empfiehlt etwa die Deutsche Forschungsgemeinschaft (DFG) inzwischen, dass Studien vorab registriert sowie alle Daten und Materialien veröffentlicht werden sollten. Doch das müsse auch die Karrieren der Forschenden voranbringen. Sonst haben weiterhin diejenigen das Nachsehen, die nicht auf Teufel kommen raus publizieren. ★

QUELLEN

Attig, C.: Wir brauchen Anreize zur schnelleren Verbreitung von offener und reproduzierbarer Forschung. Stellungnahme der Jungmitglieder der Deutschen Gesellschaft für Psychologie (DGPs). Online 28. Juli 2020 (siehe: www.dgps.de/uploads/media/DGPs_Stellungnahme_Anreize_open_science.pdf)

Bem, D. J.: Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology* 100, 2011

Ritchie, S.: Science fictions. Exposing fraud, bias, negligence and hype in science. Bodley Head, 2020 (siehe Rezension auf S. 80 in diesem Heft)

Simmons, J. P. et al.: False-significant psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22, 2011

Dieser Artikel im Internet: www.spektrum.de/artikel/1805798