SZIENTOMETRIE

Jede Menge Murks

Viele wissenschaftliche Studien lassen sich nicht reproduzieren. Das wirft Fragen zum Forschungsbetrieb auf – und zur Veröffentlichungspraxis von Fachzeitschriften.

Von Ed Yong

Es mutet wie Ironie an, dass ausgerechnet eine Studie über Vorahnungen bei vielen Psychologen das ungute Gefühl auslöste, ihr Fachgebiet befinde sich in Schwierigkeiten. Der Sozialpsychologe Daryl Bem von der Cornell University in Ithaca (New York) hatte eine Aufsehen erregende Untersuchung durchgeführt. Er präsentierte Studenten 48 Wörter und bat sie anschließend darum, alle zu notieren, die sie im Gedächtnis behalten hatten. Später erhielten die Studenten eine zufällige Auswahl der Wörter, verbunden mit dem Auftrag, die Begriffe abzutippen. Dabei zeigte sich ein verblüffender Effekt: Einige Teilnehmer konnten sich im ersten Teil des Experiments besonders gut an jene Wörter erinnern, die sie im zweiten Teil des Experiments übten. Die Wirkung ging also der Ursache voraus.

Bem veröffentlichte seine Ergebnisse zusammen mit acht weiteren Experimenten im »Journal of Personality and Social Psychology«. Er wollte damit Phänomene beweisen, die er als übersinnliche oder »Psi«-Effekte bezeichnete. Erwartungsgemäß stellten zahlreiche Wissenschaftler seine Behauptungen in Frage. Drei Teams versuchten unabhängig voneinander, die Effekte zu reproduzieren. Es gelang ihnen nicht. Als sie dieses Negativergebnis nun ihrerseits veröffentlichen wollten, stießen sie auf große Hürden: Zunächst wollte es niemand drucken. Für die Wissenschaftler war das ein Alarmsignal.

Positive Ergebnisse in der Psychologie sind wie Gerüchte – leicht zu verbreiten, aber schwer zurückzunehmen. Sie prägen den Inhalt der meisten Fachzeitschriften, was kein Wunder ist, denn die Journale berichten mit Vorliebe über neue, spannende Studien. Versuche, diese zu reproduzieren, bleiben dagegen oft unveröffentlicht, insbesondere wenn sie scheitern. Zahllose Negativergebnisse finden ihren Weg nicht aus den Laborbüchern hinaus. »Es gibt Experimente, von denen jeder weiß, dass sie sich nicht nachvollziehen lassen, aber dieses Wissen gelangt nicht in die Fachliteratur«, meint Eric-Jan Wagenmakers, Kognitionswissenschaftler an der Universität von Amsterdam. »Ich kenne Studenten, die ihre gesamte Doktorandenzeit darauf verwendeten, ein bestimmtes Phänomen zu reproduzieren, daran scheiterten und schließlich die Hochschule verließen, weil sie nichts vorzuweisen hatten.«

Das Experiment am gewünschten Ergebnis ausrichten

Solche Probleme gibt es in allen Wissenschaftsbereichen. In der Psychologie werden sie jedoch durch einige tief verwurzelte Normen verschärft. Es ist zum Beispiel üblich geworden, einen Versuchsablauf so hinzubiegen, dass er praktisch ein Positivergebnis garantiert. Wenn dieses Resultat dann einmal veröffentlicht ist, versuchen nur wenige Forscher, den Versuch exakt zu wiederholen. Stattdessen führen sie »konzeptuelle Replikationen« durch, sie prüfen also ähnliche Hypothesen mit anderen Methoden. Kritiker warnen, dass dies dem Bau eines Kartenhauses auf schwankendem Grund gleiche.

Einige schwere Betrugsfälle haben in der Vergangenheit deutlich gemacht, wie verhängnisvoll sich solche Praktiken auswirken können. Nach verbreiteter Einschätzung fliegen gefälschte Ergebnisse oft deshalb nicht auf, weil es so schwierig ist, die entsprechenden Experimente exakt zu wiederholen. Doch seit einiger Zeit verstärken sich die Anstrengungen, diesen Missstand zu beheben. Viele Psychologen haben Initiativen ins Leben gerufen, um das Ausmaß des Problems zu erfassen, Reproduktionsversuche zu unterstützen und eine Plattform zu schaffen, auf der die Ergebnisse solcher Versuche veröffentlicht werden können. »Seit etwa einem Jahr sprechen viel mehr Leute darüber und nehmen Anstoß daran«, sagt Joseph Simmons, Psychologe an der University of

AUF EINEN BLICK

DIE JAGD NACH DEM EFFEKT

1 Fachzeitschriften veröffentlichen bevorzugt positive Forschungsergebnisse – also solche, die eine aufgestellte **Hypothese** bestätigen. Besonders ausgeprägt ist das in der Psychologie.

2 Viele dieser **Positiv-Studien** erweisen sich als nicht reproduzierbar. Die gescheiterten Wiederholungsversuche bleiben jedoch oft unveröffentlicht.

3 Verantwortlich dafür sind die Arbeitspraktiken von Wissenschaftlern, aber auch die **Auswahlkriterien der Journale**. Einige Forscher wollen nun mit verschiedenen Initiativen gegensteuern.

Pennsylvania (USA). »Ich hoffe, die Dinge wenden sich bald zum Besseren.«

Die Psychologie steht mit diesen Schwierigkeiten nicht allein da. Auch in der biomedizinischen Forschung sieht es nicht viel besser aus. In einem mittlerweile berühmten Artikel argumentierte der Epidemiologe John Ioannidis, der zurzeit an der Stanford School of Medicine (USA) arbeitet, dass der statistischen Logik zufolge »die meisten veröffentlichten Forschungsergebnisse falsch sind«. Der Sozialwissenschaftler Daniele Fanelli von der University of Edinburgh in Schottland untersuchte 4600 Studien aus diversen Wissenschaftszweigen und fand heraus, dass der Anteil der positiven Ergebnisse im Jahr 2007 um 22 Prozent höher lag als 1990. Vor allem die Psychologie und die Psychiatrie stechen heraus: Hier ist die Wahrscheinlichkeit, dass ein veröffentlichtes Ergebnis positiv ausfällt, fünfmal so hoch wie in der Weltraumforschung, die am anderen Ende des Spektrums liegt. Und es zeichnet sich keine Besserung ab. Im Jahr 1959 belegte der Statistiker Theodore Sterling, dass 97 Prozent der Studien in vier wichtigen Psychologiejournalen statistisch signifikante Positivergebnisse gemeldet hatten. 1995 führte er die Analyse erneut durch und stellte keine Veränderung fest.

Wenn Experimente wiederholt werden, kommt oft etwas anderes heraus als beim ersten Mal.

SPEKTRUM DER WISSENSCHAFT / MANFRED ZENTSCH

Einer der Gründe dafür, dass positive Resultate in der Psychologie so stark überrepräsentiert sind, sei die Bevorzugung »leicht abgedrehter« Studien, meint Chris Chambers, Psychologe an der Cardiff University in Wales. Die Ergebnisse müssten spannend, plakativ oder sogar unplausibel sein, um veröffentlicht zu werden. Simmons sieht die Ursache zum Teil im Auswahlprozess: »Beim Begutachten von Manuskripten fordern wir von den Autoren häufig den Nachweis, dass ihre Befunde neu oder interessant sind. Die Echtheit der Ergebnisse hingegen müssen sie nur selten belegen.«

Simmons gehört zu denen, die es wissen müssen. Kürzlich veröffentlichte er in der Fachzeitschrift »Psychological Science« eine nicht ganz ernst gemeinte Abhandlung, die angeblich bewies, dass der Beatles-Hit »When I'm Sixty-Four« die Zuhörer um 1,5 Jahre jünger macht. Simmons wollte damit demonstrieren, wie »unerträglich einfach« man statistisch signifikante Ergebnisse produzieren kann, um jede noch so schräge Hypothese zu untermauern. Viele Psychologen legen die Schlüsselfaktoren ihrer Studien, etwa die Zahl der Teilnehmer, die zu messenden Variablen und die Auswertungsverfahren, quasi nebenbei fest. Vielleicht tun sie das bloß aus Unachtsamkeit, aber es gibt ihnen die Freiheit, Versuche und Daten so lange zu verbiegen, bis ein positives Ergebnis herauskommt.

Der Psychologe Leslie John von der Harvard Business School in Boston (USA) befragte etwa 2000 Kollegen und fand heraus, dass mehr als die Hälfte von ihnen bei wissenschaftlichen Untersuchungen zunächst die Signifikanz der Ergebnisse prüften, bevor sie entschieden, ob sie weitere Daten erheben sollten oder nicht. Das taten sie so lange, bis sich ein positives Ergebnis abzeichnete. Vier von zehn Befragten hatten selektiv Studien veröffentlicht, die »funktioniert hatten«. Die meisten waren der Meinung, dass diese Vorgehensweise vertretbar sei. »Viele tun das, weil sie es so gelernt haben«, sagt der Psychologe Brent Roberts von der University of Illinois in Urbana-Champaign (USA).

Kein Interesse an Wiederholungsversuchen

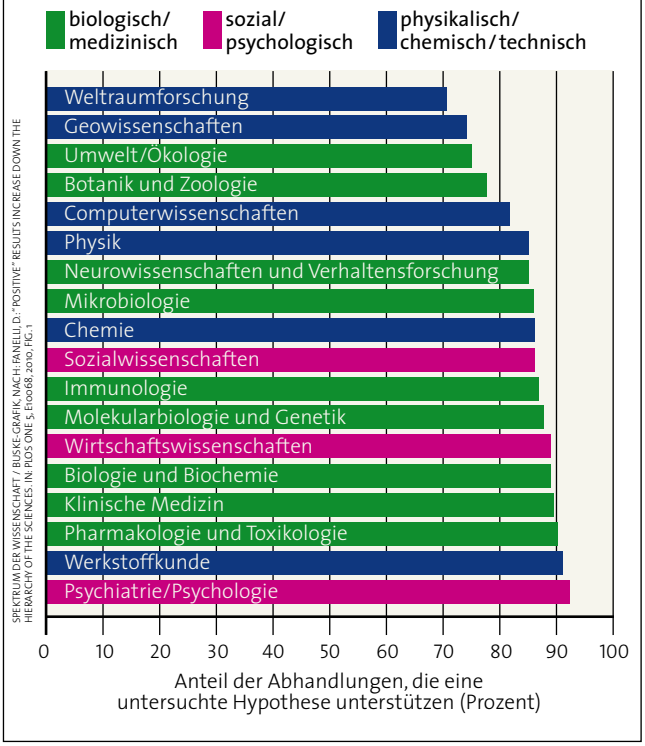
So liegt die ganze Beweislast bei denen, die die Studien reproduzieren wollen – eine mühsame und undankbare Aufgabe. Das sieht man beispielhaft an der eingangs erwähnten Studie von Daryl Bem. Die drei Forschergruppen, die erfolglos versucht hatten, seine Ergebnisse nachzuvollziehen, reichten ihre Negativresultate gemeinsam zur Publikation ein. Von der Redaktion des »Journal of Personality and Social Psychology« erhielten sie daraufhin die Auskunft, dass die Zeitschrift keine direkten Reproduktionen veröffentliche. Die Redaktionen von »Science« und »Psychological Science« äußerten sich ähnlich. Das »British Journal of Psychology« ließ die Abhandlung von Fachleuten prüfen, lehnte sie jedoch ebenfalls ab; Bem war einer der Gutachter. Am Ende erschien die Arbeit in »PLoS One«, einem Journal, dass laut eigenem Bekunden technisch einwandfreie Studien veröffentlicht, nicht nur neue.

»Ich habe alles Mögliche getan, um das Reproduzieren meiner Studie zu unterstützen«, erklärt Bem, der zu seinen umstrittenen Ergebnissen steht und sämtliche einschlägigen Methoden und Tests ins Internet gestellt hat. Er fügt jedoch hinzu, dass eine einzelne publizierte Arbeit, die zu einem Negativergebnis komme, für ihn noch nicht aussagekräftig sei. »Es ist noch zu früh«, meint er, »es kann Jahre dauern, herauszufinden, warum eine Wiederholung gescheitert oder gelungen ist. Man braucht dafür eine Metaanalyse vieler Versuche.«

Der Kognitionspsychologe Stéphane Doyen von der Université Libre de Bruxelles stieß auf ähnliche Probleme, nachdem es ihm und seinen Kollegen nicht gelungen war, ein klassisches Experiment zu reproduzieren. John Bargh, Psychologe an der Yale University in New Haven (USA), hatte gezeigt, dass Menschen langsamer gehen, wenn sie zuvor unbewusst mit altersbezogenen Wörtern in Berührung kommen. Doyens Arbeitsgruppe scheiterte daran, dieses Ergebnis zu bestätigen, und wollte damit an die Öffentlichkeit gehen – kassierte aber von mehreren Fachzeitschriften eine Absage. Schließlich erschien auch ihre Arbeit in »PLoS One«. Daraufhin schrieb Bargh einen erzürnten Blogbeitrag, in dem er Do-

Das Positive betonen

Eine fachübergreifende Analyse der wissenschaftlichen Literatur im Jahr 2010 belegte die Tendenz, nur »positive« Studien zu veröffentlichen – also solche, die die geprüfte Hypothese stützen. Am stärksten ausgeprägt ist dieser Hang in der Psychiatrie und der Psychologie. Insgesamt wurden 2434 Studien untersucht.



SPEKTRUM DER WISSENSCHAFT / BUSKE-GRAFIK, NACH: FANELLI, D.: "POSITIVE" RESULTS INCREASE DOWN THE HIERARCHY OF THE SCIENCES. IN: PLOS ONE 5, E10068, 2010, FIG. 1

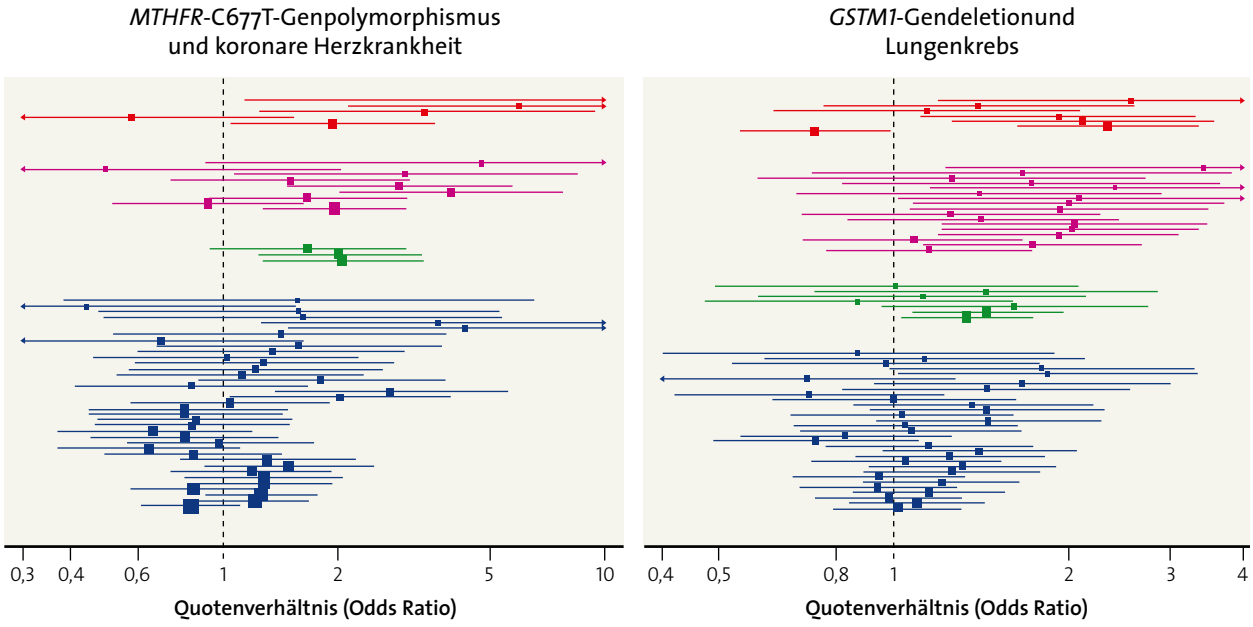
Starke Wirkungen aus Fernost

Studien, die neue Thesen prüfen oder unbekannte Phänomene untersuchen, arbeiten oft mit relativ kleinen Stichproben. Deshalb kann der von ihnen gefundene Effekt in einem relativ breiten Bereich streuen. In die Fachpublikationen gelangen vor allem solche Arbeiten, bei denen ein deutliches Resultat herausgekommen ist, auch wenn dieses möglicherweise auf Zufall beruht. Folgeuntersuchungen, die das Ergebnis zu reproduzieren versuchen, stellen oft einen schwächeren Effekt fest – insbesondere, wenn sie mit größeren Stichproben arbeiten und ihr Resultat daher näher am wahren Mittelwert liegt.

Doch nicht nur statistisch bedeutsame Variablen wie der Stichprobenumfang beeinflussen das Ergebnis einer Untersuchung. Auch die Nationalität der Forscher spielt eine Rolle, wie Zhenglun Pan vom Shandong Provincial Hospital (China) und seine Kollegen entdeckt haben. Sie analysierten zahlreiche Arbeiten, deren Ziel darin bestanden hatte, medizinische Studien aus der Fachliteratur zu reproduzieren. Das verblüffende Ergebnis: Bei Arbeiten von asiatischen Forschern kam oft ein größerer Effekt heraus als bei Untersuchungen, die Forscher aus westlichen Ländern durchgeführt hatten. Am deutlichsten war dies bei chinesischen Wissenschaftlern: Ihre Studien hatten die deutlichsten Resultate ergeben.

Die Gründe hierfür sind unklar. Vermutlich haben sie mit einer regional unterschiedlich stark ausgeprägten Bevorzugung von Positivergebnissen (»Publikationsbias«) zu tun, möglicherweise auch mit kulturellen Eigenheiten. *fs*

Pan, Z.: Local Literature Bias in Genetic Epidemiology: An Empirical Evaluation of the Chinese Literature. In: Public Library of Science Medicine, 2: e334, 2005



Forscher aus verschiedenen Ländern haben den Zusammenhang zwischen bestimmten Erbanlagen und Krankheiten untersucht. Jeder waagrechte Strich steht für eine Studie. Das Quadrat markiert den jeweils ermittelten Wert dafür, wie die Genvariante mit der Krankheit assoziiert ist. Quotenverhältnis = 1: nicht korreliert; > 1: positiv korreliert; < 1: negativ korreliert. Die Länge der Striche ist ein Maß für die Unsicherheit, die Größe des Quadrats steht für den Umfang der Studie. Orange und violett: chinesische Autoren; grün: andere asiatische Autoren; blau: Autoren aus anderen Ländern.

SPEKTRUM DER WISSENSCHAFT / BUSKE-GRAFIK, NACH: PAN Z.: LOCAL LITERATURE BIAS IN GENETIC EPIDEMIOLOGY (...). IN: PLOS MED 2: E334, 2005, FIG. 2

yen und seine Mitarbeiter als inkompetent bezeichnete. Später legte er sich auch mit dem Autor dieses Artikels an, der über die Kontroverse berichtet hatte. Heute erklärt Bargh seine heftige Reaktion wie folgt: Er habe die These, dass unbewusste Denkvorgänge eine wichtige Rolle spielen, wachsender Skepsis ausgesetzt gesehen und den Eindruck gewonnen, dass hier dem Fachgebiet geschadet werde.

Natürlich kann ein einzelner missglückter Wiederholungsversuch ein Forschungsergebnis kaum widerlegen. Sein Negativergebnis kann ganz banale Ursachen haben. Wenn der zu prüfende Effekt beispielsweise sehr schwach ist, lässt er sich schon aus statistischen Gründen schwer nachweisen. Oder die Teilnehmer eines Wiederholungsexperiments unterscheiden sich von denen des Originalversuchs. Oder einem Team fehlen schlicht die handwerklichen Fähigkeiten, ein Experiment zu reproduzieren.

»Das Durchführen subtiler Versuche ähnelt in vielerlei Hinsicht der Regie bei einer Theateraufführung«, sagt Daniel Kahneman, nobelpreisgekrönter Psychologe an der Princeton University. Simple Details wie der Wochentag oder die

Farbe der Tapete können die Ergebnisse einer psychologischen Untersuchung beeinflussen, doch tauchen sie niemals im Methodenteil eines Fachartikels auf. Bargh argumentiert beispielsweise, Doyens Team habe die Versuchsteilnehmer zu vielen altersbezogenen Wörtern ausgesetzt, so dass diese auf den verborgenen Zweck des Experiments aufmerksam wurden. In Bahnungs- oder Primingstudien müsse man einen unbewusst einwirkenden Reiz äußerst genau justieren, erläutert Kahnemann: intensiv genug, dass er noch wirken kann, aber wiederum nicht so ausgeprägt, dass er den Versuchsteilnehmern auffalle. Ob das gelungen sei, zeige sich erst, wenn der gefundene Effekt in hunderten Experimenten bestätigt wurde.

»Man beweist nicht mit B, dass A richtig ist. **Man wiederholt dafür A**«

Eine Prüfmethode, die nur bejahen kann

Dutzende Forscherteams ließen sich von John Barghs klassischem Experiment dazu anregen, eigene Primingversuche durchzuführen. Bei solchen Experimenten untersuchen die Wissenschaftler, ob ein vorangehender Reiz gewisse Gedächtnisinhalte aktiviert und dadurch die Verarbeitung nachfolgender Reize beeinflusst beziehungsweise den Gemütszustand oder das Verhalten prägt. Es zeigte sich beispielsweise, dass Versuchsteilnehmer ihre Gesprächspartner ernster nahmen und soziale Probleme als schwerwiegender empfanden, wenn sie vorher ein schweres Manuskript-Klemmbrett gehalten hatten. Und wurden Personen mit Begriffen konfrontiert, die mit Sauberkeit und Reinlichkeit assoziiert waren, beurteilten sie niederträchtige Taten anschließend nachsichtiger.

Dass unbewusste Denkvorgänge eine wichtige Rolle spielen, lässt sich also mit verschiedenen Experimenten und empirischen Methoden bestätigen. Solche konzeptuellen Replikationen sind für die Psychologie sehr nützlich. »Man sollte meinen, dass eine konzeptuelle Replikation schwerer wiegt als eine exakte Wiederholung, denn sie ist ein stärkerer Beweis dafür, dass sich der Effekt verallgemeinern lässt«, sagt der Psychologe Eliot Smith von der Indiana University Bloomington, USA, der das »Journal of Personality and Social Psychology« mitherausgibt. Andere Psychologen sind da skeptischer. »Man kann kein Konzept nachahmen«, meint etwa Chambers, »das ist zu subjektiv. Es ist völlig offen, wie ähnlich etwas sein muss, um als konzeptuelle Replikation zu gelten.« Außerdem erzeuge dieses Vorgehen eine »logische Doppelmoral«. Beeinflusst beispielsweise ein schweres Klemmbrett unbewusst das menschliche Urteilsvermögen, dann könnte man das als indirekte Bestätigung des Bargh-Experiments werten, demzufolge Menschen nach dem Priming mit altersbezogenen Begriffen langsamer gehen. Hat das Gewicht des Klemmbretts jedoch keinen Einfluss, dann würde niemand behaupten, Barghs Befund sei indirekt widerlegt worden.

Die konzeptuelle Replikation ermöglicht es also, Versuchsresultate angeblich zu bestätigen, aber nicht, sie zu entkräften. Auf diese Weise erlaubt sie den Forschern, ein Kartenhaus aus zweifelhaften Ergebnissen zu errichten, die sich gegenseitig stützen. »Sie ist die wissenschaftliche Verkörperung des Bestätigungsfehlers – also der Neigung, Informationen so auszuwählen, dass sie die eigenen Erwartungen erfüllen«, meint Brian Nosek, Sozialpsychologe an der University of Virginia. »Die Psychologie ist in gewisser Weise auf konzeptuelle Replikationen angewiesen, doch diese ersetzen keine direkten Reproduktionen. Man beweist nicht mit B, dass A richtig ist. Man wiederholt dafür A.«

In dem Umfeld, das die Publikations- und Forschungspraktiken der Psychologen erzeugen, können Fehltritte leicht unbemerkt bleiben. Im November 2011 wurde dem Sozialpsychologen Diederik Stapel von der niederländischen Tilburg University, der bis dahin als aufsteigender Stern seines Fachgebiets gegolten hatte, wissenschaftlicher Betrug in großem Umfang vorgeworfen, den er schließlich auch eingestand. Stapel hatte mehrere Aufsehen erregende Studien veröffentlicht, die unter anderem belegten, dass eine unordentliche Umgebung – etwa ein heruntergekommener Bahnhof – diskriminierendes Verhalten fördert. Es gab zahlreiche Faktoren, die eine Wiederholung der Versuche erschwerten, und das half ihm, seinen Betrug zu verstecken. Die wissenschaftliche Untersuchungskommission schrieb: »All diese äußerst stimmig wirkenden Resultate hätten eigentlich nachdenklich machen müssen, doch sie wurden begeistert

Reproduktion als Dienstleistung

Wo viele Menschen das gleiche Problem haben, dauert es meist nicht lange, bis jemand eine Geschäftsidee daraus entwickelt. So auch beim Prüfen von Forschungsergebnissen. Untersuchungen haben gezeigt, dass zahlreiche biomedizinische Arbeiten, die in der Fachliteratur erscheinen, sich nicht reproduzieren lassen. Nun bietet das US-Unternehmen »Science Exchange« an, wissenschaftliche Experimente gegen Gebühr zu wiederholen. Bestätigt sich der gefundene Effekt dabei, macht dies die Originalstudie vertrauenswürdiger.

Interessierte Forscher können ihre Arbeit bei »Science Exchange« einreichen. Das Unternehmen sucht dann per Zufallsprinzip Experten aus, die den Versuch unabhängig und im Blindversuch nachstellen. Die Ergebnisse werden an die Autoren der Originalstudie zurück übermittelt. Diese können das Datenmaterial als zusätzliche Publikation in einer Sonderausgabe der Fachzeitschrift »PloS One« veröffentlichen. Erfolgreich reproduzierte Studien erhalten ein Zertifikat. *fs*

Mehr unter: www.reproducibilityinitiative.org

W I S wissenschaft in die schulen!

Didaktische Materialien für den Unterricht zum Thema »Verstehen und Lesen von wissenschaftlichen Artikeln« kostenfrei herunterladen unter:

www.wissenschaft-schulen.de/lesekompetenz

aufgenommen ... Wer erfolglos versuchte, die Ergebnisse zu reproduzieren, ging davon aus, dass ihm einfach Herrn Stapels Begabung fehle.« Die Fälschungen betreffen mindestens 30 Fachartikel.

Stapels Geschichte gleicht derjenigen der Psychologen Karen Ruggiero und Marc Hauser von der Harvard University in Cambridge, Massachusetts. Auch sie veröffentlichten jeweils stark beachtete Ergebnisse zu Diskriminierung und Moralverhalten. Ruggiero wurde im Jahr 2001 des Forschungsbetrugs überführt, Hauser befand man 2010 des Fehlverhaltens für schuldig. Beide wurden von Informanten aus ihrem engeren Umfeld bloßgestellt – wie übrigens auch Diederik Stapel. »Wenn der Wissenschaftsapparat sich wirklich selbst korrigiert, warum hat dieser Korrekturmechanismus dann bei keinem einzigen von den dreien funktioniert?«, fragt Nosek.

Viele Psychologen suchen nun nach neuen Wegen, um das Reproduzieren von Studien zu unterstützen. Im vergangenen Jahr haben Hal Pashler, Psychologe an der University of California in San Diego, und seine Kollegen die Website www.psychfiledrawer.org erstellt, auf der Psychologen unveröffentlichte Reproduktionsversuche publizieren können, egal ob diese erfolgreich waren oder nicht. Die Website stieß auf große Zustimmung, verfügt bisher (Dezember 2012) aber nur über 18 Einträge. Es gibt kaum Anreize für eine Veröffentlichung: Jeder Eintrag setzt die Wissenschaftler der Kritik ihrer Kollegen aus, trägt aber nicht dazu bei, ihre Publikationsliste aufzuwerten.

Wenn die Jungen die Alten kontrollieren

Der Sozialpsychologe Matthew Lieberman von der University of California, Los Angeles, schlägt einen anderen Ansatz vor: »Die führenden Hochschulen könnten von ihren Psychologiestudenten verlangen, eine Studie in ihrem Fachgebiet zu reproduzieren.« Auf diese Weise, meint er, könnten die Studenten ihre Fähigkeiten schulen und wertvolle Veröffentlichungen sammeln – und die Wissenschaft würde von der Fleißarbeit enorm profitieren.

Wagenmakers wiederum vertritt die Auffassung, dass Reproduktionsversuche ähnlich wie klinische Studien vorher angemeldet werden sollten, um das Zurechtbiegen von Daten zu erschweren. Zudem solle ihre Veröffentlichung unabhängig vom Ergebnis erfolgen. Der Kontakt zu den Autoren des Originalversuchs, oder sogar eine Zusammenarbeit mit ihnen, könne späteren Streitereien um die eingesetzten Methoden vorbeugen.

Unterdessen bezweifeln einige Wissenschaftler immer noch, dass überhaupt ein Problem besteht. Selbst der kritische Sozialpsychologe Brian Nosek betont, dass es keine belastbaren Schätzungen zur Häufigkeit falscher Positivergebnisse gibt. Ende 2011 rief er ein Team von Psychologen zusammen, das sämtliche Studien zu reproduzieren versucht, die seit 2008 in den drei wichtigsten Psychologie-Fachzeitschriften veröffentlicht wurden. Die Wissenschaftler wollen sich so genau wie möglich an die Originalarbeiten halten und versuchen, mit deren Autoren zusammenzuarbeiten. Das Ziel lautet nicht, einzelne Forscher mit ihren Arbeiten an den Pranger zu stellen, sondern für das gesamte Forschungsgebiet »erste Hinweise auf die Erfolgswahrscheinlichkeit von Reproduktionen zu erhalten«, wie sich Nosek ausdrückt.

Manche Forscher wollen es gar nicht so genau wissen und sind an dem Ergebnis nicht interessiert. Psychologe Hal Pashler erwartet indessen, dass seine Befürchtungen bestätigt werden – dass sich also die informellen Flurgespräche über nicht nachvollziehbare Studien und über zahllose fehlgeschlagene Versuche, deren Protokolle in Schubladen verschwinden, als wahr erweisen. »Dann kann es keiner mehr leugnen.« ~

DER AUTOR

Ed Yong lebt als freier Schriftsteller in London und betreibt den Blog »Not Exactly Rocket Science« (sinngemäß »Nicht wirklich geniale Wissenschaft«).

QUELLEN

Bem, D.J.: Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect. In: Journal of Personality and Social Psychology 100, S. 407–425, 2011

Doyen, S. et al.: Behavioral Priming: It's All in the Mind, but Whose Mind? In: PLoS One 7, e29081, 2012

Fanelli, D.: Negative Results are Disappearing from Most Disciplines and Countries. In: Scientometrics 90, S. 891–904, 2012

Ioannidis, J.P.: Why Most Published Research Findings are False. In: PLoS Medicine 2, e124, 2005

Simmons, J.P. et al.: False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. In: Psychological Science 22, S. 1359–1366, 2011

WEBLINK

Diesen Artikel sowie weiterführende Informationen finden Sie im Internet: **www.spektrum.de/artikel/1178943**

© Nature Publishing Group
www.nature.com
Nature 485, S. 298–300, 17. Mai 2012