


PSYCHOLOGIE HEUTE

GESELLSCHAFT

Versuch und Irrtum

Psychologie ist eine empirische Wissenschaft: ohne Experimente keine Erkenntnisse. Doch die Kritik an den Methoden mehrt sich. Wie verlässlich ist die Forschung wirklich?

VON JOCHEN PAULUS

 01. FEB 2017

Draußen lärmt eine Motorsäge, drinnen sehen Studenten eine Dokumentation über das bescheidene Leben armer Juden. Anschließend müssen sie bewerten, wie interessant der Film war und wie sehr er andere berühren würde. Am Ende entschuldigt sich der Versuchsleiter für den Krach und lässt die Studenten noch notieren, ob dieser ihr Urteil beeinflusst hat. Die meisten sind davon überzeugt. Doch sie irren sich: Sahen Versuchspersonen den Film ohne Lärm, beurteilten sie ihn ganz genauso.

Die Filmvorführung mit Motorsäge war nur eines der Experimente, mit denen Richard Nisbett und Timothy Wilson vor fast 40 Jahren bewiesen: Es nützt oft wenig, Menschen zu fragen, was in ihnen vorgeht. Denn sie wissen es häufig nicht. Es braucht Experimente. Ohne sie lässt sich kaum feststellen, was Ursache welcher Wirkung ist. Ein Beispiel: Warum kann Alkohol aggressiv machen? Liegt das an seiner pharmakologischen Wirkung? Oder werden manche Angetrunkene ausfällig, weil sie ihre

Rüpeleien auf den Schnaps schieben können? Um das zu untersuchen, bat Laurent Bègue von der Universität Grenoble 117 Männer zu einem angeblichen Geschmackstest.

Jeder bekam einen Fruchtcocktail serviert, der angeblich viel, wenig oder keinen Alkohol enthielt. Allerdings stimmten die Angaben, die die Teilnehmer erhielten, meist nicht. Manche bekamen beispielsweise keinen Alkohol, glaubten es aber, und umgekehrt. Im Verlauf des Experiments wurde jeder Proband mehrmals von einem vermeintlichen anderen Teilnehmer provoziert, der in Wirklichkeit ein Komplize des Versuchsleiters war. Abschließend kippte der Provokateur jede Menge Salz und Tabasco in einen Kartoffelbrei, den der Teilnehmer essen musste. Dieser durfte seinerseits die Mahlzeit für den Komplizen würzen. Und siehe da: Je mehr Alkohol ein Teilnehmer im Blut zu haben glaubte, umso mehr ließ er seinen Rachegeilüsten freien Lauf – auch wenn er tatsächlich keinen Alkohol intus hatte.

Mit Experimenten lassen sich einzelne Faktoren wie der wahrgenommene und der tatsächliche Alkoholspiegel gezielt beeinflussen und auf eine zuvor formulierte Hypothese prüfen. Im wirklichen Leben kommen zu viele Faktoren zusammen, die sich nicht isoliert betrachten und unter wiederholbaren Bedingungen manipulieren beziehungsweise kontrollieren lassen, wie es bei solchen Untersuchungen der Fall ist. Experimente finden deshalb oft in einer sehr künstlichen Umgebung statt – in einem Labor zum Beispiel, in dem allein der Versuchsleiter bestimmt, was der Proband sieht und hört. Doch auch im Alltag lässt sich experimentieren, solange der Forscher die jeweils wichtigen Variablen kontrollieren kann.

Berühmt für solche sogenannten Feldexperimente ist Robert Cialdini, ein inzwischen emeritierter Psychologieprofessor der *Arizona State University*. Eines seiner Experimente fand im *Petrified Forest* statt, einem Nationalpark in Arizona, dessen Attraktion zahllose versteinerte

Holzstücke sind. Leider stahlen Besucher immer wieder solche Kostbarkeiten. Die Parkverwaltung hatte darum ein Schild aufgestellt, das die Diebstähle beklagte, die sich auf 14 Tonnen im Jahr summierten. Cialdini hielt das für genau die falsche Botschaft. Die Besucher folgerten daraus, dass Diebstahl im Park normal sei, und verhielten sich entsprechend, so seine Vermutung. Um diese zu überprüfen, stellte er vor manchen Waldwegen solche Schilder auf, vor anderen keine. Als er nach einigen Wochen nachzählte, wie viele der von ihm markierten Versteinerungen verschwunden waren, zeigte sich: Bei den gutgemeinten Schildern hatte sich die Zahl der Diebstähle verdreifacht.

Zweifel an der Übertragbarkeit

Die Methode des Experiments ist die übliche Vorgehensweise in der Psychologie. Doch in den vergangenen Jahren sind so viele psychologische Experimente in die Kritik geraten, dass zahlreiche Forscher eine Krise ihres Faches und seiner Glaubwürdigkeit konstatieren. Es gibt vor allem zwei wichtige Kritikpunkte.

Da ist zum einen die Frage nach der Übertragbarkeit aus der Laborsituation: Oft finden Experimente vor einem Computerbildschirm statt, und nicht selten werden die Versuchspersonen – die meist Studenten sind – schlicht gefragt, wie sie in einer bestimmten Situation reagieren würden. Was die Antworten über das echte Leben aussagen, bleibt zweifelhaft. Doch vor allem Fragebogenstudien lassen sich schnell durchführen und veröffentlichen. Wer Professor werden will, ist darauf angewiesen, denn die Berufungskommissionen der Universitäten legen großen Wert auf eine lange Publikationsliste. „Die Probleme, Verhalten zu beobachten, selbst wenn es möglich wäre, können es schwieriger machen, das Datenvolumen auszustoßen, das akademischer Erfolg heute erfordert“, kritisierte beispielsweise ein Team um den Psychologieprofessor Roy Baumeister, damals an der *Florida State University*, im Fachblatt *Perspectives on Psychological Science*.

Die schnelle Massenproduktion von Studien ließ zudem Zweifel an den Ergebnissen aufkommen. 270 Forscher haben daher in einem Großversuch 100 Experimente wiederholt, die 2008 in drei angesehenen Fachzeitschriften erschienen waren. Die Bilanz veröffentlichten sie im August 2015 im Wissenschaftsmagazin *Science*: Nur 36 Prozent der Ergebnisse ließen sich demnach mit der zuvor in der Originalstudie berichteten statistischen Bedeutsamkeit bestätigen. In der Sozialpsychologie, die sich mit Beziehungen zwischen Menschen beschäftigt, waren es sogar lediglich 25 Prozent. Die replizierenden Forscher hatten die Aufgabe gehabt, ein Experiment von Kollegen so genau wie möglich zu wiederholen. Dafür stellte ein Teil der Autoren der ursprünglichen Untersuchungen auch die Materialien zur Verfügung, mit denen sie geforscht hatten, etwa Texte und Filme – ein Faktor übrigens, der die Replizierbarkeit der Experimente positiv beeinflusste.

Seither ist international viel darüber diskutiert worden, welche Schlussfolgerungen die *Science*-Offensive erlaubt. Der Präsident der amerikanischen *Association for Psychological Science* etwa sprach von einer „sehr guten Studie“. Doch nicht alle Forscher sind überzeugt, dass die Replikationen genau genug waren. Ein Team um den Harvard-Psychologen Daniel Gilbert nennt als Beispiel eine Originalstudie, bei der US-Studenten ein Video gezeigt worden war, in dem andere US-Studenten über Quoten für Afroamerikaner bei der Universitätszulassung diskutierten. Wiederholt wurde der Versuch jedoch in Amsterdam – mit dem Originalvideo aus den USA, obwohl sich solche Fragen in den Niederlanden womöglich anders stellen. Gilbert argumentiert, dass die Ergebnisse sich womöglich wegen solcher Abweichungen nicht replizieren ließen. Ähnliche Einwände erhebt er gegen fünf weitere Wiederholungsexperimente.

Die Organisatoren der *Science*-Studie halten dagegen: Die Hälfte der von Gilbert monierten Unterschiede bei sechs Studien sei von den ursprünglichen Forschern abgesegnet worden. Und eine weitere

kritisierte Replikation habe die Resultate einer US-Rassismus-Studie bestätigt – in Italien. Zwar räumen auch die Kritiker ein, dass man die Experimente eventuell noch genauer wiederholen müsste. Eine „exakte“ Replikation könne es aber ohnehin niemals geben. Schließlich seien die Umstände zwangsläufig immer andere, vom Untersuchungsleiter über die Teilnehmer bis zu den technischen Bedingungen, dem Wetter, dem historischen Kontext und so weiter. Dennoch liefere eine hinreichend ähnliche erfolgreiche Wiederholung einen Hinweis darauf, wie verallgemeinerbar ein Ergebnis sei. Und eine Abweichung der Ergebnisse sei immer Anlass, sich auf die Suche nach den Gründen zu begeben, warum es sie gibt, und diese dann zu überprüfen.

Auch viele andere Untersuchungen, die beim Großversuch für *Science* nicht dabei waren, ließen sich nicht bestätigen. Verhalten sich Menschen großzügiger, wenn sie subtil an Gott erinnert werden? Wird mehr Testosteron ausgeschüttet, wenn man eine Pose einnimmt, die Macht signalisiert? Sind Kinder und Erwachsene mit zwei Muttersprachen im Vorteil, wenn es gilt, bei bestimmten Aufgaben falsche Antworten zu vermeiden? Macht der Anblick von Geld Menschen egoistischer? Zu alledem wurden Experimente durchgeführt und deren Resultate mit Überzeugung verkündet, doch teilweise großangelegte Versuche, die Ergebnisse zu bestätigen, scheiterten später.

Die Dinge sind oft nicht, wie sie scheinen

„Ich liebe die Sozialpsychologie immer noch“, schreibt der vielfach preisgekrönte Psychologieprofessor Michael Inzlicht vom *Toronto Laboratory for Social Neuro-science* in einem Blogbeitrag zum Thema. „Aber ich bin bestürzt, dass so viele von uns die kleinen (und nicht so kleinen) Anzeichen dafür ignorieren oder zu rechtfertigen versuchen, dass die Dinge einfach nicht richtig sind, dass sie nicht sind, was sie scheinen.“

Inzlicht hat viel zum Phänomen der Selbsterschöpfung (Engl. *ego depletion*) geforscht. In solchen Experimenten erhalten Versuchspersonen zunächst eine Aufgabe, bei der sie sich selbst kontrollieren müssen. Sie sollen beispielsweise eine langweilige historische Schilderung so vorlesen, als wäre sie hochdramatisch, und später ein Buch mit halbnackten Frauen und Männern durchblättern. Die Erschöpften sehen dabei länger hin als andere, deren Selbstkontrolle nicht zuvor geschwächt wurde. Viele Experimente nach der gleichen Logik kommen zu ähnlichen Ergebnissen. Aber stimmen die Resultate? „Das Problem ist, dass es Selbsterschöpfung vielleicht gar nicht gibt“, sagt Inzlicht. Denn ein großer Replikationsversuch des grundlegenden Effekts „mit mehr als 2000 Teilnehmern fand nichts, nada, null“.

Dabei wurde die Selbstkontrolle am Computer getestet, weil der Versuch in 24 Laboren weltweit identisch ablaufen sollte. Zunächst musste ein Teil der Versuchspersonen immer dann einen Knopf drücken, wenn auf dem Bildschirm ein Wort mit dem Buchstaben „e“ aufleuchtete – außer wenn das „e“ maximal zwei Buchstaben entfernt von einem Vokal stand. Wer das wie in dem Versuch über siebeneinhalb Minuten macht, dessen Selbstkontrolle ist geschwächt, so die Überlegung. Im zweiten Teil des Experiments leuchteten immer drei Ziffern gleichzeitig auf, von denen zwei identisch waren und die dritte nicht. Die Teilnehmer sollten per Tastendruck die abweichende Ziffer benennen. Das war nicht ganz so einfach: Handelte es sich dabei etwa um eine 2, die aber an erster Stelle stand, mussten die Versuchspersonen nicht die erste Taste drücken, sondern die zweite. Das erfordert Selbstkontrolle und sollte Probanden daher noch schwerer fallen, wenn sie sich vorher mit den Buchstaben abgemüht hatten. So war es zumindest in der Originalstudie. Doch dieses Ergebnis bestätigte sich zum Kummer Inzlichts nicht. Trotzdem ist er dankbar für die Erfahrung: Der Misserfolg lehre wichtige Lektionen, schreibt er in seinem Blog – „wenn wir sie hören wollen“.

Warum lassen sich psychologische Experimente oft nicht erfolgreich wiederholen? Eine Antwort auf diese Frage liefert der Soziologe David Peterson am Beispiel von Studien aus der Babyforschung, die an mehreren US-Universitäten realisiert wurden. In der Theorie hätten die Wissenschaftler in diesen Untersuchungen immer eine Hypothese entwickeln und sie dann in einem durchgeplanten Experiment überprüfen müssen.

Tatsächlich führten sie die Studie aber nach Petersons Beobachtung oft nur durch, wenn die ersten Resultate gut aussahen. Anderenfalls seien die Versuchsbedingungen geändert worden – so lange, bis die Ergebnisse brauchbar erschienen. Ein Jungforscher erläuterte Peterson die Strategie so: „Wir haben ein paar unausgegorene Ideen. Wir machen ein paar Experimente. Was für Daten dabei auch immer herauskommen – wir tun so, als hätten wir danach gesucht.“ So entstünden oft selektiv herausgegriffene Zufallsbefunde, die andere Forscher nicht replizieren können, so Peterson.

Wie verbreitet solche unsauberen Methoden sind, ist umstritten. Eine amerikanische Umfrage unter forschenden Psychologen kam auf extrem hohe Werte. Eine etwas anders angelegte Wiederholung, an der sich gut 1100 deutschsprachige Wissenschaftler beteiligten, zeigte weniger, aber immer noch bedenklich viele der zweifelhaften Praktiken auf. So gestanden mehr als 40 Prozent der Teilnehmer, schon einmal nur einen Teil ihrer Daten zu einem Befund veröffentlicht zu haben. Fast die Hälfte hatte mindestens einmal fälschlicherweise angegeben, ein überraschendes Ergebnis genau so erwartet zu haben.

Ist gegenseitige Überwachung und Kontrolle die Lösung?

Weil auch die Mogler nicht in jeder Untersuchung mogeln, ist der Anteil der betroffenen Experimente am gesamten Studienaufkommen geringer. Einige problematische Vorgehensweisen betreffen aber offenbar

immerhin zehn Prozent. Klaus Fiedler von der Universität Heidelberg, einer der Autoren der Umfrage, argumentiert allerdings, dass diese teils auch gerechtfertigt sein können – etwa das Weglassen von Daten, um die Leser nicht zu langweilen. Überhaupt hält Fiedler die Lage für nicht so ernst wie viele Kollegen. „Wo ist die Krise?“, fragte er bei der großen Tagung experimentell arbeitender Psychologen im vergangenen März in Heidelberg, die sich unter anderem mit diesem Streitthema beschäftigte.

Fiedler war auch der Vorsitzende einer Kommission, die für die Deutsche Gesellschaft für Psychologie (DGPs) den im Oktober 2015 vorgelegten Bericht zur „Qualität in der psychologischen Forschung“ erarbeitete. In dem Papier werden Replikationsstudien zwar befürwortet, doch warnt die Kommission, „den Fortschritt kreativer Wissenschaft durch defensive, von Misstrauen getriebene Kontrollmaßnahmen zu ersticken“. Bei öffentlichen Vorwürfen und persönlicher Kritik sollen „ganz besonders strenge Kriterien der Validität und sorgfältigen Prüfung“ angelegt werden. „Im Gegensatz zu der Vorstellung eines gegenseitigen Überwachungs- und Kontrollnetzes“ solle sich jeder einzelne Wissenschaftler an die Maxime der Selbstüberwachung und Selbstkontrolle als Grundprinzip halten.

Die wahrscheinlich wirksamste Einzelmaßnahme zur Verbesserung der Qualität von Studien wird in dem DGPs-Papier zwar unter anderem empfohlen, aber nicht verlangt: die Registrierung von Experimenten, bevor sie umgesetzt werden. Bei diesem Verfahren beschreiben die Forscher beispielsweise im Internet genau, wie sie ihre Untersuchung durchführen und auswerten wollen. Damit soll verhindert werden, dass sie anschließend so lange herumprobieren, bis die Ergebnisse passen, oder unangenehme Resultate einfach unerwähnt bleiben.

Führende Medizinzeitschriften wie beispielsweise das *New England Journal of Medicine* und *The Lancet* verlangen schon seit Jahren eine vorherige Registrierung von klinischen Studien. Inzwischen ermutigen

auch einzelne psychologische Publikationen wie *Psychological Science* und *Experimental Psychology* ihre Autoren zu diesem Schritt, denn so würden die Ergebnisse psychologischer Experimente vermutlich verlässlicher. Das wäre wichtig, denn ohne Experimente werden sich psychologische Fragen auch in Zukunft nicht klären lassen. Eine alternative Methode gibt es schlicht nicht.

Wie sich vertrauenswürdige Resultate erzielen lassen, zeigt das Beispiel einer Zusammenarbeit von 23 Wissenschaftlerteams aus acht Ländern. Gemeinsam gelang es ihnen kürzlich, ein spannendes, aber schlecht gesichertes Forschungsergebnis zu bestätigen: Vor einem Vierteljahrhundert hatte Jonathan Schooler, heute an der *University of California* in Santa Barbara, Versuchspersonen ein kurzes Video von einem Bankraub vorgeführt. Später bat er sie, aus einer Sammlung von Fotos das Bild des Diebes auszuwählen. Die Hälfte der Teilnehmer hatte dessen Aussehen vorher schriftlich beschreiben müssen. Doch anstatt den Verbrecher dadurch leichter wiederzuerkennen, störte es die visuelle Erinnerung offenbar, wenn sie die Gesichtszüge verbal memoriert hatten. Diese Versuchspersonen erkannten den Täter danach schlechter wieder als Teilnehmer, die sich zwischendurch mit einer anderen Aufgabe beschäftigt hatten.

Die Studie wurde oft zitiert und steht in vielen Lehrbüchern; bei einigen, allerdings oft kleinen Wiederholungen in den folgenden Jahren bestätigten sich die Ergebnisse aber nicht immer. Deshalb rief die Fachzeitschrift *Perspectives on Psychological Science* Forscher vor einigen Jahren zur Wiederholung des Experiments auf. Die Replikationsvorhaben wurden vorab erfasst, damit kein Resultat verschwiegen werden konnte. Das ursprüngliche Ergebnis hielt der Wiederholung stand. Es empfiehlt sich offenbar nicht, Zeugen eines Verbrechens vor einer Gegenüberstellung um eine genaue Beschreibung zu bitten. Ob diese so abgesicherte Erkenntnis im Zweifel auch abseits des Labors, im echten Leben gilt, lässt sich daraus nicht schließen. Diese

Frage könnte ein entsprechendes Feldexperiment beantworten, dessen wissenschaftliche Standards sich im Polizeialltag allerdings schwerlich einhalten lassen.

Literatur

Victoria K. Alogna u. a.: Registered replication report: Schooler & Engstler-Schooler (1990). *Perspectives on Psychological Science*, 2014, 9/5, 556–578. DOI: 10.1177/1745691614545653

Christopher J. Anderson u. a.: Response to comment on “Estimating the reproducibility of psychological science”. *Science*, 351/ 6277, 2016, 1037. DOI: 10.1126/science.aad9163

Roy F. Baumeister u. a.: Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, 2/4, 2007, 396–403. DOI: 10.1111/j.1745-6916.2007.00051.x

Laurent Bègue u. a.: A message in a bottle: Extrapharmacological effects of alcohol on aggression. *Journal of Experimental Social Psychology*, 2009, 45/1, 137–142. DOI: 10.1016/j.jesp.2008.07.018

Klaus Fiedler, Norbert Schwarz: Questionable research practices revisited. *Social Psychological and Personality Science*, 2016, 7/1, 45–52. DOI: 10.1177/1948550615612150

Daniel Gilbert u. a.: Comment on “Estimating the reproducibility of psychological science”. *Science*, 351/6277, 2016, 1037. DOI: 10.1126/science.aad7243

Michael Inzlicht: Reckoning with the Past. michaelinzlicht.com/getting-better/2016/2/29/reckoning-with-the-past, 29. Februar 2016.

Don A. Moore: Preregister if you want to. *American Psychologist*, 71/3, 2016, 238–239. DOI: 10.1037/a0040195

Richard Nisbett, Timothy Wilson: Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84/3, 1977, 231–259. DOI: 10.1037/0033-295X.84.3.231

Open Science Collaboration: Estimating the reproducibility of psychological science. *Science*, 349/6251, 2015. DOI: 10.1126/science.aac4716

David Peterson: The baby factory. Difficult research objects, disciplinary standards, and the production of statistical significance. *Socius: Sociological Research for a Dynamic World*, 2016/2, 1–10. DOI: 10.1177/2378023115625071

Den kompletten Artikel können Sie bei uns kaufen oder freischalten.
