

Das Elend der Statistik

In einem internationalen Großprojekt haben Wissenschaftler hundert psychologische Studien noch einmal durchgeführt. Das Ergebnis ist verblüffend. Von Martin Koch



Grafik: 123rf/route55

Jeder Mensch, der für sich beansprucht, Wissenschaft zu betreiben, ist an ein strenges Kriterium gebunden: Er muss seine Ergebnisse so darstellen, dass sie von anderen Forschern jederzeit mit dem gleichen Ergebnis wiederholt werden können, von unvermeidlichen Messfehlern einmal abgesehen. Darin unterscheidet sich die Wissenschaft von allen nichtwissenschaftlichen Formen des Denkens, die sich gewöhnlich mit anekdotischen Belegen begnügen.

Beispiel Astrologie. Dass deren Grundmodell korrekt sei, begründen manche Astrologen damit, dass sie unter anderem den Fall der Berliner Mauer und den Anschlag auf das World Trade Center vorhergesagt hätten. Nachprüfen kann man solche Behauptungen in der Regel nicht, man kann nur an sie glauben. Mitunter sind Astrologen allerdings so unvorsichtig, ihre Prophezeiungen schriftlich festzuhalten. Dann wird offenbar: Die meisten Vorhersagen sind so vage formuliert, dass man im Nachhinein fast immer ein Ereignis finden kann, welches sich mit etwas Fantasie als deren Erfüllung präsentieren lässt.

In der Wissenschaft stoßen astrologische wie auch andere esoterische Aussagen deshalb auf Skepsis und Ablehnung. Zumal darin häufig kausale Zusammenhänge zwischen Dingen postuliert werden, zwischen denen es nach wissenschaftlichem Ermessen keinen kausalen Zusammenhang geben kann.

Gleichwohl sind auch Wissenschaftler nicht davor gefeit, ihre Erkenntnisansprüche zu überdehnen. Denn sie wissen: Mit schrillen kausalen Hypothesen kommt man in die Schlagzeilen und erlangt so zumindest vorübergehend eine gewisse öffentliche Bekanntheit. Darauf hoffte wohl auch der niederländische Sozialpsychologe Diederik Stapel, der im Jahr 2000 als Professor an die Universität Groningen berufen wurde und 2010 den Posten des Dekans an der Tilburg School of Social and Behavioral Sciences übernahm. Während dieser Zeit kamen er und ein Kollege auf die seltsame Idee, auf dem Hauptbahnhof von Utrecht, der wegen eines Streiks mit Müll (und Grafittis) übersät war, einige Dutzend Niederländer nach ihrer Meinung zu Homosexuellen und Muslimen zu befragen. Eine Woche später wiederholten die Forscher die Befragung auf dem mittlerweile gesäuberten Bahnhof. Über ihre Ergebnisse berichteten sie im April 2011 im renommierten US-Fachblatt »Science«. Bei Anwesenheit von Schmutz und Unrat, hieß es da, würden Homosexuelle und Muslime viel stärker mit negativen Eigenschaften belegt als in einer sauberen Umgebung. Oder drastisch formuliert: Dreck und Müll macht Menschen schwulenfeindlicher und rassistischer!

Zeitungen auf der ganzen Welt griffen diese »Botschaft an die Poli-

tik« begierig auf. »Der Spiegel« berichtete darüber ebenso wie die »Frankfurter Rundschau« und die »Neue Zürcher Zeitung«. Was niemand ahnte: Stapel hatte seine Studie mit gefälschten Daten unterlegt. Und das nicht zum ersten Mal. In einer anderen Untersuchung wollte er herausgefunden haben, dass Fleischesser egoistischer und sozialer seien als Vegetarier. Schon der Anblick eines saftigen Steaks machte Versuchspersonen aggressiver, behauptete er. Viele Vegetarier jubelten – zu früh. Die Daten waren manipuliert. Nachdem Stapel sein Fehlverhalten eingestanden hatte, musste er im September 2011 seine Ämter niederlegen. 2013 stand er vor Gericht. Doch das Strafverfahren gegen ihn wurde mit der Auflage, 120 Sozialstunden zu leisten, eingestellt. Damit schieden für viele die Sache erledigt. Nicht so für den US-Wirtschaftsnobelpreisträger Daniel Kahneman. Statt die Stapel-Affäre als bedauerlichen Einzelfall abzutun, forderte er die Sozialpsychologen auf, endlich Ordnung in ihre Studien zu bringen: »Euer Fach ist heute das Paradebeispiel für fragwürdige Forschung in der Psychologie. Hinter dieser Forschung steht ein großes Fragezeichen, und es ist eure Verantwortung, es wieder loszuwerden.«

Nimmt man diese Aussage ernst, dann scheint zumindest in den anderen Disziplinen der Psychologie alles in bester Ordnung zu sein. Dem ist jedoch mitnichten so, wie eine Studie zeigt, die jetzt in »Science« (doi: 10.1126/science.aac4716) veröffentlicht wurde. Sie ist das Resultat eines wahren Mammutprojekts, das der Psychologe Brian Nosek von der University of Virginia in Charlottesville (USA) initiiert hat. Im Rahmen der sogenannten Open Science Collaboration gewann er 270 Forscher aus fünf Kontinenten, deren Aufgabe darin bestand, hundert zufällig ausgewählte psychologische Studien zu wiederholen. Sämtliche Studien waren 2008 in renommierten Fachzeitschriften erschienen. Das Ergebnis fiel nach dreijähriger Arbeit

enttäuschend aus: Bei mehr als der Hälfte der überprüften Studien bzw. Experimente konnte das ursprüngliche Resultat nicht bestätigt werden. So hatten 97 Prozent der Originalstudien statistisch signifikante Ergebnisse erbracht, bei den Wiederholungen lag die Quote nur bei 36 Prozent. Je deutlicher ein Ergebnis im Original ausgefallen war, desto eher führte auch die Wiederholung auf das gleiche, in der Regel allerdings weniger ausgeprägte Resultat. Überraschende Befunde ließen sich dagegen kaum reproduzieren.

»In einer entsprechenden Datenmenge kann man immer statistisch belegbare Zusammenhänge aufspüren, wenn man nur genügend viele Hypothesen ausprobiert.«

Peter C. Austin, University of Toronto

In der Zeitschrift »Psychological Science« wurde beispielsweise behauptet, dass sich verheiratete Frauen an ihren fruchtbaren Tagen zu allein stehenden Männern hingezogen fühlten. Die Erklärungen, die man aus Sicht der Evolution für diesen Befund ersonnen hat, klingen durchaus plausibel. Nur: Bei der Wiederholung der Studie ging der erwähnte Effekt im Rauschen der Daten fast vollständig unter.

Dass man so viele psychologische Erkenntnisse nicht habe bestätigen können, bedeute nicht automatisch, dass sie falsch seien, betont Nosek. Verdachtsmomente bleiben dennoch bestehen. Zu oft schon wurden strittige Resultate durch Betrug oder Täuschung gewonnen. Das trifft natür-

lich nicht auf alle jetzt wiederholten Studien zu. Bei vielen dürften Zufallsereignisse oder sonstige subtile Einflüsse zu abweichenden Resultaten geführt haben. Immerhin beschäftigt sich die psychologische Wissenschaft mit einem Gegenstand, der methodisch und experimentell nur schwer zu fassen ist: der Mensch.

In den vergangenen hundert Jahren, so geht aus einer weiteren Untersuchung hervor, wurden nur etwa ein Prozent der psychologischen Experimente wiederholt. Diese Zurückhaltung ist verständlich, denn sogenannte Replikationsstudien versprechen keinen zusätzlichen Erkenntnisgewinn und mithin keine akademische Anerkennung. »Wiederholungsforscher« gelten deshalb als eine Art Handwerker, denen nur wenig Kreativität abverlangt werde. Die zahlreichen Forschungsskandale der letzten Jahre zeigen jedoch, dass es nicht nur auf die theoretische Reproduzierbarkeit von Hypothesen, sondern ebenso auf deren praktische Überprüfung ankommt. Es wäre mithin sinnvoll, auch hierfür künftig mehr Forschungsmittel bereitzustellen. Zumal gerade psychologische Erkenntnisse ein großes mediales Echo finden und oftmals dazu beitragen, reduktionistische Menschenbilder zu formen.

Die Wiederholung als »Goldstandard« der Forschung – das gilt nicht nur für die Psychologie, sondern auch für andere Wissenschaften. Bereits 2005 hatte der Epidemiologe John Ioannidis von der Stanford University hierzu einen Aufsatz veröffentlicht – mit dem Titel: »Die meisten wissenschaftlichen Studien sind falsch.« Denn viele Forscher hätten die Hoffnung, dass man selbst eine schmale Datenbasis mit einer komplizierten Statistik noch retten könne. Das sei jedoch ein Trugschluss. Aus Studien mit einer zu geringen Teilnehmerzahl könne man keine belastbaren Aussagen ableiten. Außerdem neigten viele Wissenschaftler dazu, die von ihnen gefundenen Effekte maßlos zu überschätzen. Beispiele dafür hat Ioannidis in der Literatur reich-

lich gefunden, und man findet sie in Ratgebern bis heute: hoch dosiertes Vitamin E reduziert die Sterblichkeit bei Herz-Kreislauf-Erkrankungen, Östrogene schützen wirksam vor Alzheimer, Beta-Karotin beugt einem Sonnenbrand vor ...

In der Regel sind solche Schutzeffekte minimal oder, wie aus Wiederholungsstudien hervorgeht, gar nicht vorhanden. Dennoch werden Menschen erheblich verunsichert und veranlasst, viel Geld für wirkungslose oder gar schädliche Gesundheitsprodukte auszugeben.

Wie einfach es ist, mit etwas Fantasie und Statistik selbst die skurrilsten Hypothesen zu belegen, demonstrierte vor Jahren der kanadische Mediziner Peter C. Austin von der University of Toronto. Obwohl er dabei an die Astrologie anknüpfte, kam er ganz ohne Datenmanipulation aus. Menschen mit Sternzeichen Skorpion, so behauptete er im »Journal of Clinical Epidemiology«, hätten ein um 80 Prozent gesteigertes Risiko, an lymphoide Leukämie zu erkranken. Zu dieser auf den ersten Blick erstaunlichen, aber letztlich unsinnigen Aussage gelangte Austin dadurch, dass er die Daten von mehreren Millionen Klinikeinweisungen in der Provinz Ontario analysierte. Genauer gesagt suchte er darin nach Zusammenhängen zwischen bestimmten Sternzeichen und einer von 223 Diagnosen. Seinen Erfolg verdankte er dem Zufall. Und das nicht nur bei Skorpionen. Bei Menschen mit Sternzeichen Löwe ergab sich eine signifikante Häufung von Blutungen im Magen-Darm-Trakt, bei Fischen waren Herzkrankheiten überproportional vertreten.

»Je intensiver wir nach Mustern suchen, desto eher werden wir sie auch finden«, so Austin. Es sei daher ein Irrweg, wenn Forscher zuerst Daten sammelten und anschließend ihre Hypothesen formulierten. »In einer entsprechenden Datenmenge kann man immer statistisch belegbare Zusammenhänge aufspüren, wenn man nur genügend viele Hypothesen ausprobiert.« Mit dieser Methode und etwas Glück ließe sich zum Beispiel auch belegen, dass Männer, die täglich Bier trinken und auf Sex verzichten, weniger Haarausfall haben.

Obwohl Austin mit seiner Studie nur vor den Tücken der Statistik hatte warnen wollen, wurden seine Behauptungen von manchen Medien für bare Münze genommen. »So wirken die Sterne auf Ihre Gesundheit!« titelte etwa die »Bild«-Zeitung und listete auf, bei welchen Sternzeichen sich welche Krankheiten häuften. Natürlich müssten diese Zusammenhänge in weiteren Studien bestätigt werden, so das Blatt. Ausnahmsweise wäre eine Wiederholungsstudie hier gänzlich überflüssig. Denn Austin selbst hatte mehrmals nachdrücklich erklärt: »Weder ich noch meine Mitarbeiter glauben, dass es zwischen dem Sternzeichen eines Menschen und dessen Gesundheit einen kausalen Zusammenhang gibt.«

Wissenschaft?

Die Astrologie gilt zu Recht als unwissenschaftlich. Doch auch viele wissenschaftliche Studien liefern ebensowenig reproduzierbare Ergebnisse. Die auffälligsten Mängel dieser Art wiesen in den letzten Jahren psychologische Studien auf. Da fand mancher Forscher bei der statistischen Auswertung von Daten eine zufällige Häufung, die zu seiner Hypothese passte oder ihm eine solche erst nahelegte.

Daneben – und das gilt leider auch für die sogenannten exakten Wissenschaften – gibt es immer wieder Fälle, wo in Fachpublikationen gefälschte oder erfundene empirische Daten verwendet werden. Bekannte Beispiele sind die angeblichen menschlichen Klone des südkoreanischen Mediziners Hwang Woo Suk und die vermeintlichen Nanotransistoren des deutschen Physikers Jan Hendrik Schön. Beide Forscher kostete dieser Betrug den Job. StS