

Forschungsethische Herausforderungen in der Psychologie

Im Jahr 2011 wurden die Sozial- und Verhaltenswissenschaften durch 2 Ereignisse erschüttert: im September 2011 wurde bekannt, dass Diederik Stapel, ein niederländischer Universitätsprofessor der Sozialpsychologie im großen Stil Daten gefälscht und erfunden hatte. Über einen Zeitraum von fast 8 Jahren wurden ca. 60 Artikel mit fingierten Daten veröffentlicht. Stapel gestand, als Nachwuchswissenschaftler Zweifel an seinen Daten anmerkten und er vom Rektor der Universität damit konfrontiert worden war. Ein halbes Jahr zuvor, im Februar desselben Jahres, veröffentlichte ein Team um Daryl J. Bem, einem US-amerikanischen Sozialpsychologen, eine aufsehenerregende Studie: Studenten mussten 48 Wörter ansehen und sich danach all jene notieren, die sie sich gemerkt hatten. Danach bekamen sie eine zufällige Auswahl der 48 Wörter präsentiert und mussten sie üben. Das Bemerkenswerte: die Studenten waren scheinbar eher dazu in der Lage sich jene Wörter im ersten Teil des Experiments zu merken, die sie im zweiten Teil üben mussten – die Wirkung ging der Ursache voraus. Können wir ein wenig in die Zukunft blicken? Haben wir Vorahnungen bezüglich zukünftiger Ereignisse? Dieser von Bem als „Psi“ bezeichnete Effekt klingt verblüffend, aber er mag auch Zweifel wecken. Diesen gingen mehrere Forscherteams nach und es stellte sich heraus: solche Effekte konnte man nicht wiederfinden. Das Problem: im Gegensatz zu Bems bahnbrechender Entdeckung schafften es die Replikationsversuche nicht in das renommierte Fachjournal *Journal of Personality and Social Psychology* (JPSP). Die Psychologie hat ein Problem: sie braucht Ergebnisse, nämlich positive Ergebnisse. Und sie hat noch ein zweites: Replikation, eine Grundzutat zur Qualitätssicherung, ist nicht gerne gesehen, weil es ihr an der Neuwertigkeit fehlt. Wenn Replikationen dann zusätzlich keine positiven, bestätigenden Ergebnisse liefern, ist eine Veröffentlichung nicht zu erwarten.

Kehren wir erneut zum mittlerweile ehemaligen Universitätsprofessor Diederik Stapel zurück. Warum fälschte er seine Daten? Ein Grund mag darin liegen, dass unter Forschern ein Wettkampf um Veröffentlichungen besteht. Wissenschaftler sind dann erfolgreich, wenn sie viel publizieren. Aber Fachzeitschriften berichten lieber über neue, bahnbrechende Entdeckungen und halten sich nicht mit Reproduktionsversuchen alter Studien auf. Forscher müssen also neue, unbekannte Themenbereiche erschließen und Ergebnisse liefern. Das Unbekannte aber hat einen Haken: es mag nicht immer ein Topf mit Gold am Ende des Regenbogens stehen, manchmal sind Thesen einfach zu verwerfen und Studien ergebnislos. Diese Misserfolge – und genau hier liegt der Fehler – schaffen es meist nicht in Fachzeitschriften. Dabei sind solche Studien weder ergebnislos noch Misserfolge, stattdessen zeigen sie uns auf, dass eine gewisse Fragestellung mit den erhobenen Daten verneint werden kann. Doch Forscher brauchen Forschungsgelder, sie stehen unter Publikationsdruck. Da mag es schon helfen, Ergebnisse zu beschönigen, Daten zu verändern, um plötzlich doch den erwarteten positiven Effekt wiederzufinden. Solche Effekte schaffen es dann auch häufiger durch die Peer-Review, einem

forschungsinternen Qualitätssicherungsprozess, in welchem Experten des jeweiligen Themengebiets die zu veröffentlichende Studie analysieren und kritisch hinterfragen. Warum haben es positive Effekte leichter? Dafür liefert die Sozialpsychologie ironischerweise die Antwort: der *confirmation bias*. Konsistente Geschichten lesen sich besser. Wenn die Ergebnisse nicht die Annahmen bestätigen, werden Gesichter in Falten gelegt. Aber wenn die Ergebnisse zu den Vermutungen passen, dann scheint ja alles klar zu sein.

Die Folge? Wissenschaftler inkludieren fragwürdige Werkzeuge in ihre Methoden: man spricht von *Questionable Research Practices*. So werden beispielsweise Versuchsabläufe genau so ausgerichtet, dass ein Positivergebnis möglichst leicht eintreten kann. Wenn dieses Ergebnis dann veröffentlicht ist, ist das Thema oftmals erledigt. Replikationen im klassischen Sinne schaffen es, wie bereits erwähnt, nicht in die Zeitschriften. Daher werden meist nur ähnliche Hypothesen mit ähnlichen Methoden geprüft, eine direkte Infragestellung des Ergebnisses passiert allerdings meist nicht.

Ein weiteres fragwürdiges Werkzeug zur Ergebnissicherung ist das sogenannte *p-Hacking*. Der p-Wert bezeichnet die statistische Wahrscheinlichkeit dafür, dass ein gefundenes Ergebnis auch ohne den vermuteten Effekt durch Zufall zustande kommt. Berechnet wird er auf Basis einer Annahme: es besteht zwischen zwei Variablen (z.B. Kaffeekonsum und Leistungsfähigkeit) keine Abhängigkeit. Hierbei spricht man von einer Nullhypothese. Getestet wird, mit welcher Wahrscheinlichkeit ein Ergebnis genauso extrem ausfallen würde wie das beobachtete Ergebnis, wenn zwischen den Variablen wirklich keine Abhängigkeit gegeben ist. Sinkt die Wahrscheinlichkeit auf unter 5% wird üblicherweise davon ausgegangen, dass man doch von einer Abhängigkeit ausgehen muss und die beiden Variablen in Wirklichkeit zusammenhängend sind. Dann spricht man von einem signifikanten Ergebnis. Über die Stärke des Zusammenhangs im eigentlichen Sinne wird aber keine Aussage getroffen. Besonders, wenn die Analysen geringe, aber nicht signifikante Wahrscheinlichkeiten hergeben, bietet sich für den Wissenschaftler die Möglichkeit, durch geringe Veränderung der Daten (z.B. bestimmte Werte als Ausreißer deklarieren und ausschließen) doch noch die statistische Signifikanz zu erreichen – damit kommt man der Publikation schon näher. Alternativ werden die ursprünglichen Fragestellungen auch im Nachhinein geändert und entsprechend der signifikanten Ergebnisse angepasst. Bin ich ursprünglich davon ausgegangen, dass ein Zusammenhang zwischen Kaffeekonsum und Leistungsfähigkeit besteht? Meine Ergebnisse passen nicht dazu? Nun nehme ich an, dass kein Zusammenhang besteht. Wer viel Kaffee trinkt, lernt nicht besser! *Glück gehabt, jetzt wurden meine Annahmen doch noch bestätigt...*

Ein Beispiel für die Vielfältigkeit statistischer Analysetechniken bietet eine Frage, die 29 Forscherteams gestellt wurde: zeigen Schiedsrichter dunkelhäutigen Fußballspielern häufiger eine rote Karte? Das Ergebnis: 29 Ergebnisse. Während einzelne Analysen eine bis zu dreifach erhöhte

Wahrscheinlichkeit für rote Karten bei dunkelhäutigen Spielern fanden, fand der Großteil der Forscherteams eine Wahrscheinlichkeit, die circa um das 1,3-fache erhöht war. Andere Teams sahen sogar einen umgekehrten Effekt. Klar ist: Einzelergebnisse sind mit Vorsicht zu interpretieren, die Wissenschaft lebt von der Reproduktion ihrer Experimente. Aussagekraft entsteht nicht durch eine bahnbrechende Entdeckung, sie entsteht durch die mehrfache Bestätigung der Entdeckung im Zuge kritischer Hinterfragung der Ergebnisse.

Kann man die psychologische Wissenschaft noch retten? Einen Rettungsring wirft sicherlich die *Open-Science-Bewegung*, die es sich zum Ziel gesetzt hat, die Qualität der Forschung durch erhöhte Transparenz zu sichern. Dies erfolgt beispielsweise dadurch, dass Wissenschaftler dazu ermutigt werden, die inhaltlichen und methodischen Abläufe vor (!) der eigentlichen Datenerhebung festzulegen und zu veröffentlichen. Dieser Prozess wird als Präregistrierung bezeichnet. So können nachträgliche Schönungen verhindert und der bisher unzugängliche Umgang mit den Daten aus der „Black Box“ befreit werden. Klar ist, dass dies nur in Kooperation mit Fachzeitschriften zu mehr Transparenz und einem Image-Rückgewinn führen kann. Diese müssen Forschern Anreize bieten, sodass diese dazu ermutigt werden, ihre Forschungsvorhaben vorab zu registrieren. Immerhin schränkt dies die Flexibilität im nachträglichen Umgang mit den Daten und Hypothesen ein, man begibt sich als Forscher in Zeiten hohen Publikationsdrucks damit auf dünnes Eis. Glücklicherweise bieten mittlerweile 158 Journals zumindest teilweise die Möglichkeit zur Publikation präregistrierter Artikel. Dies bedeutet nicht, dass in diesen Journalen ausschließlich solche Artikel publiziert werden und garantiert ebenso wenig, dass nicht bestätigte Forschungshypothesen gleichwertige Behandlung finden. Aber es ist ein Anfang und signalisiert, dass zumindest innerhalb der psychologischen Wissenschaftsdisziplin mehr und mehr Bewusstsein über den Mangel an transparentem Vorgehen herrscht.