

Nicht reproduzierbare Wissenschaft – Teil 2

geschrieben von Chris Frey | 23. Mai 2022

Dr. Jay Lehr

Der empirische Wissenschaftler führt kontrollierte Experimente durch und führt genaue, unvoreingenommene Aufzeichnungen über alle beobachtbaren Bedingungen zum Zeitpunkt der Durchführung des Experiments. Wenn ein Forscher ein wirklich neues oder bisher unbeobachtetes Naturphänomen entdeckt hat, sollten andere Forscher – mit Zugang zu seinen Aufzeichnungen und einigen selbst entwickelten Geräten – in der Lage sein, diese Entdeckung zu reproduzieren oder zu bestätigen. Wenn genügend Bestätigungen vorliegen, erkennt die wissenschaftliche Gemeinschaft schließlich an, dass das Phänomen real ist, und passt die bestehende Theorie an die neuen Beobachtungen an.

Die Validierung der wissenschaftlichen Wahrheit erfordert eine Replikation oder Reproduktion. Replizierbarkeit bedeutet in der Regel, dass das Ergebnis eines Experiments in einer unabhängigen Studie von verschiedenen Forschern mit unterschiedlichen Daten erzielt wird, während Reproduzierbarkeit bedeutet, dass verschiedene Forscher die gleichen Daten, Methoden und/oder Computercodes verwenden, um zu den gleichen Ergebnissen zu gelangen.

Doch heute funktioniert der wissenschaftliche Prozess der Replikation und Reproduktion nicht mehr richtig. Ein großer Teil der wissenschaftlichen Behauptungen in der veröffentlichten Literatur ist nicht repliziert oder reproduziert worden. Man schätzt, dass die Mehrheit dieser veröffentlichten Behauptungen, die nicht repliziert oder reproduziert werden können, in Wirklichkeit falsch sind.

Eine außerordentliche Anzahl wissenschaftlicher und sozialwissenschaftlicher Disziplinen liefert nicht mehr zuverlässig wahre Ergebnisse, ein Zustand, der gemeinhin als Irreproduzierbarkeits-Krise bezeichnet wird. Eine beträchtliche Mehrheit von 1500 aktiven Wissenschaftlern, die kürzlich von der Zeitschrift *Nature* befragt wurden, bezeichnete die akute Situation als „Krise“. Die völlig unangemessenen Leistungsanreize in der wissenschaftlichen Welt tragen einen Großteil der Schuld an diesem katastrophalen Versagen.

Politiker und Bürokraten handeln in der Regel, um ihre eigenen Interessen zu maximieren, anstatt als uneigennützig Diener des Gemeinwohls zu agieren. Dies gilt insbesondere für Wissenschaftler, Peer-Review-Gutachter und Regierungsexperten. Die verschiedenen Teilnehmer am wissenschaftlichen Forschungssystem dienen alle ihren eigenen Interessen, da sie den Anreizen des Systems erliegen.

Gut publizierende Universitätsforscher verdienen sich durch die Veröffentlichung aufregender neuer positiver Ergebnisse eine Festanstellung, eine Beförderung, einen Wechsel an eine renommierte Universität, Gehaltserhöhungen, Stipendien, berufliches Ansehen und öffentliche Wertschätzung. Die gleichen Anreize wirken sich auf die Herausgeber von Zeitschriften aus, die durch die Veröffentlichung von als aufregend empfundenen neuen Forschungsergebnissen Anerkennung für ihre Zeitschrift und persönliche Auszeichnungen erhalten – auch wenn die Forschungsergebnisse nicht gründlich überprüft wurden.

Geldgeber wollen spannende Forschung finanzieren, und staatliche Geldgeber haben den zusätzlichen Anreiz, dass spannende Forschung mit positiven Ergebnissen die Erweiterung des Auftrags ihrer Organisation unterstützt. Amerikanische Universitätsverwaltungen wollen Forschungsprojekte fördern, von denen sie profitieren, indem sie die Gemeinkosten – häufig den größten Teil des Förderbetrags – erhalten. Als jemand, der dies aus erster Hand erfahren und gesehen hat, wird es den Leser verblüffen, welcher großer Teil der meisten Forschungszuschüsse als Gemeinkosten an die Universität geht, anstatt die eigentlichen Forschungskosten zu unterstützen.

All diese Anreize belohnen veröffentlichte Forschung mit neuen positiven Behauptungen, aber nicht unbedingt reproduzierbare Forschung. Forscher, Herausgeber, Geldgeber, Bürokraten, Universitätsverwaltungen – sie alle haben einen Anreiz, nach scheinbar aufregenden neuen Forschungsergebnissen zu suchen, die Geld, Status und Macht bringen. **Es gibt nur wenige oder gar keine Anreize, ihre Arbeit zu überprüfen.** Vor allem haben sie wenig Anreiz, die Forschung zu reproduzieren und zu überprüfen, ob die aufregende Behauptung Bestand hat, denn wenn dies nicht der Fall ist, verlieren sie Geld, Status und Prestige.

[Hervorhebung vom Übersetzer]

Die Anreize der wissenschaftlichen Welt für neue Erkenntnisse und nicht für reproduzierbare Studien haben einen drastischen Einfluss darauf, was zur Veröffentlichung eingereicht wird. Wissenschaftler, die versuchen, ihre Karriere auf der Überprüfung alter Erkenntnisse oder der Veröffentlichung negativer Ergebnisse aufzubauen, werden wahrscheinlich keinen beruflichen Erfolg haben. Das Ergebnis ist, dass Wissenschaftler keine negativen Ergebnisse zur Veröffentlichung einreichen. Einige negative Ergebnisse landen in der Aktenschublade. Andere verwandeln sich irgendwie in positive Ergebnisse, weil die Forscher bewusst oder unbewusst ihre Daten und ihre Analysen frisieren (als wissenschaftliche Modellierer nennen wir das „Tuning“, ein Fachwort für Betrug). Sie führen auch nicht viele Replikationsstudien durch oder veröffentlichen sie, da die Anreize der wissenschaftlichen Welt auch diese Aktivitäten nicht belohnen.

Das Konzept der statistischen Signifikanz wird so sehr strapaziert, dass buchstäblich Hunderte, wenn nicht Tausende von nutzlosen Arbeiten, die

diese Signifikanz für sich beanspruchen, überall erscheinen.

Forscher versuchen festzustellen, ob sich die von ihnen untersuchten Zusammenhänge von dem unterscheiden, was allein durch Zufall erklärt werden kann, indem sie Daten sammeln und Hypothesentests durchführen, die auch als Tests der statistischen Signifikanz bezeichnet werden. Meistens beginnen sie damit, die Wahrscheinlichkeit zu testen, dass es keinen tatsächlichen Zusammenhang zwischen zwei Variablen gibt, was als „Nullhypothese“ bezeichnet wird. Wenn diese Hypothese nicht zutrifft und es wahrscheinlich ist, dass ein Zusammenhang besteht, wird eine andere Hypothese aufgestellt. Wie gut die Daten die „Nullhypothese“ (kein Zusammenhang) unterstützen, zeigt eine Statistik, die als p-Wert bezeichnet wird. Liegt der p-Wert unter 5 % oder 0,05, wird davon ausgegangen, dass ein Zusammenhang zwischen den untersuchten Variablen bestehen könnte.

Die zentrale Rolle der Regierungen in der Wissenschaft, sowohl bei der Finanzierung wissenschaftlicher Forschung als auch bei der Verwendung wissenschaftlicher Forschung zur Rechtfertigung von Vorschriften, trägt in hohem Maße dazu bei, dass die fadenscheinige statistische Signifikanz in der gesamten akademischen Welt zunimmt. Innerhalb einer Generation hat sich die statistische Signifikanz von einer nützlichen Abkürzung, die Forscher in der Landwirtschaft und der Industrie benutzten, um zu entscheiden, ob sie ihre derzeitigen Verfahren beibehalten oder zu etwas Neuem übergehen sollten, zu einer Voraussetzung für Regulierung, staatliche Zuschüsse, Festanstellungen und jede andere Form von wissenschaftlichem Prestige entwickelt und ist auch für Veröffentlichungen unerlässlich.

Viele weitere Wissenschaftler verwenden eine Vielzahl von statistischen Verfahren mit mehr oder weniger sträflicher Nachlässigkeit, darunter:

- *unrichtige statistische Methodik

- *voreingenommene Datenmanipulation, die zu den gewünschten Ergebnissen führt

- *Auswahl von Messwerten, die statistisch signifikant sind, und Ignorieren von Messwerten, die nicht signifikant sind

- *unzulässige Manipulationen von Forschungsverfahren

Wieder andere führen statistische Analysen durch, bis sie ein statistisch signifikantes Ergebnis finden, und veröffentlichen dieses Ergebnis. Dies wird als „p-hacking“ bezeichnet. Viel zu viele Forscher geben ihre Methoden unklar an und lassen den uninformatierten Leser annehmen, dass sie tatsächlich einem strengen wissenschaftlichen Prozess gefolgt sind.

Der heimtückischste aller wissenschaftlichen Betrügereien ist p-HARKING. Das bedeutet, dass ein Wissenschaftler eine Hypothese erst dann

aufstellt, wenn er alle Daten gesammelt hat, die zu dem gewünschten Ergebnis führen. Ein offensichtlicheres Wort dafür ist CHEATING [to cheat = schummeln]. Nicht reproduzierbare Forschungshypothesen, die durch HARKING zustande kommen, treiben ganze Disziplinen in den Abgrund.

Publikationsverzerrungen und „p-harking“ haben die wissenschaftliche Forschung insgesamt in Mitleidenschaft gezogen. Darüber hinaus zeigen Umfragen seit Jahrzehnten, dass Forscher wahrscheinlich keine negativen Ergebnisse ihrer Studien veröffentlichen werden.

Eine falsche Forschungsbehauptung kann zur Grundlage für eine ganze Reihe von Veröffentlichungen werden, die durchweg falsch sind und dennoch zur anerkannten Wahrheit werden. Wir können nicht genau sagen, welche Forschungsarbeiten von diesen Fehlern betroffen sind, solange Wissenschaftler nicht alle veröffentlichten Forschungsarbeiten replizieren. Wir verfügen jedoch über ausgefeilte statistische Strategien, die es uns ermöglichen, bestimmte Behauptungen zu diagnostizieren, die eine staatliche Regulierung unterstützen. Eine solche Methode – ein Härtetest für statistische Betrügereien – ist die Darstellung von p-Werten, die im Handbuch der National Association of Scholars, SHIFTING SANDS, ausführlich beschrieben wird. Ein kurzes Papier zurück, das ich nicht zu sehr empfehlen kann.

Hinweis: Teile dieses Aufsatzes wurden mit Genehmigung der National Association of Scholars und ihrer Autoren Peter Wood, Stanley Young, Warren Kindzierski und David Randall aus dem NAS-Buch SHIFTING Sands entnommen.

Autor: CFACT Senior Science Analyst [Dr. Jay Lehr](#) has authored more than 1,000 magazine and journal articles and 36 books. Jay's new book *A Hitchhikers Journey Through Climate Change* written with Teri Ciccone is now available on Kindle and Amazon.

Link: <https://www.cfact.org/2022/05/16/irreproducible-science-part-two/>

Übersetzt von [Christian Freuer](#) für das EIKE