

# Die unvermeidliche Evolution schlechter Wissenschaft

geschrieben von Ed Yong, The Atlantic | 16. Oktober 2016

Wie auch Andere hier schon angemerkt haben, liegt das Problem darin, dass das, was für individuelle Wissenschaftler gut ist, nicht notwendigerweise auch für die Wissenschaft als Ganzes gut ist. Gegenwärtig ist die Karriere eines Wissenschaftlers abhängig von der Veröffentlichung so vieler Studien wie möglich in den prestigeträchtigsten Journalen. Mehr als alles andere verschafft ihnen dies Prestige, [finanzielle] Zuwendungen und Jobs.

Jetzt stelle man sich vor, man sei ein Forscher, der mit diesem System spielen will. Folgendes wird man tun: man fertige viele kleine und statistisch auf schwachen Füßen stehende Studien an. Man optimiere dann seine Verfahren, um positive Ergebnisse sicherzustellen. Falls man negative Ergebnisse erhält, kehre man sie unter den Teppich. Man versuche niemals, frühere Ergebnisse zu checken; man verfolge nur neue und begeisternde Ergebnisse. Dies ist kein Larifari. Wir wissen, dass es solche Praktiken im Überfluss gibt. Sie sind großartig, um Veröffentlichungen zu lancieren, aber sie verunreinigen auch die wissenschaftliche Aufzeichnung mit Ergebnissen, die *tatsächlich nicht wahr sind*. Richard Horton, Herausgeber von *The Lancet*, schrieb einmal: „Niemand wird Anreize bekommen, um Recht zu haben. Stattdessen bekommen Wissenschaftler Anreize, *produktiv* zu sein“.

Dies ist kein neuer Gedanke. In den siebziger Jahren schrieb der Sozialwissenschaftler Donald Campbell, dass jedwede Qualitäts-Metrik korrumpiert werden kann, falls die Leute anfangen, die Metrik selbst zu priorisieren über die Wesenszüge, die sie vermeintlich reflektiert. „Wir fanden, dass dieses Argument selbst dann funktioniert, falls Individuen nicht *versuchen*, ihre Metrik zu maximieren“, sagt Smaldino.

Er und McElreath zeigten dies hier, indem sie ein mathematisches Modell erzeugten, in dem simulierte Labore miteinander im Wettbewerb standen und sich entwickelten – man denke an SimAcademia. Die Labore wählten Dinge, die sie studieren wollten, ließen Experimente laufen, um ihre Hypothesen zu testen und versuchten, ihre Ergebnisse zu veröffentlichen. Sie variieren hinsichtlich der Stärke der Bemühungen zum Austesten ihrer Gedanken, was darauf einwirkt, wie viele Ergebnisse sie bekommen und wie zuverlässig diese Ergebnisse sind. Da gibt es einen Zielkonflikt: Mehr Mühe bedeutet richtigere, aber weniger Veröffentlichungen.

Im Modell wie in der Akademia sind positive Ergebnisse leichter zu veröffentlichen als negative, und Labore, die mehr veröffentlichen, bekommen mehr Prestige, Förderung und Studenten. Auch sie geben ihre Verfahren weiter. Mit jeder Generation stirbt eines der ältesten Labore,

während sich eines der produktivsten reproduziert und Nachwuchs hervorbringt, der den Forschungsstil seines Doktorvaters nachmacht. Das ist das Äquivalent eines Studenten aus einem erfolgreichen Team, der ein eigenes Labor eröffnet.

Mit der Zeit und über viele Simulationen gleiten die virtuellen Labore unaufhaltsam in Richtung weniger Mühe, schlechtere Verfahren und fast vollständig unzuverlässiger Ergebnisse. Das Wichtige dabei ist: Anders als der hypothetische Forscher, den ich zuvor heraufbeschworen habe, versucht keiner dieser simulierten Wissenschaftler aktiv zu schummeln. Sie wandten keine Strategie an und verhielten sich integer. Und doch glitt die *Gemeinschaft* in Richtung schlechterer Verfahren. Das Modell zeigt: eine Welt, die Wissenschaftler für Publikationen über allem anderen belohnt – also eine dieser Welt nicht ganz unähnliche – führt auf natürliche Weise zu schwächerer Wissenschaft.

„Das Modell kann sogar optimistisch sein“, sagt Brian Nosek vom Center of Open Science, weil es nicht unsere unglückliche Tendenz in Betracht zieht, den Status Quo zu rechtfertigen und zu verteidigen. Er merkt zum Beispiel an, dass Studien im Bereich der sozialen und biologischen Wissenschaften im Mittel beklagenswert schwach ausgestattet sind – sie sind zu klein, um zuverlässige Ergebnisse zu erbringen.

Geringe statistische Belastbarkeit ist ein offensichtliches Symptom schwacher Forschung. Sie ist leicht zu berechnen, und die Menschen haben darüber seit den sechziger Jahren geredet. Und doch, über 50 Jahre später, hat sich daran *überhaupt nichts* verbessert. Tatsächlich „gibt es nach wie vor *aktiven* Widerstand gegen Bemühungen, die statistische Belastbarkeit zu erhöhen, und zwar seitens der Wissenschaftler selbst“, sagt Nosek. „Wenn der Wunsch nach Veröffentlichung höher ist als der Wunsch nach Richtigkeit, werden Forscher geringe statistische Belastbarkeit verteidigen, obwohl dies Null Qualität für die Wissenschaft bedeutet“.

**Wichtig: „Solange es die Anreize gibt, wird es Belohnungen für jene geben, die das System täuschen können, ob sie das nun absichtlich machen oder nicht“.**

Wissenschaftler haben jetzt mit den Konsequenzen dieser Stagnation zu kämpfen. In vielen Bereichen, darunter Neurowissenschaft, Genetik, Psychologie, Ökologie und Biomedizin, gibt es Gerede über eine Reproduzierbarkeits-Krise, in der schwache und schlechte Studien die Welt mit zweifelhaften Ergebnissen überschwemmt haben. „Wir bringen viel Zeit damit zu, uns über die Kultur der Wissenschaft zu beklagen, aber verbale Diskussionen erlauben es den Leuten, darüber zu reden“, sagt Smaldino. „Ein formelles Modell erlaubt es deutlicher zumachen, worüber man spricht“.

Zum Beispiel haben sich viele Wissenschaftler auf Replikation konzentriert – also der Wiederholung von Studien, um zu sehen, ob deren

Ergebnisse stehen – als eine Art, die Zuverlässigkeit der Wissenschaft zu verbessern. Aber das würde dem Modell von Smaldino und McElreath zufolge nichts besser machen. Ihre Labore *könnten* Zeit damit verbringen, Arbeiten der Vergangenheit zu reproduzieren, und falls diese Versuche scheitern, würden die ursprünglichen Autoren einen erheblichen Rufschaden erleiden. Aber das spielte keine Rolle, „weil es viel mehr Ergebnisse gibt als überhaupt reproduziert werden können“, sagt Smaldino. Langfristig sind Labore damit durchgekommen, wenn sie schlechte Verfahren angewendet hatten, sogar falls andere sie mit ihren dubiosen Ergebnissen konfrontierten.

„Solange es die Anreize gibt, wird es Belohnungen für all jene geben, die das System beschummeln können, ob sie das nun absichtlich tun oder nicht“, sagt Smaldino. Um die Wissenschaft zu verbessern, müssen sich die Anreize ändern.

Und diese Änderungen müssen für alle Bereiche gelten, aber es müssen keine großen Veränderungen sein, sagt Nosek. Beispiel: Wenn Wissenschaftler nach Promotionen streben, werden sie oftmals aufgefordert, eine vollständige Liste ihrer Studien vorzuweisen. Niemand hat die Zeit, alle diese Studien zu *lesen*, und darum greifen die Mitglieder des Komitees zu an sich unpassenden Dingen wie etwa der Anzahl der Studien oder dem Prestige der Journale. „Eine einfache Änderung wäre es, den Kandidaten aufzufordern drei Artikel einzusenden, welche das Komitee lesen und detailliert bewerten kann“, sagt Nosek. „Das wären Anreize für die Kandidaten, drei herausragende Arbeiten zu erstellen“.

Aber in UK hat man ein solches System bereits aufgelegt, um Wissenschaftler zu beurteilen, und Andrew Higginson sowie Marcus Munafò, zwei Psychologen der Universities of Exeter bzw. Bristol würden verneinen, dass dies besser ist. Sie haben ein anderes mathematisches Modell angewendet, um vorherzusagen, wie sich Wissenschaftler verhalten sollen, um den Wert ihrer Veröffentlichungen für ihre Karriere zu maximieren. Und sie kamen zu dem Ergebnis, dass falls die Leute auf der Grundlage einer kleinen Anzahl von hoch einflussreichen Veröffentlichungen beurteilt werden, es für sie die beste Strategie ist, all ihre Bemühungen auf geringerwertige Studien zu konzentrieren, die sich nur an neuesten Ergebnissen ausrichten ohne die Ergebnisse älterer Studien zu checken. Als Folge davon wird die Hälfte dessen, was sie veröffentlichen, falsch sein.

Es gibt andere Lösungen. Einige Wissenschaftler haben sich für ein System der „Vor-Registrierung“ [pre-registration] stark gemacht, wobei die Arbeit vor dem Hintergrund ihrer Gedanken und Pläne bewertet wird, *bevor* irgendeine Arbeit tatsächlich angefertigt wird. Sie verpflichten sich, den Plan buchstabengetreu auszuführen, und die Journale verpflichten sich, die Ergebnisse zu veröffentlichen, egal welcher Art diese sind. Das reduziert die Kapazität und den Anreiz, mit Studien Schindluder zu treiben, um die eigenen Chancen auf die Durchführung

einer Studie zu erhöhen. Außerdem verschiebt sich der Schwerpunkt weg von augenfälligen Ergebnissen hin zu soliden, zuverlässigen Verfahren. Fast 40 Journale veröffentlichen diese Art von *Registered Reports*, und es gibt Bemühungen, diese enger an Zuwendungen zu binden (hier), so dass eine einzelne Begutachtung der Verfahren in einer Studie Zuwendungen und Veröffentlichung garantiert.

Das Hauptgewicht auf Transparenz zu legen, kann ebenfalls helfen, sagt Simine Vazire, eine Psychologin an der University of California. „Falls die Autoren verpflichtet werden, mehr Details über ihre Forschungen bekannt zu geben, werden die Journale und die Begutachter besser in der Lage sein, die Qualität dieser Studien zu evaluieren, und es wird viel schwieriger für die Autoren sein, das System vorzuführen“.

Spitzenjournale wie *Nature* und *Science* ermutigen Autoren tatsächlich, transparenter zu sein hinsichtlich Daten und Verfahren, während sie Checklisten ausgeben, um es einfacher für die Herausgeber zu machen, die statistischen Qualitäten neuer Studien zu inspizieren. Und das Center for Open Science von Nosek hat Standards für Transparenz, Offenheit und Reproduzierbarkeit erstellt (hier), welche Journale und Geldgeber übernehmen können, sowie Auszeichnungen für gutes Verhalten (hier).

Ultimativ „ist die Änderung von Anreizen über das gesamte Wissenschafts-Ökosystem ein Problem der Koordination“, sagte Nosek. „Institutionen, Geldgeber, Herausgeber, Gesellschaften und die Forscher selbst müssen ihre Erwartungen allesamt etwas ändern. Ohne dies wird keine Änderung effektiv sein“.

Munafo ist voller Hoffnung. „Wir haben Fortschritte gemacht, von der Beschreibung des Problems hin zum Verständnis von dessen Natur“, sagt er. „Das ist ein gutes Zeichen. Hoffen wir, dass es uns Hinweise gibt, wo wir Anreiz-Strukturen am effizientesten ändern können. Wir stecken mitten in einem faszinierenden natürlichen Experiment, wobei viele Innovationen eingeführt oder angestoßen werden. Was funktioniert und was nicht, und was populär ist und was nicht, bleibt abzuwarten“.

„Ich möchte nicht übermäßig pessimistisch sein“, sagt Smaldino. „Es gibt viele wirklich gute Wissenschaftler, die nach Arbeiten mit hoher Qualität streben. Es gibt immer mehr Individuen, die erkennen, dass Qualität tatsächlich bedeutsam ist. Ich hoffe einfach, dass Sentiment vorherrscht“.

Link:

<http://www.theatlantic.com/science/archive/2016/09/the-inevitable-evolution-of-bad-science/500609/>

Übersetzt von Chris Frey EIKE