



Einsichten eines Wissenschaftsnarren (53)

Wissenschaftsbetrug ist selten. Stimmt das eigentlich?

Der Narr hat recherchiert: Wissenschaftliches Fehlverhalten – also vor allem Plagiarismus, Fälschungen und Fabrikation von Daten – ist viel häufiger als wir uns eingestehen. Gegenmaßnahmen sind also dringend angesagt.

Fast wöchentlich lesen wir von Fällen wissenschaftlichen Fehlverhaltens. Häufig spielen darin renommierte Journale und prominente Wissenschaftler eine Rolle. Unablässig versorgt uns die Website *Retraction Watch* von Ivan Oransky und Adam Marcus mit entsprechenden „Enthüllungsnachrichten“ und deren Hintergründen. Und auch im *Laborjournal* findet sich fast in jeder Ausgabe eine Story über ein Labor, in dem es nicht mit rechten Dingen zugeht.

»Selten decken Universitäten, Fördergeber oder Journale solche malignen Machenschaften auf.«

Meist wurden solche Machenschaften ruckbar, nachdem ein Artikel mit manipulierten, gefälschten oder gar erfundenen Daten aufgefliegen war. Ans Licht der wissenschaftlichen Öffentlichkeit bringen dies oft Whistleblower. Oder aufmerksame Leser, die ihre Zweifel an der Verlässlichkeit von Abbildungen anonym auf PubPeer veröffentlichen.

Auffällig selten dagegen decken Universitäten, Fördergeber oder Journale solche malignen Machenschaften auf.

Recht häufig greifen dann auch die Wissenschaftsseiten der Tageszeitungen derartige Nachrichten aus den moralischen Niederungen der Wissenschaft auf. Aktuell versorgen sie uns etwa mit Berichten über fragwürdige Publikationen aus den „Ställen“ des Nobelpreisträgers Gregg Semenza, dem „Entdecker“ des Hypoxie-induzierten Faktors (HIF), oder des Neurowissenschaftlers und derzeitigen

Präsidenten der Stanford University Marc Tessier-Lavigne. Dabei geht es nicht nur um Arbeiten aus der Grundlagenforschung: Gerade COVID-19 hat uns einen wahren Tsunami von Artikel-Retraktionen beschert – angeführt etwa von Studien aus *The Lancet* und dem *New England Journal of Medicine*, die auf komplett erfundenen Daten beruhten.

Wir selbst goutieren Berichte über wissenschaftliches Fehlverhalten oft mit wohligem Gruseln, doch groß aufregen tun wir uns darüber nicht. Wir haken es vielmehr unter der Rubrik „Jede Branche hat ihre schwarzen Schafe“ ab – ist das doch menschlich, allzu menschlich. Oder wir tun es – wenn es beispielsweise um Massen-Retraktionen von Artikeln aus sogenannten „Paper Mills“ geht, die gegen Gebühr komplett erfundene Artikel verfassen – als exotische Phänomene des Wissenschaftsbetriebes in uns fernen Weltgegenden ab. Bei uns spielt Wissenschaftsbetrug – also das Plagiiere sowie Fälschungen oder Fabrizieren von Daten – doch keine wichtige Rolle.

Wenn überhaupt, dann vielleicht Plagiarismus – allerdings eher in nicht-naturwissenschaftlichen Fächern. Solches Abschreiben ist zwar unschön, aber doch ein eher geringfügiges Vergehen. In ihren Pressemitteilungen listet die DFG für das Jahr 2022 ganze sechs Fälle, in denen Wissenschaftler wegen wissenschaftlichem Fehlverhalten gerügt und für ein paar Jahre von der Förderung ausgeschlossen wurden. Zeigt das nicht, wie selten Betrügeereien in unserem Wissenschaftsbetrieb sind?

Auch der Narr hat sich bis vor kurzem dieser bequemen Illusion hingegeben. Und geglaubt, dass es fast exklusiv sogenannte „fragwürdige wissenschaftliche Praktiken“ sind, die unsere Aufmerksamkeit verdienen. Also das Weglassen von Befunden, die eine Story nicht mehr ganz so glatt erscheinen lassen. Oder die Durchführung multipler Tests, bis man auf einen stößt, der die ersehnte statistische Signifikanz ergibt – auch als p-Hacking bekannt und beliebt. Oder das Formulieren von Hypothesen, nachdem man die Ergebnisse bereits kennt – aber so tut, als wären die Versuche durchgeführt worden, um genau diese Hy-

pothesen zu testen (sogenanntes „HARKing – Hypothesizing after the results are known“).

Oder das Bearbeiten von Banden auf Western-Blots mit Photoshop? Oder das Manipulieren von Ergebnissen in Spreadsheets? Die Verwendung von Kontrollergebnissen, die gar nicht zum aktuellen Experiment gehören? Nicht bei uns, und auch nicht im Nachbarlabor!

Doch warum sind wir uns da eigentlich so sicher? Denn eigentlich spricht sehr viel dafür, dass wissenschaftliches Fehlverhalten weit häufiger ist, als wir uns das eingestehen. Eine aufwendige und methodisch exzellente Arbeit ergab gerade erst, dass acht Prozent von etwa 7.000 niederländischen Wissenschaftlern, die auf eine anonyme Umfrage zu Forschungspraktiken geantwortet hatten, zwischen 2017



Foto: BIH/Thomas Rafalzyk

Ulrich Dirnagl

ist experimenteller Neurologe an der Berliner Charité und ist Gründungsdirektor des QUEST Center for Transforming Biomedical Research am Berlin Institute of Health. Für seine Kolumne schlüpft er in die Rolle eines „Wissenschaftsnarren“ – um mit Lust und Laune dem Forschungsbetrieb so manche Nase zu drehen.

Sämtliche Folgen der „Einsichten eines Wissenschaftsnarren“ gibt es unter www.laborjournal.de/rubric/narr

und 2020 mindestens einmal Daten gefälscht und/oder erfunden hatten! In Medizin und Biowissenschaften waren es sogar über zehn Prozent. Mehr als die Hälfte gab außerdem zu, häufig (!) fragwürdige Wissenschaftspraktiken anzuwenden.

Haben niederländische Wissenschaftler etwa mehr kriminelle Energie als deutsche? Vermutlich schon alleine deshalb nicht, weil die Niederlande mit dem Fall „Diederik Stapel“ einen wissenschaftlichen Betrugsskandal erlebten, der die gesamte Nation bis ins Mark erschütterte – und der überdies weitreichende Konsequenzen im niederländischen Wissenschaftssystem zur Folge hatte. Zum Beispiel einen nationalen Plan unter Beteiligung der Universitäten und Fördergeber mit dem Ziel, offene Wissenschaft zu fördern („Open Science“). Oder den Start einer Reform des akademischen Karriere- und Belohnungssystems („Every talent counts“). Um beides beneiden wir unsere Nachbarn mittlerweile.

So schockierend die Ergebnisse der niederländischen Umfrage sind, so sehr passen sie doch ins Bild. Beispielsweise können mittlerweile wissenschaftliche Abbildungen automatisiert auf Manipulationen untersucht werden. Und tatsächlich zeigt die Anwendung dieser Techniken, dass mehr als vier Prozent aller biomedizinischen Publikationen Graphen und Abbildungen enthalten, die hochgradig verdächtig auf maligne Manipulationen sind – etwa die Verschiebung von Banden, Duplikationen, nicht plausible Fehlerbalken und so weiter. Diese Zahlen werden auch durch Arbeiten bestätigt, in denen Menschen die Abbildungen untersuchten.

Gleichzeitig jedoch hat ein Wettlauf begonnen zwischen Software, die kaum noch zu erkennende „Deep Fakes“ von wissenschaftlichen Grafiken erzeugen kann, sowie Software, die in der Lage ist, genau diese zu erkennen. Die steigende Zahl von Retraktionen sowie die vermehrten Berichte über nachgewiesenen Wissenschaftsbetrug können uns ja immer nur die Spitze des Eisbergs von tatsächlich realisiertem Fehlverhalten anzeigen – vermutlich mit einem Bias in Richtung der krasseren Verstöße.

»Was Stapel letztlich das Genick brach, waren seine eigenen Studenten.«

Hieraus auf die wirkliche Größe des Problems, also auf die Gesamtmasse des Eisbergs zu schließen, ist nicht möglich. Aber klar ist, dass dieser viel größer sein muss als das, was sichtbar aus dem Wasser ragt: Handelt es sich doch um sanktioniertes, wenn nicht gar justiziables Verhalten. Deshalb dürften vermutlich auch Umfragen wie die erwähnte niederländische ebenfalls zu niedrige Prävalenzen von Verstößen aufzeigen.

Aufgeschreckt und verunsichert durch die Zahlen aus den Niederlanden hat sich der Narr kürzlich selbst in der Literatur umgetan (siehe wie immer unter dimagl.com/lj) und fand überraschend viele Belege – Umfrage-Ergebnisse, Stichproben, systematische Reviews und anderes mehr –, die in ihrer Totalität nur einen

Schluss zulassen: Wissenschaftliches Fehlverhalten jenseits von HARKing und p-Hacking – also Plagiarismus, Falsifikation und Fabrikation von Daten – ist viel häufiger, als wir uns eingestehen.

Erhellende Hinweise darauf, warum das so ist, finden sich übrigens in der Autobiographie des bereits erwähnten Wissenschaftsbetrügers Diederik Stapel. Er beschreibt, wie leicht es ihm gefallen ist, durch nicht offenlegte Selektion von Daten und Analyseverfahren die „Stories“ seiner Paper interessanter zu machen – und sie dadurch in renommierten Journalen publizieren zu können. So fing er an, sich in der Psychologie einen Namen zu machen – die Tenure war greifbar. Der Übergang von der Selektion zur Manipulation seiner Daten war dann fließend. Niemand an der Uni und auch kein Reviewer fragte nach oder wollte Daten sehen. Das ging alles so einfach und glatt, dass Stapel allmählich dazu überging, Studienergebnisse komplett zu erfinden. Seine Studenten führten die Befragungen durch, und er hübschte die Daten dann in großem Stil auf. Damit wurden Stapels Ergebnisse letztlich so spektakulär, dass *Science* und *Nature* sie mit Handkuss nahmen. So erfand er zum Beispiel Daten, deren Auswertung zeigte, dass in einer vermüllten Umgebung Befragte eher zu rechtsextremen Antworten neigen als in einer sauberen. Über solche Ergebnisse berichtete sogar die *New York Times*, und in kurzer Zeit wurde er zum Shootingstar der Psychologie!

An einer anderen Stelle seiner Autobiographie beschreibt Stapel, dass er sich fühlte wie ein Kind, das man allein im Bobbon-



Code im Sack
20 €

laborjournal.de ↔ service ↔ shop



laden zurückgelassen hatte. Einzig mit dem Hinweis, doch bitte keine Süßigkeiten zu stibitzen. Was Stapel letztlich indes das Genick brach, waren seine eigenen Studenten. Die konnten sich zwar zunächst freuen, Koautoren auf tollen Papern zu sein, fanden es aber nach einiger Zeit befremdlich, die Daten nie selbst auswerten zu dürfen, sondern stattdessen immer nur bereits vom Chef prozessierte Daten zu Gesicht zu bekommen.

Sicher, der Fall Diederik Stapel ist extrem. Und wie er seine Verfehlungen ganz lässig auf das System abwälzt, das ihm das alles zu leicht gemacht habe, ist natürlich wohlfeil. Aber nichtsdestotrotz kann man an seiner Karriere sehr schön die wesentlichen allgemeinen Elemente des modernen Wissenschaftsbetrugs studieren:

» das auf einer Journal-Reputationsökonomie basierende akademische Belohnungssystem;

» Journale, die spektakuläre Studien so liden vorziehen;

» mit der Qualitätskontrolle überforderter Reviewer;

» Berufungskommissionen und universitäre Gremien, die sich von Storys und Selbstvermarktern blenden lassen;

» fragwürdige Wissenschaftspraktiken, die als normal gelten und nicht sanktioniert werden als „Einstiegsdrogen“;

» mangelhafte Diskussions- und Führungskultur in der Arbeitsgruppe;

» sowie methodische Inkompetenz bei allen Beteiligten.

»Wissenschaftsbetrug wird nicht nur selten aufgedeckt, sondern noch seltener geahndet.«

In den meisten Forschungseinrichtungen finden sich zu jeder Zeit mehrere Elemente dieses toxischen Gemisches. Wenn aber alle zusammenkommen, ist es nur noch eine Frage der Zeit, bis einzelne Wissenschaftler der Versuchung erliegen, ihrer wissenschaftlichen Karriere ein bisschen nachzuhelfen und Abkürzungen zu nehmen. Nur wenn sie es allzu doll treiben, wie eben Herr Stapel, müssen sie damit rechnen aufzufliegen. Und auch dann sind die Konsequenzen, falls es überhaupt zu Sanktionen kommt, recht überschaubar.

Besteht die Lösung des Problems also darin, Wissenschaftsbetrug härter zu sanktionieren? Schaden würde das sicher nicht. Schließlich kann man die Fälle, in denen bislang Strafen verhängt wurden, an einer Hand abzählen. Wissenschaftsbetrug wird also nicht nur selten aufgedeckt, sondern noch seltener geahndet.

Müssen wir mehr gute wissenschaftliche Praxis lehren und trainieren? Auch das ist eine gute Idee, aber sehr viel nützen wird es wohl nicht. Es gibt ja auch keine Kurse, in denen Schülerinnen und Studenten erklärt wird, dass Banküberfall und Urkundenfälschung gegen gesellschaftliche Normen verstoßen, daher verboten sind und konsequenterweise bestraft werden. Wissenschaftsbetrüger wissen, was sie tun – und tun es nicht aus Unkenntnis ihnen unbekannter Regeln.

Brauchen wir vielleicht eine Wissenschaftspolizei, die unangekündigte Kontrollen von Western Blots und Festplatten in Laboren durchführt? Ganz sicher nicht! Moderne Wissenschaft ist viel zu komplex, als dass sie durch solche Visiten kontrollierbar wäre. Ganz abgesehen davon, dass die dadurch entstehende Big-Brother-Atmosphäre alles andere als förderlich für gutes Forschen wäre.

Ein viel naheliegender Ansatz zur Abhilfe ist es, sich dem Kern des Problems anzunehmen und das toxische Karriere- und Bewertungssystem zu reformieren – also Forscher nicht auf Basis fragwürdiger Metriken, sondern mit Fokus auf Forschungsqualität, Inhalte und dem tatsächlichen wissenschaftlichen oder gesellschaftlichen Impact zu beurteilen. Das ist in der Tat der Königsweg, und in Ansätzen findet das zum Glück derzeit auch statt. Die von der Europäischen Union initiierte Coalition for Reforming Research Assessment (CoARA), der die DFG übrigens bereits beigetreten ist, wird hierbei eine wichtige Rolle spielen. Allerdings geschieht all dies eher im Schneckentempo, sodass ein schnellerer Fix erstrebenswert wäre.

Vielleicht gibt es den sogar! Wissenschaftsbetrug ist nämlich nur dort möglich, wo einzelne die Auswertung und Analyse von Forschungsdaten monopolisiert haben – und zudem häufig noch die methodische Kompetenz im unmittelbaren Umfeld fehlt. Nur wenn Western Blots lediglich von einer Person angefertigt und ausgewertet werden, und auch niemand sonst mit der nötigen Kompetenz draufschaut, dann können diese mittels Photoshop manipuliert werden. Analoges gilt für Datenreihen und die darauf angewendeten Analyseverfahren: Manipulationen sind dann besonders gut möglich, wenn nur eine Person die Datenbanken oder Spreadsheets verwaltet und selbstverfasste Codes darüber laufen lässt – und wenn die Ergebnisse nicht von anderen kontrolliert werden. Was ja schon alleine wegen der ehrlichen Fehler, die wir alle leider häufig machen, notwendig wäre. Wenn aber dann noch der Gruppenleiter einsam vor dem Rechner sitzt und die Ergebnisse in eine Story verwandelt, kann es durchaus passieren, dass ihm ein übereifriger Mitarbeiter ein „faules Ei“ vorlegt – oder dass umgekehrt er selbst allzu „kreativ“ aus den Ergebnissen eine Story bastelt.

In den Arbeitsgruppen braucht es folglich eine funktionierende Struktur und Arbeitskultur, dann ist wissenschaftliches Fehlverhalten praktisch ausgeschlossen. Problematisch wird es nämlich immer dann, wenn Arbeitsgruppen zu groß werden, die Expertise zu fragmentiert ist oder punktuell gar komplett fehlt. Leider sind das Bedingungen, die gerade in der biomedizinischen Forschung nicht wirklich selten sind.

Wie kann da Abhilfe geschaffen werden? Auf jeden Fall durch Thematisierung und Fokus auf gute Arbeitskultur und Gruppenleitung, wo immer das möglich ist. Dazu natürlich in der Ausbildung, wo diese Inhalte meist zu kurz kommen. Zwar bieten viele Unis im Rahmen der Personalentwicklung ein „Führungskräfte-training“ an, darin müsste aber ein stärkerer Fokus auf die Wichtigkeit von offener und kollaborativer Arbeitsweise als Bollwerk gegen wissenschaftliches Fehlverhalten gelegt werden.

»Ein breiter Diskurs über die Arbeitskultur in wissenschaftlichen Arbeitsgruppen würde uns sicher schon weiterbringen.«

Aber auch bei Berufungen und Tenurisierung sollten wir uns stärker mit dem Thema befassen. Die zuständigen Kommissionen könnten Kandidaten gezielt über Größe, Struktur und den Interaktionen in ihrer Arbeitsgruppe befragen. Sie könnte noch Gespräche mit ehemaligen (oder auch noch aktiven) Mitgliedern der Arbeitsgruppe führen. An mancher Stelle wird dies bereits praktiziert, zum Beispiel bei EU-LIFE, einer Allianz von renommierten europäischen Forschungsinstituten. Bei klinischen Berufungen ist es übrigens gängige Praxis, dass Berufungskommissionen die Abteilung der Bewerber aufsuchen und sich vor Ort einen Einblick in deren „Arbeitsweise“ verschaffen. Ein Vorschlag zur Steuerung der Arbeitsgruppensdynamik durch die Universitäten wäre zum Beispiel das Kappen der leistungsorientierten Mittelvergabe (LOM) ab einer gewissen Gruppengröße.

Ob all dies jedoch eher närrische und letztlich unrealistische oder ineffektive Maßnahmen sind, und ob damit wissenschaftliches Fehlverhalten tatsächlich vermindert werden könnte, muss offen bleiben. Allein ein breiter Diskurs über die Arbeitskultur in wissenschaftlichen Arbeitsgruppen und wie wir sie verbessern können, würde uns aber sicher schon weiterbringen.

Weiterführende Literatur und Links finden sich wie immer unter: <http://dirnagl.com/lj>.