Zeitschrift: Schweizerische Zeitschrift für Soziologie = Revue suisse de sociologie = Swiss journal of sociology

Herausgeber: Schweizerische Gesellschaft für Soziologie

Band: 31 (2005)

Heft: 1

Artikel: "Betrug und Täuschung in der Wissenschaft" : Datenfälschung, Diagnoseverfahren, Konsequenzen

Autor: Diekmann, Andreas

DOI: https://doi.org/10.5169/seals-815064

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 13.12.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

«Betrug und Täuschung in der Wissenschaft». Datenfälschung, Diagnoseverfahren, Konsequenzen

Andreas Diekmann*

1 Ist Datenfälschung überhaupt ein Problem?

Charles Babbage, Mathematiker in Cambridge und einer der Pioniere beim Bau von Rechenmaschinen, hat 1830 eine Denkschrift unter dem bemerkenswerten Titel «On the Decline of Science in England» vorgelegt. In dieser kritischen Abrechnung mit den Verkrustungen und der Günstlingswirtschaft der «Royal Academy» werden vier Typen von Fälschungen an Beispielen erläutert, an denen es auch zu seinen Zeiten ganz offensichtlich nicht mangelte. Diese Typologie umfasst 1. «hoax», 2. «forgery», 3. «trimming» und 4. «cooking». «Hoax» ist eine Fälschung, um die Gutgläubigen zu blamieren, «forgery» die Datenfälschung, «trimming» die Retusche der Messfehler und «cooking» die Selektion der passenden Daten. In diesem Beitrag wird es hauptsächlich um «forgery», die Fälschung von Daten und Ergebnissen gehen.

Sind Fälschungen überhaupt ein Problem? Wenn postmoderne Wissenschaftsphilosophen von der konstruktivistischen Aushandlung wissenschaftlicher Befunde sprechen oder Paul F. Feyerabend (2001) fröhlich «Wider den Methodenzwang» streitet, dann könnte doch zu einem «Anything Goes» auch gleich die «Methode» der Fälschung zählen? Die Wissenschaftsgeschichte bietet zudem reichlich Beispiele dafür, dass paradoxerweise Fälschungen der Wahrheit zum Durchbruch verholfen haben, wie die Korrekturen und Mogeleien sogar eines Galilei oder Newton demonstrieren (Broad und Wade, 1984; DiTrocchio, 1994; Westfall, 1973). Die Wissenschaftsphilosophie des «Kritischen Rationalismus» von Karl R. Popper sollte mit Fälschungen auch kein grosses Problem haben. Hypothesen, die sich nur durch betrügerische Datenmanipulation scheinbar bewährt haben, sollten im evolutionären Wettstreit der Theorien keine wirkliche Überlebenschance haben. So lautet auch eine gebräuchliche Beschwichtigungsformel, dass Betrug bei unwichtigen Frage-

* Prof. Andreas Diekmann, Professor für Soziologie, Scheuchzerstrasse 68/70, ETH-Zentrum, SEW E 26, CH–8092 Zürich; Tel.: +41 1 632 55 59; E-Mail: andreas.diekmann@soz.gess.ethz.ch.

Der Obertitel ist dem klassischen Werk von Broad und Wade (1984) entnommen. Ich bedanke mich bei Wojtek Przepiorka für die Recherche im Social Science Citation Index und die Unterstützung bei der Anfertigung von Grafiken. Für Anregungen und Hinweise gilt mein Dank Prof. Rudolf Hagemann, Halle. Es handelt sich um die überarbeitete Fassung eines Vortrags vor der Deutschen Akademie der Naturforscher Leopoldina und an der ETH Zürich (veröffentlicht im Jahrbuch 2003 der Akademie, vgl. Diekmann 2004a).

stellungen zwar möglicherweise nie entdeckt, aber dann eben auch irrelevant sei. Bei wichtigen Fragen aber wird der Schwindel über kurz oder lang auffliegen. Freilich gilt das nur auf lange Sicht. Auch ein optimistischer Popperianer wird zugeben, dass Fälschungen das Ableben falscher Hypothesen verzögern können. Lügen haben oftmals ziemlich lange Beine.

Vor allem aber schaffen Betrug und Fälschung ein Vertrauensproblem. Wer die Regeln wissenschaftlicher Ethik missachtet und z. B. Daten fälscht, untergräbt damit gleichzeitig das Vertrauen in die wissenschaftliche Forschung schlechthin. Die Einhaltung wissenschaftlicher Ethik lässt sich als Kollektivgut charakterisieren, das durch betrügerisches, nicht-kooperatives Verhalten Schaden nimmt. Vertrauen, und vielleicht noch schlimmer: das Vertrauen von Wissenschaftlern in die Ergebnisse von Kollegen, kann sehr rasch zerstört werden, wenn im Wettbewerb um knappe Ressourcen, um Forschungsgelder, Karrieren und Ansehen, die Normen wissenschaftlicher Redlichkeit erodieren (z. B. DFG, 1998; Mosimann et al., 1995).

Es ist aber nicht nur der kollektive Vertrauensverlust in Wissenschaft und Öffentlichkeit zu beklagen. Auch die einzelnen Konsequenzen wissenschaftlicher Fälschungen, z. B. in der biomedizinischen Forschung, sind keineswegs gering zu veranschlagen. Wenn z. B. eine Behandlungsmethode mit gefährlichen Nebenwirkungen zugelassen wird, weil Untersuchungen betrügerisch waren, dann kann der Schaden für Leib und Leben der Patienten beträchtlich sein. So verschwieg die Tabakindustrie Studien, die frühzeitig das Suchtpotenzial von Nikotin erkannten. Man hielt aber die Ergebnisse nicht nur geheim, sondern nutzte sie auch bewusst – quasi zur Verstärkung der «Kundenbindung». Nicht zuletzt auf diesen Wissenschaftsbetrug stützten sich die grossen Schadenersatzklagen gegen die Tabakindustrie (Hurt und Robertson, 1998).

Zur Vermeidung von Wissenschaftsbetrug wird man sich nicht allein auf die Internalisierung wissenschaftlicher Redlichkeit, d. h. auf die Einübung wissenschaftlicher Tugenden verlassen, obwohl dieser Weg auch nicht ganz zu verachten ist. Im Allgemeinen aber sind Wissenschaftler nicht mehr oder weniger moralisch als Wirtschaftsleute, Politiker, Kirchenvertreter oder Angehörige anderer Professionen. Wie alle Menschen reagieren auch Wissenschaftlerinnen und Wissenschaftler auf positive und negative Anreize. Es sind vor allem die institutionellen Regeln der Kontrolle, Selbstkontrolle und die Sanktionspraktiken, welche die Ausbreitung von Betrug und Täuschung in einer zunehmend wettbewerbsorientierten Gemeinschaft von Wissenschaftlern verhindern können.

Sanktionen für «wissenschaftliches Fehlverhalten», wie es heute etwas euphemistischer heisst, riskiert man jedoch nur dann, wenn der Schwindel auffliegt. «Auch die Nürnberger hängen keinen, es sei denn sie hätten ihn». Deshalb ist das Problem der Kontrolle ein zentraler Punkt. Und zwar nicht nur zur Verhinderung von Betrug und Fälschung, sondern auch um Irrtümern durch methodisch unsaubere Praktiken begegnen zu können.

Dieses Thema, die Frage nach den Möglichkeiten der Aufdeckung von Betrug und Täuschung, steht im Mittelpunkt dieses Artikels. Zunächst werde ich einige Fallbeispiele aus den Naturwissenschaften diskutieren. Behauptungen über den Umfang von Wissenschaftsbetrug nach dem Motto «Randerscheinung oder Spitze eines Eisbergs?» sind heute reine Spekulation. Deshalb möchte ich zweitens zeigen, dass die Methoden der Sozialforschung zumindest Hinweise auf das Ausmass von Fälschungen liefern können. Drittens stellt sich die Frage, ob und warum die Natur- und die Sozialwissenschaften in unterschiedlicher Weise von Problemen des Datenbetrugs betroffen sind. Viertens werde ich Methoden der Diagnose von Datenfälschung ansprechen. Die Weiterentwicklung solcher Methoden könnte die Kontrolle der Datenqualität wesentlich unterstützen. Abschliessend werden einige institutionelle Vorkehrungen zur Diskussion gestellt, die dazu beitragen können, das Vorkommen betrügerischer Datenmanipulationen und fehlerhafter Datenanalysen zu verringern.

2 Nicht nur die Biomedizin: Die Betrugsfälle Herrmann und Schön

Die Diskussion über Fälschungen begann in den USA bereits in den 1970er-Jahren. In Europa tat man das Problem von Fälschungen lange Zeit als Randerscheinung ab. Aufgeschreckt wurden Wissenschaft und Öffentlichkeit erst 1997 durch die spektakulären Fälschungen der Forschungsgruppe von Friedhelm Herrmann, einem der damaligen Stars der deutschen Krebsforschung. Glätteten sich die Wogen danach wieder und glaubte man, dass vor allem die biomedizinische Forschung betroffen sei, wurde man rasch eines Besseren belehrt. Im Mai 2002 wurde der Verdacht geäussert, im Herbst des Jahres lag der Nachweis für einen der grössten Fälschungsskandale in der physikalischen Grundlagenforschung vor, die Fälschungen des jungen deutschen Physikers Jan Hendrik Schön in den US-amerikanischen Bell Labs in New Jersey. Zu Beginn seiner Tätigkeit in den Bell Labs 28 Jahre jung, umfasste die Publikationsliste bei Entdeckung des Betrugs mehr als 100 Titel, darunter einige Artikel mit nobelpreisverdächtigen Ergebnissen in Science und Nature. Natürlich ist man im Nachhinein immer klüger. Die besten Diagnosen stellt bekanntlich der Pathologe. Liest man aber die Berichte der Untersuchungskommissionen zum Wissenschaftsbetrug von Friedhelm Herrmann und Jan Hendrik Schön, dann scheinen tatsächlich einige Sicherungen durchgebrannt zu sein. Die «durchgebrannten Sicherungen» sind hier vor allem Herausgeber von Fachzeitschriften, interne Kontrollen im Team, Gutachter und Ko-Autoren. Natürlich können Herausgeber normalerweise nicht die Fälschung von Rohdaten oder Ergebnissen erkennen. Sie können aber die Aufbewahrung und Dokumentation der experimentellen Daten verlangen. Bei mehreren untersuchten Veröffentlichungen von Schön waren die Daten nicht mehr auffindbar.

Abbildung 1: Typische Datenfälschung in einer Publikation von F. Herrmann

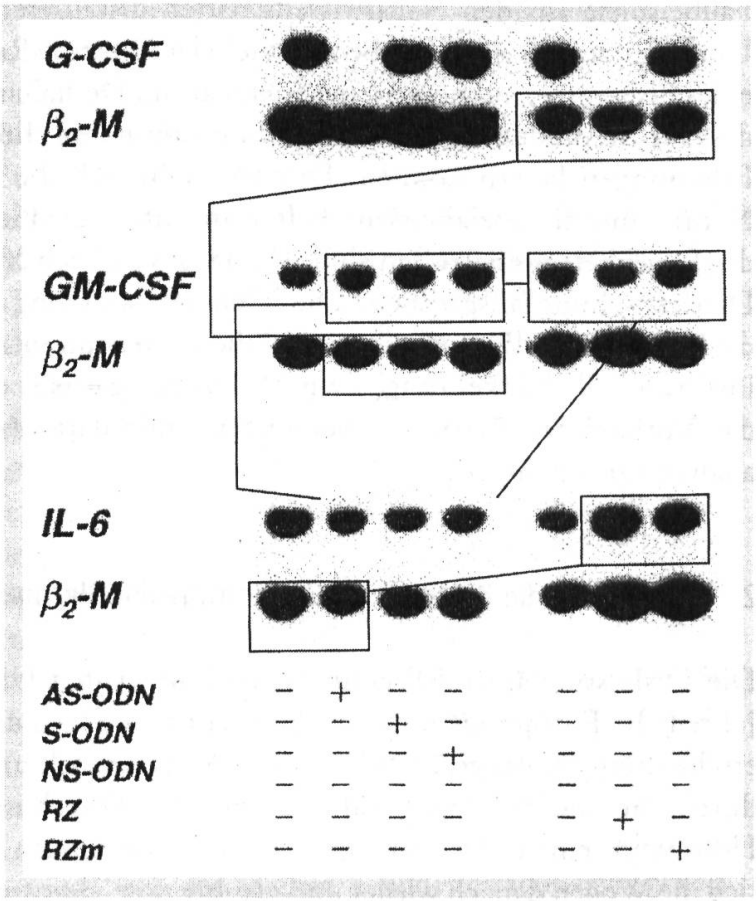


Abbildung 1: Ausschnitt aus Abb. 5 der Publikation TF 421, ergänzt um einige Rahmen und Pfeile, die die verschiedenen Duplikationen innerhalb dieser eindeutig gefälschten Abbildung verdeutlichen. Man sieht, dass einige der «Banden» genannten Flecken untereinander eine so hohe Ähnlichkeit aufweisen, dass auszuschliessen ist, dass sie, wie vorgegeben, aus unterschiedlichen experimentellen Bedingungen hervorgegangen sind. Die vollständige Dokumentation zu Abb. 5 aus TF 421 inklusive Legende findet sich in Anlage 8-279.

Quelle: Abschlussbericht der Taskforce F. H., Rapp et al., 1999.

Ebenso können Gutachter nicht wissen, ob Daten gefälscht wurden. Grobe Auffälligkeiten und Inkonsistenzen aber, die bei Friedhelm Herrmann, Jan Hendrik Schön und in vielen anderen Fällen von Wissenschaftsbetrug vorlagen, hätten Gutachter bei genauerem Hinsehen entdecken können. Und was für die Gutachter gilt, gilt erst recht für Koautoren. Allerdings fallen Unstimmigkeiten oft erst dann auf, wenn die Suche durch einen Verdacht gelenkt wird. Kann man von Gutachtern mit begrenztem Zeitbudget verlangen, dass sie bei ihrer Arbeit, dem «peer-review», die Möglichkeit von Fälschungen immer mitbedenken?[1]

In den inkriminierten Arbeiten von Herrmann bestanden die Fälschungen vorwiegend in der Manipulation von Grafiken (Abbildung 1). So genannte «Banden» oder «Blots», sie sind sozusagen die Fingerabdrücke bestimmter Moleküle, wurden

1 Die Krebsforscherin Karin Mölling berichtete auf einem Berliner Symposium, dass ihr erst nach mehrfacher Durchsicht eines Artikels von Herrmann aufgefallen sei, dass an drei Stellen die gleichen Abbildungsteile zu erkennen waren (Ronzheimer, 1998).

in passender Weise in andere Bilder hineinkopiert. Die zur Aufdeckung der Fälschungen gebildete «Task Force» fand in akribischer Arbeit ca. 200 solcher Manipulationen, oftmals mehrere in den Abbildungen des gleichen Artikels (Rapp et al., 2000). Ko-Autoren und Gutachter hatten die Manipulationen in keinem der zahlreichen Artikel über einen Zeitraum von mehr als einem Jahrzehnt bemerkt. Alarm schlug schliesslich der junge und mutige Post-Doktorand Eberhard Hildt, nachdem ihm Widersprüche zwischen Laborexperimenten und den zur Publikation vorgesehenen Befunden aufgefallen waren.

Die Entdeckung weitgehend übereinstimmender Abschnitte von Messkurven brachte auch Jan Hendrik Schön zur Strecke (Beasley et al., 2002). Die Abbildungen aus Nature von Oktober 2001 und aus Science von Dezember 2001 stammten aus kurz nacheinander publizierten Artikeln mit den gleichen Ko-Autoren. In Abbildung 2 wurden die Ausschnitte aus den Originalgrafiken stark vergrössert. Mit blossem Auge ist die weitgehende Übereinstimmung der Verläufe nicht einfach zu erkennen. Die Beasley-Kommission deckte bei der Untersuchung des Fälschungsverdachts eine Reihe weiterer Unstimmigkeiten auf. Natürlich ist eine Kommission im Nachhinein im Vorteil, da sie das Material unter der Perspektive des Fälschungsverdachts sichtet, während Ko-Autoren in der Regel einander vertrauen, dass sie wissenschaftlich korrekt arbeiten. Die Ko-Autoren waren bei den Experimenten nicht zugegen, aber hätten sie nicht doch einige der Unstimmigkeiten bemerken können, von denen die Beasley-Kommission nach einer kritischen Analyse der Artikel berichtete?[2]

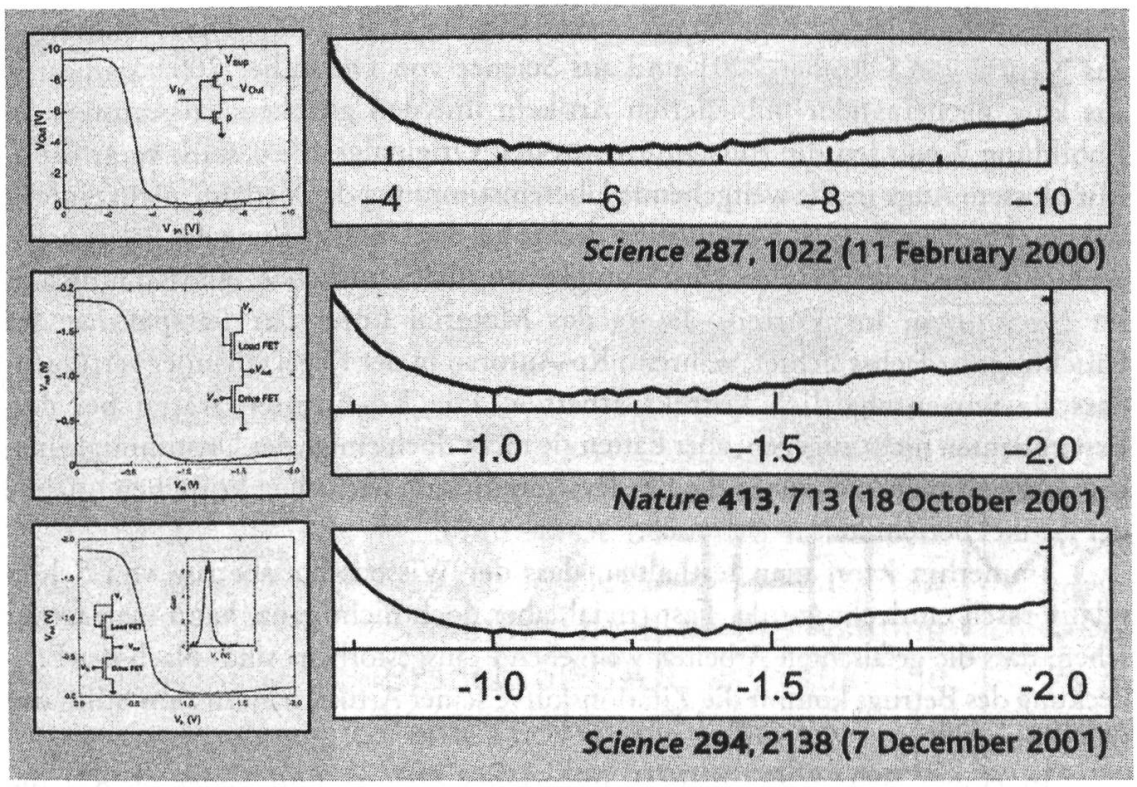
Immerhin kann man festhalten, dass der Wissenschaftsbetrug von Schön relativ rasch entdeckt wurde. Fast trivial, aber doch nicht ganz, kann man ferner sehen, dass die gefälschten Arbeiten weitgehend «ausgestorben» sind. Nach der Entdeckung des Betrugs kommt die Zitationskurve seiner Artikel nahezu zum Stillstand (Abbildung 3).

In einem Beitrag für Nature analysieren Stewart und Feder (1987) die gefälschten Forschungsarbeiten des jungen Kardiologen John Darsee. In vielen der gefälschten Publikationen in z. T. führenden Fachzeitschriften zeigten sich zahlreiche Fehler und Diskrepanzen, die einem aufmerksamen Leser bei einer genaueren Durchsicht der Artikel hätten auffallen können. Hier ist eine Kostprobe aus dem New England Journal of Medicine (Abbildung 4). In dem abgebildeten

2 Die von den Bell-Labs eingesetzte Kommission (Beasley et al., 2002) hat alle Ko-Autoren vom Vorwurf der Fälschung eindeutig freigesprochen. Der Freispruch erstreckt sich aber nicht auf eine eventuelle professionelle Mitverantwortung. Im Bericht heisst es: «The Committee find all coauthors of Hendrik Schön in the work in question completely cleared of scientific misconduct.» Und weiter heist es: «There is no implication here of scientific misconduct; the issue is one of professional responsibility.» Hier stellt die Kommission die Frage, ob «Bertram Batlogg, as the distinguished leader of the research, took a sufficiently critical stance with regard to the research in question ...» Meines Erachtens hätte die ETH Zürich dieser Frage im Interesse der Beteiligten in einem transparenten Verfahren nachgehen sollen.

Stammbaum findet man eine Familie mit vier Kindern im Alter von acht, sieben, fünf und vier Jahren. Der Vater ist 17, hätte also sein erstes Kind mit acht oder neun und das zweite Kind mit neun oder zehn Jahren gezeugt haben müssen. Da das Alter des Vaters noch zwei Mal im Text auftritt, handelt es sich kaum um einen Druckfehler.

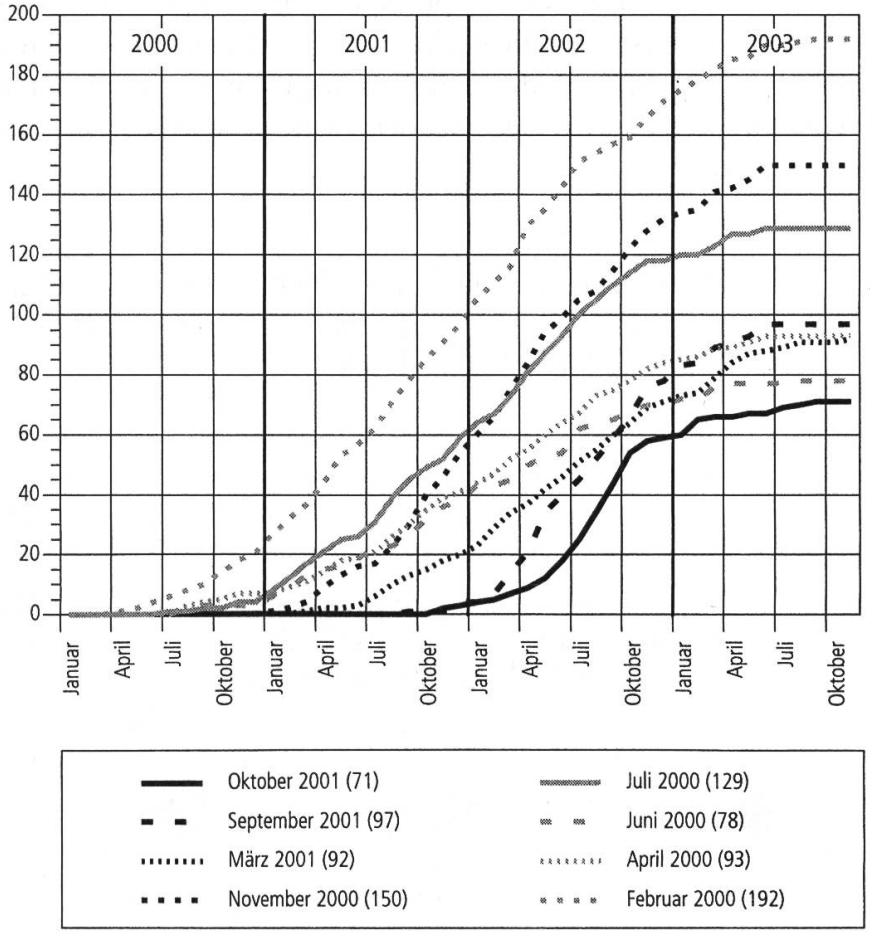
Abbildung 2: Fälschung der Abbildung von Messwerten in Publikationen von J. H. Schön



Striking resemblance. Published data from studies of different devices revealed a similarity in recorded «noise.» Schön says the bottom figure was sent to Science by mistake (see correction, p. 1400).

Quelle: Figure 1: Aus, Service, Robert F., 2002 (MAY): Pioneering Physics Papers Under Suspicion for Data Manipulation, Science 296: 1376–1377.
Oben: Schön J.H., Berg S., Kloc C., Batlogg B., 2000 (FEB): Ambipolar pentacene field-effect transistors and inverters, Science 287: 1022–1023.
Mitte: Schön J.H., Meng H., Bao Z., 2001 (OCT): Self-assembled monolayer organic field-effect transistors, Nature 413: 713–716.
Unten: Schön J.H., Meng H., Bao Z., 2001 (DEC): Field-effect modulation of the conductance of single molecules, Science 294: 2138–2140.

Abbildung 3: Zitationen von Schön-Artikeln im Science Citation Index (SCI)



Kumulierte Anzahl Zitationen in Klammern.

1. Schön J.H., Meng H., Bao Z., 2001 (OCT): Self-assembled monolayer organic field-effect transistors, Nature 413: 713–716.
2. Schön J.H., Kloc C., Batlogg B., 2001 (SEP): High-temperature superconductivity in lattice-expanded C-60, Science 293: 2432–2434.
3. Schön J.H., Dodabalapur A., Bao Z., Kloc C., Schenker O., Batlogg B., 2001 (MAR): Gate-induced superconductivity in a solution-processed organic polymer film, Nature 410: 189–192.
4. Schön J.H., Kloc C., Batlogg B., 2000 (NOV): Superconductivity at 52 K in hole-doped C-60, Nature 408: 549–552.
5. Schön J.H., Kloc C., Dodabalapur A., Batlogg B., 2000 (JUL): An organic solid state injection laser, Science 289: 599–601.
6. Schön J.H., Kloc C., Batlogg B., 2000 (JUN): Fractional quantum Hall effect in organic molecular semiconductors, Science 288: 2338–2340.
7. Schön J.H., Kloc C., Haddon R.C., Batlogg B., 2000 (APR): A superconducting field-effect switch, Science 288: 656–658.
8. Schön J.H., Berg S., Kloc C., Batlogg B., 2000 (FEB): Ambipolar pentacene field-effect transistors and inverters, Science 287: 1022–1023.

Es ist immerhin verblüffend, wie plump Wissenschaftsfälschungen konstruiert werden. Oder aber noch schlimmer, die plumpen Fälschungen werden entdeckt und bilden die Spitze des Eisbergs, die raffinierten Fälschungen kommen dagegen selten oder nie ans Tageslicht. Auf jeden Fall zeigt ein Grossteil der bekannt gewordenen Fälschungen, dass die bestehenden Institutionen der Selbstkontrolle offenbar unzureichend sind.

Abbildung 4: 17-jähriger Vater mit vier Kindern im Alter von acht, sieben, fünf und vier Jahren in einer Publikation von Darsee und Heymsfield im New England Journal of Medicine

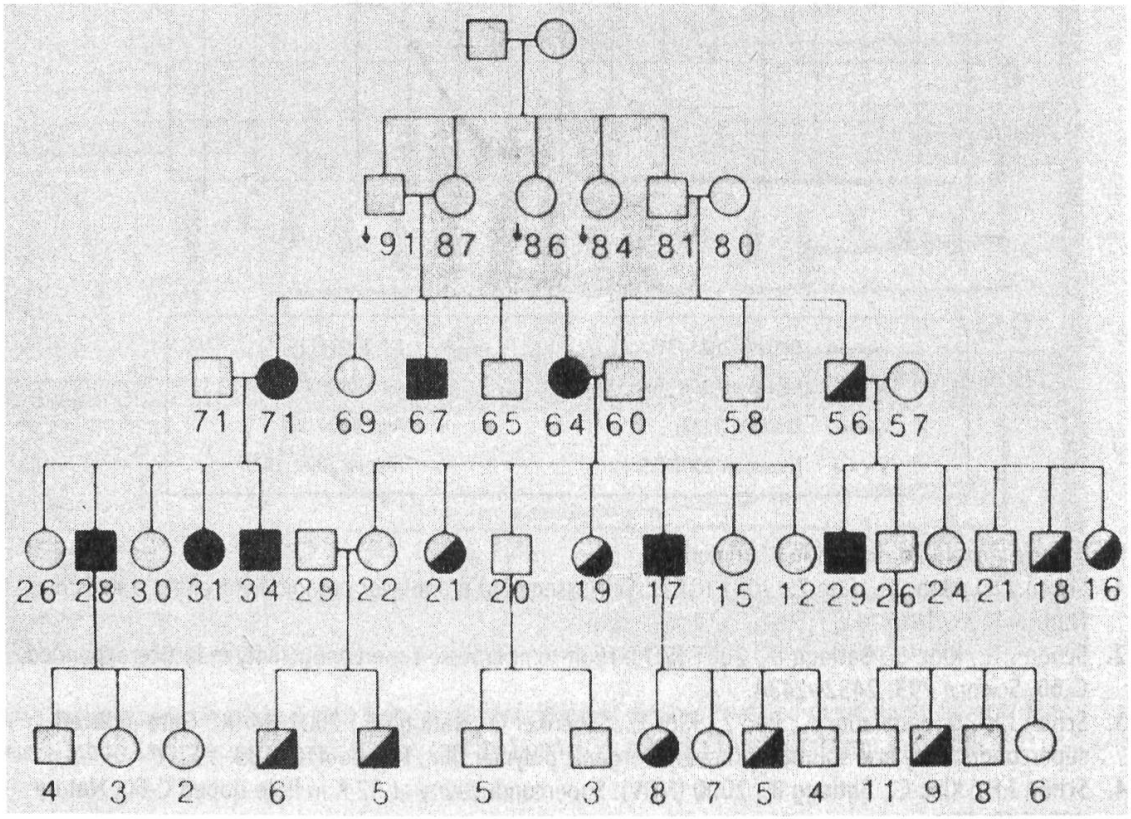


Fig. 2 Pedigree reproduced from ref. 87. The age of each family member at the start of the investigation is given below each symbol. Arrows (second row) denote those who had died before the start of the investigation. Note the 17-year-old male (row 4) with children ages 8,7,5 and 4; at the age of 8 or 9 (probably 8) he impregnated the mother of his oldest child.

3 Schätzung von Fälschungen mit Methoden der Sozialforschung

Handelt es sich um Randerscheinungen oder die Spitze eines Eisbergs (vgl. Mayntz, 1999)? Hat das Ausmass an Wissenschaftsbetrug in einem stärker auf Konkurrenz und Karriere bezogenen Wissenschaftssystem zugenommen?

Der Wissenschaftshistoriker DiTrocchio (1994), der ebenfalls «Big Science» verantwortlich macht, differenziert nochmals nach dem Motiv. Die vormodernen Schwindler hätten aus der ehrenhaften Absicht des festen Glaubens an ihre Theorie gehandelt. Hier spricht man auch lieber von Mogeleien, die als legitimes Mittel galten, der Theorie im Kreise der Widersacher und Skeptiker Anerkennung zu verschaffen. In «Big Science» geht es den Fälschern dagegen nicht um die Wahrheit von Theorien, sondern schlicht und schnöde um Forschungsgelder und Karrieren.

Leider sind die gegensätzlichen Vermutungen bezüglich des Ausmasses von Fälschungen reine Spekulation. Zwar wird bei einem exponenziellen Wachstum der Anzahl von Wissenschaftlern und Publikationen auch die absolute Anzahl von Betrügereien zugenommen haben. Ob dies aber auch für den Anteil von Publikationen mit gefälschten Daten gilt, kann niemand mit Sicherheit bezeugen.

Nach Robert K. Mertons (1957) Kriminalitätstheorie steigt die Wahrscheinlichkeit abweichender Handlungen, wenn die Diskrepanz zwischen Zielen und legalen Mitteln der Zielerreichung wächst. Bei wachsender Konkurrenz um die Verteilung knapper Forschungsmittel und Positionen würde man entsprechend eine Zunahme betrügerischer Aktivitäten erwarten. Da aber keine empirisch gesicherten Angaben zum Ausmass von Fälschungen vorliegen, bleibt dies eine zwar plausible, aber unbewiesene Hypothese.

Niemand kann heute auch nur ungefähr das Ausmass von Datenbetrug in den einzelnen Wissenschaftsdisziplinen angeben. Wenn man wollte, könnte man aber mit geeigneten Methoden durchaus genauere Hinweise erhalten.

In der Kriminologie werden seit langem mit den Methoden der empirischen Sozialforschung so genannte Dunkelfelduntersuchungen durchgeführt. Anonymisierte Befragungen zur selbst berichteten Delinquenz oder viktimologische Studien liefern ungefähre Schätzungen zum Ausmass krimineller Handlungen, die von Kriminalstatistiken nur selektiv erfasst werden. Erstaunlicherweise hat man diese Methoden, von ganz wenigen Ausnahmen abgesehen, nicht auf die Untersuchung von Wissenschaftsbetrug angewendet.

Eine solche Ausnahme ist eine Umfrage der Herausgeber der Zeitschrift Chemical Engineering aus dem Jahre 1980, die allerdings nicht nach den strengen Kriterien der Methoden der empirischen Sozialforschung durchgeführt wurde. Gleichwohl liefert sie einige Hinweise. U. a. wurde die folgende hypothetische Frage gestellt:

«Auf der Suche nach einem neuen katalysierenden Stoff hat Jays Team die Spannbreite der Möglichkeiten auf die Alternativen A und B eingegrenzt. Vermutlich

Fehler behaftete Daten legen die Wahl von B nahe. Aber Jay, sein Team und sein Chef, der ein Experte auf dem Gebiet ist, sind sicher, dass A die bessere Wahl ist. Jays Chef bittet ihn, die Daten zu fälschen und einen Bericht zu schreiben, der die Wahl von A favorisiert» (Hughson und Kohn, 1980, 137, Übersetzung A. D.).

4318 Leser von Chemical Engineering haben geantwortet. Kaum zu glauben, aber der Anteil der Chemiker, die bereit waren, einen Report auf der Basis gefälschter Daten zu schreiben, betrug immerhin 23,2 Prozent! Allerdings ist die Stichprobe nicht kontrollierbar und selbstselektiv wie bei allen Leserumfragen, bei denen der Fragebogen einfach in der Zeitschrift publiziert wurde.

Meist werden Fälschungen durch Indizien und nicht durch Geständnisse aufgedeckt. Ist es also naiv, Wissenschaftler einfach zu fragen, ob sie schon einmal Daten gefälscht haben? Selbst wenn man Anonymität zusichert, wird man wohl kaum korrekte Antworten erhalten. Nun gibt es aber in der Sozialforschung eine Methode für heikle Fragen, die bei korrekter Durchführung absolute Anonymität garantiert. Eine verständige Person weiss, dass niemand anders, nicht einmal der Interviewer, jemals die Antwort dekodieren kann. Die Methode heisst Randomized-Response-Technik oder kurz RRT.[3] Bei dieser Technik wird zunächst ein Zufallsexperiment mit z. B. zwei Ausgängen durchgeführt. Entscheidend ist, dass nur die befragte Person das Ergebnis des Zufallsexperiments kennt. Z. B. kann man die Person bitten, verdeckt eine Münze zu werfen. Bei «Kopf» soll die Person «ja» angeben, bei Zahl dagegen auf die gestellte Frage antworten. Niemand kann dann die Bejahung einer heiklen Frage individuell zuordnen. Man kann aber den Anteil von Personen in der Stichprobe, welche die heikle Frage bejaht haben, durch eine einfache Rechnung abschätzen. Mit einer Variante dieser Methode hat eine Gruppe von Ökonomen kürzlich die Kolleginnen und Kollegen auf einem Kongress zu verschiedenen Formen wissenschaftlichen Fehlverhaltens befragt (List et al., 2001). Verteilt wurde ein schriftlicher Fragebogen, der nach Beantwortung anonym in eine Sammelbox gelegt oder anonym per Post retourniert werden konnte. Die Wahrscheinlichkeit eines «Zufalls-Ja» wurde mit einem Achtel festgelegt. Eine der Fragen lautete: «Haben Sie jemals im Rahmen der Forschung Daten gefälscht?» In der RRT-Stichprobe von 140 Personen ergab sich ein geschätzter Anteil von Ja-Antworten auf die Fälschungsfrage von immerhin 4,5 Prozent. Einen fast genauso grossen Wert, nämlich 4,3 Prozent, erhielt die Forschungsgruppe für die Vergleichsstichprobe von 94 schriftlich befragten Kongressteilnehmern, deren Angaben ohne zusätzliche Anonymisierung durch «randomized response» erhoben wurden. Natürlich besagt die Schätzung nicht, dass 4,5 Prozent der Publikationen auf gefälschten Daten beruhen. Gefragt wurde ja, ob jemals eine Fälschung begangen wurde. Das konnte auch ein einmaliges Delikt sein, z. B. bei einer Forschungsarbeit,

3 Die Technik wurde von Warner (1965) eingeführt. Zu einem kurzen Überblick siehe Diekmann, 2003, 418–424 oder ausführlicher Fox und Tracy, 1986.

die nie publiziert wurde. Ausserdem war die Stichprobe relativ klein und die Schätzungen sind entsprechend fehlerbehaftet. Dennoch liefern sie einen Hinweis, dass auch eine der Todsünden in der Forschung, die bewusste Fälschung von Daten, möglicherweise doch nicht so extrem selten auftritt. Vor allem aber weist die Studie einen Weg, dass es mit clever angewandten Methoden der Sozialforschung gelingen kann, etwas Licht in die Dunkelziffer des Wissenschaftsbetrugs zu bringen.

4 Fälschungen in den Sozialwissenschaften

Anders als in den Naturwissenschaften sind spektakuläre Datenfälschungen in den Sozialwissenschaften eine Rarität.[4] Dies obwohl jährlich in den Fachzeitschriften der Ökonomie, Politikwissenschaften, Soziologie und anderen empirischen Sozialwissenschaften tausende von Artikeln mit Tabellen multivariater Schätzungen mit ökonometrischen Methoden erscheinen.

Eigentlich erstaunlich, dass nur äusserst wenige Fälschungen, z. B. solcher Tabellen, in den empirischen Sozialwissenschaften bekannt geworden sind. In der Ethnologie gibt es einige Fälle von Feldbeobachtungen, die frei erfunden wurden. Carlos Castañedas «Don Juan» ist ein Beispiel neueren Datums. In einem Aufsatz von Gerhard Fröhlich (2001) wird Castañeda, der übrigens mit seinen Phantasiegeschichten bei Harold Garfinkel in Soziologie promoviert wurde, ziemlich treffend als der «Karl May der Ethnologie» tituliert (vgl. Dürr, 1987). Zu den wenigen bekannt gewordenen Fällen von Datenbetrug in den empirischen Sozialwissenschaften, die mit statistischen Schätzmethoden arbeiten, zählen die erfundenen Zwillinge des britischen Psychologen Cyril Burt. Sir Cyril Burt war bis in die sechziger Jahre des vergangenen Jahrhunderts der führende britische Psychologe, dessen Studien zur Intelligenzforschung massgeblichen Einfluss auf das britische Schulsystem ausgeübt hatten. Burt verglich u. a. die Intelligenzleistungen gemeinsam und getrennt aufgewachsener eineiiger Zwillinge, um Anhaltspunkte für das Ausmass der Vererbung von Intelligenz zu erhalten. 1955 konnte Burt für 21 Paare von getrennt aufgewachsenen Zwillingen eine Korrelation von 0,771 und für 83 Paare gemeinsam aufgewachsener Zwillinge den Wert von 0,944 angeben. Dieser Befund spricht für die These eines sehr hohen Ausmasses der Erblichkeit von Intelligenz.

4 Der Ombudsmann der DFG Hans-Heinrich Trute (2003) hält fest: «Nach Disziplinen unterteilt stellen wir fest, dass die meisten Fälle wissenschaftlichen Fehlverhaltens sich im Bereich der Naturwissenschaften, vor allem aber in der medizinischen Forschung abspielen. Die Anrufungen aus dem Bereich der Medizin machen etwa ein Drittel aus; die der Naturwissenschaften insgesamt mehr als die Hälfte. Anrufungen aus den Bereichen der Geistes-, Sozial- und Wirtschaftswissenschaften sind hingegen deutlich seltener.» Diese Bemerkungen beziehen sich allerdings nicht nur auf Datenfälschungen. Und wohlgemerkt handelt es sich nur um die dem Ombudsmann bekannt gegebenen Fälle.

Der Haken ist nur, dass Burt seine Zwillinge weitgehend erfunden hat. Dem Psychologen Leon Kamin gebührt das Verdienst, dass ihm nach kurzer Lektüre in den Artikeln von Burt eine Merkwürdigkeit auffiel, die offenbar alle anderen Leser – Gegner wie Befürworter der Thesen von Burt – übersehen hatten.

Die gleichen Werte der Korrelation stehen bis auf die dritte Stelle hinter dem Komma in einer Publikation aus dem Jahre 1966, nur hatte Burt hier die getrennt aufgewachsenen Zwillinge bereits auf 53 Paare und die gemeinsam aufgewachsenen Zwillinge auf 95 Paare aufgestockt (Abbildung 5). Getrennt aufgewachsene eineiige Zwillinge kommen nicht gerade häufig vor. Kritiker wunderten sich, wie es Burt gelungen sei, so viele Versuchspersonen dieser seltenen Spezies aufzustöbern. Dass dann aber die Korrelationskoeffizienten der unterschiedlichen Stichproben exakt gleich waren, ist einfach extrem unwahrscheinlich. Hier haben wir im Übrigen wieder ein Beispiel für eine äusserst plumpe Fälschung, die allein durch Konsistenzprüfungen verschiedener Artikel erkennbar ist. Heute weiss man zudem aufgrund der Tagebuchaufzeichnungen von Burt, dass die publizierten Ergebnisse gefälscht waren.

Welches ist der Grund dafür, dass Datenbetrug in den Sozialwissenschaften viel seltener berichtet wird als in den Naturwissenschaften? Sind vielleicht Sozialwissenschaftler ehrlicher? Oder raffinierter? Vermutlich sind es zwei Gründe. Erstens sind die Sozialwissenschaften weniger kumulativ und replikativ orientiert. Viele, wenn nicht die meisten Studien werden nie wiederholt. Und wenn eine Replikation zu anderen Ergebnissen kommt, kann auf unterschiedliche Bedingungen verwiesen werden. Zweitens folgen sozialwissenschaftliche Untersuchungen in geringerem Masse dem hypothetisch-deduktiven Wissenschaftsideal. Viele Studien sind explorativ, entwickeln Hypothesen anhand des empirischen Materials und finden immer irgendwelche mehr oder minder berichtenswerte Befunde. «Story telling» nennt das kritisch und bissig der Soziologe und Demograph Thomas DiPrete. Kurzum, bei explorativen Arbeiten gibt es geringere oder gar keine Anreize zur Fälschung.

Wahrscheinlich leiden die Sozialwissenschaften weniger unter direkter, beabsichtigter Datenfälschung, sondern vielmehr an den Problemen fehlerhafter Datenerhebung und -auswertung (dazu auch Mayntz, 1999). Beispiele für den ignoranten und grob fahrlässigen Umgang mit statistischen Daten sind Legion, allerdings nicht nur in der sozialwissenschaftlichen, sondern auch in der biomedizinischen Forschung. Ein flapsiges Insider-Bonmot lautet vielleicht nicht ganz zu Unrecht, dass man bei zwei Dingen nicht zusehen solle: Bei der Herstellung von Würstchen und bei der Schätzung ökonometrischer Modelle.

Eine Gruppe von Ökonomen (Dewald, Thursby und Anderson, 1986) publizierte in der American Economic Review eine bemerkenswerte Studie. Es handelte sich um einen Bericht über die Replikation der Datenanalysen von Artikeln, die in einer anderen renommierten Fachzeitschrift, dem Journal of Money, Credit and

Banking erschienen waren.[5] Die Ergebnisse des Replikationsversuchs waren verheerend. Nur ein kleiner Teil der publizierten Befunde war mit den von den Autoren gelieferten Daten reproduzierbar (Tabelle 1).

Ein zweites Problem der empirischen Sozialforschung sind Fälschungen durch Dritte. Die Datenbasis empirischer Sozial- und Wirtschaftsforschung wird in der Regel von Umfrageinstituten geliefert, die wiederum mit der Erhebung der Daten Interviewer beauftragen, die in der Regel nach Akkord und nicht gerade üppig entlohnt werden. Das System setzt Anreize zur Fälschung, und jedem in der Branche ist bekannt, dass ein Teil der persönlichen Interviews ganz oder teilweise gefälscht sind, woran auch so genannte Feldkontrollen nicht viel ändern. Auch Angaben zu den so genannten Non-Response-Quoten sind kaum überprüfbar und oftmals krass nach oben verzerrt. Wenn z. B. ein Institut eine Response-Quote von 70 Prozent angibt, dann sollte man zunächst einmal fragen, wie dieser Wert errechnet wurde. Leider interessieren sich Auftraggeber, auch aus der Wissenschaft, viel zu wenig für diese Details (vgl. auch Diekmann, 2002). Weniger problematisch sieht es übrigens bei den computerunterstützten Telefoninterviews aus, die verlässlichere Daten liefern. Gelegentlich sind es aber nicht nur die Interviewer, sondern die kommerziellen Institute selbst, die bewusst oder grob fahrlässig die Fälschung von Daten ermöglichen.

Tabelle 1: Replication in Empirical Economics: The Journal of Money, Credit and Banking Project*

Empirische Arbeiten, angefragt	62	
keine Antwort	20	
Daten nicht lieferbar	20	wovon 14 verloren oder vernichtet
Daten und Programme lieferten	22	
davon:		
gut dokumentiert	1	
Replikation für Auswahl von Arbeiten	9	
(weitestgehend) replizierbar**	3	

Anmerkungen:
* JMCB ist eine renommierte und referierte Fachzeitschrift.
** Nicht replizierbar wegen z. B. Programmierfehlern, Datenfehlern, statistischen Irrtümern.
Quelle: Ergebnisse nach Dewald et al., 1986.

5 Äusserst wünschenswert und aufschlussreich wäre es, diese Studie für die gegenwärtige deutsche empirisch-sozialwissenschaftliche Forschung zu wiederholen. Es ist auch sehr gut möglich, dass sich die Situation verbessert hat. Ökonometrische Schätzungen erfolgten damals häufig mit selbst programmierter, schlecht dokumentierter und fehlerträchtiger Software. Heute wird weit mehr mit Standardsoftware gearbeitet, so dass Replikationen erleichtert werden. Siehe zum Thema Replikation auch weiter unten Teil VI.

Ist die Fälschung von Interviews überhaupt ein Problem? Eine extreme These besagt, dass Interviewfälschungen die Datenqualität nicht beeinträchtigen. Sind hundert Prozent der Interviews gefälscht, dann hat man eben eine Befragung von Interviewern und die wissen meist besser Bescheid als die Befragten selbst. Gefälschte Interviews sind gewissermassen Experteninterviews. Natürlich ist diese These unhaltbar, denn erstens sind Interviewer ein selektives Sample und zweitens können sie kaum stellvertretend für die Zielpersonen sprechen, wenn es um neue Themen und Stimmungen oder ihnen unbekannte Sachverhalte geht.[6]

5 Diagnose gefälschter Daten

Wie kommt man Datenfälschungen auf die Spur? Bleiben wir noch bei den Interviews. Bei persönlichen Interviews auf der Basis von Zufallsstichproben kann man zunächst Adressen per Zufallsverfahren ermitteln. Die Interviewer erhalten dann die Adressen und sollen die entsprechenden Zielpersonen befragen. Man spricht hier von Adressrandom. Die alternative, billigere Methode ist das Random-Route-Verfahren, bei dem der Interviewer die Zielpersonen nach bestimmten Begehungsregeln selbst ermittelt und interviewt. In der Theorie ist Random-Route gut, sofern sich Interviewer daran halten. Die Einhaltung ist aber kaum kontrollierbar. Und die Interviewer haben einen Anreiz, von den Regeln abzuweichen, wenn der Zielhaushalt schlecht erreichbar oder im Haushalt keine Person anwesend ist. Dann wird eben beim Nachbarn geläutet. Beim «Deutschen Familiensurvey» (ca. 10'000 Face-to-Face Interviews, 1988) wurde eine Teilstichprobe aus Gemeinderegistern gezogen (Adressrandom), die Befragten der zweiten Teilstichprobe wurden per Random-Route ermittelt. Dieser einmalige Methoden-Split erlaubt einen Vergleich der beim ersten Kontakt realisierten Interviews. Es sollte bei Einhaltung der Regeln zur Auswahl der Befragten eigentlich keinen grossen Unterschied geben. Tatsächlich findet man eine sehr starke Differenz. Random-Route-Interviewer realisieren beim ersten Kontakt weitaus mehr Interviews (44%) als ihre Kollegen mit den vorgegebenen Adressen der Zielpersonen aus der Gemeindestichprobe (25%, Alt, 1991).

Kriminalistischen Scharfsinn entwickelten Blohm und Wasmer (2003) bei der «Allgemeinen Bevölkerungsumfrage 2002». Es bestand die begründete Vermutung, dass es bei einem grossen Block von mehr als 400 Interviews zu Unregelmässigkeiten gekommen war. Da die persönlichen Interviews mit einem Laptop durchgeführt wurden (so genannte CAPI-Interviews), war die Interviewdauer auch für Teilmengen von Fragen exakt bestimmbar. Nun korrelierten sie die Zahl der Haushaltsmitglieder mit der Interviewdauer für diejenigen Fragen, bei denen

6 Zu Experimenten zu Verzerrungen durch gefälschte Interviews siehe Reuband, 1990; Schnell, 1991 und Schraepler und Wagner, 2003.

die Auskunftsperson über sich und die anderen Personen im Haushalt informiert hatte. Die inkriminierten Interviews unterschieden sich in diesem Punkt deutlich von korrekten Interviews. Sie korrelierten nämlich überhaupt nicht mit der Anzahl der im Haushalt lebenden Personen. Gleichgültig wie gross der Haushalt war, pendelte die Befragungsdauer ungefähr um den Wert von rund 200 Sekunden (Abbildung 5).[7]

Für die empirische Forschung ist es von erheblicher Bedeutung, wirksame Kontrolltechniken zur Aufdeckung von Fälschungen zu entwickeln. Es gibt neuere, unkonventionelle Diagnoseverfahren, die sich allerdings noch in der Erprobung befinden. Die Prinzipien dieser Verfahren und einige experimentelle Ergebnisse sollen im Folgenden kurz skizziert werden.

Statistische Daten sind in der Regel Zahlen und Ziffern. Unter gewissen Bedingungen folgen die Ziffern echter Daten Gesetzmässigkeiten, die gefälschte Daten nicht gleichermassen erfüllen.

Ein einfaches Beispiel von Hill (1998) zur Erläuterung des Prinzips. Man stelle sich folgende Aufgabe vor: Eine Person soll die Daten von 200 Münzwürfen als Abfolge von Kopf und Wappen notieren. Sie kann eine Münze werfen und echte Daten berichten; sie kann sich aber auch die Daten ausdenken. Der Experimentator behauptet nun, dass er über magische Kräfte verfüge und die Versuchsreihen unterscheiden könne.

Das ist ganz einfach. Mit sehr hoher Wahrscheinlichkeit findet man bei einem echten Zufallsexperiment von 200 Münzwürfen eine Abfolge von mindestens sechsmal Kopf oder Wappen in ununterbrochener Folge. Wer dagegen solche Daten fälscht, schreibt höchst selten sechsmal hintereinander das gleiche Symbol.

Ähnlich verhält es sich mit Ziffern, die aus natürlichen oder sozialen Prozessen resultieren. Auch hier ein Beispiel. A bietet B folgende Wette an: A setzt darauf, dass die erste Ziffer im Artikel rechts unten auf der Wirtschaftsseite der Ausgabe der Neuen Zürcher Zeitung, die am nächsten Tag erscheinen wird, im Bereich eins bis vier liegt. A verliert die Wette, wenn sich die Ziffer im Bereich fünf bis neun befindet. Bei Gewinn erhält A zehn Euro, bei Verlust zahlt A die gleiche Summe an seinen Wettpartner B. Ist das ein faires Angebot?

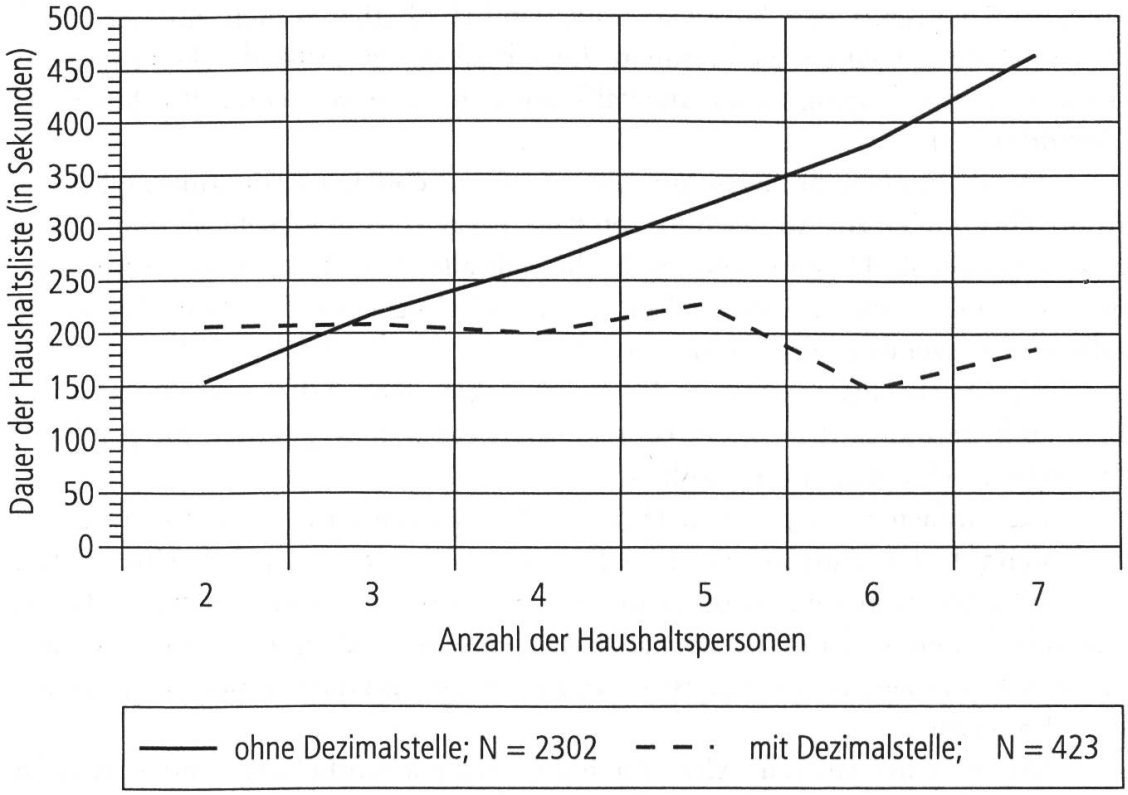
Ist es nicht. Denn die ersten Ziffern von Zahlen verschiedenster Dinge wie Hausnummern, Börsenkurse, die Fläche von Gewässern, die Flügelspannweite von Vögeln usw. sind nicht gleichverteilt. Die 1 z. B. kommt wesentlich häufiger vor als die 9. Unter bestimmten Voraussetzungen folgt die erste Ziffer vieler Daten einer logarithmischen Verteilung, der so genannten Benford-Verteilung.[8]

---

7 ZUMA hat die nicht korrekt durchgeführten Interviews umgehend aus dem Datensatz des «Allbus 2002» entfernt.

8 Die Verteilung lautet: $p_k = \log_{10}(1+1/k)$ mit $k=1,2,\ldots,9$.

Abbildung 5: Analyse von Interviews beim Allbus 2002*



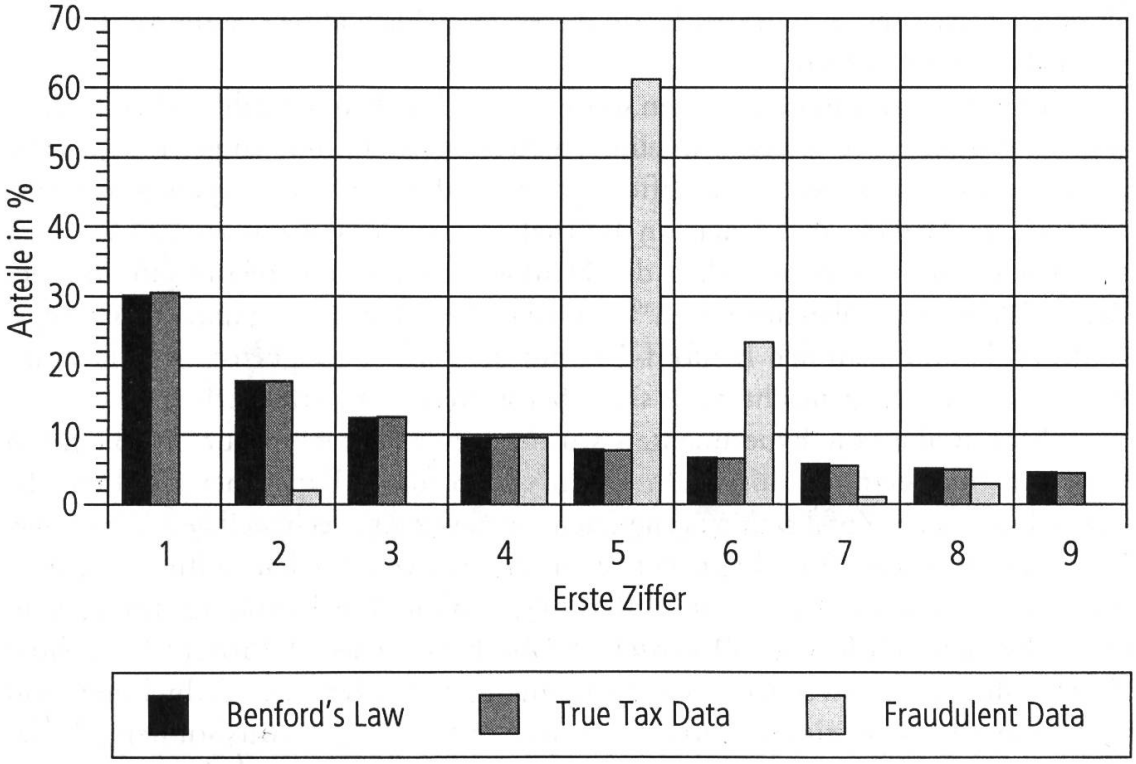
Anmerkungen:
*Anders als die korrekten Interviews wiesen die unter Fehlerverdacht stehenden Interviews eine Dezimalstelle auf. In der Analyse wurden Ausreisser nicht berücksichtigt: Dauer < 20 Sekunden, N=4; Dauer > 20 Minuten, N = 6.
Quelle: Blohm et al., 2003, 74.

«Benfords Gesetz» hat man sich z. B. zunutze gemacht, um Steuerbetrug und Bilanzfälschungen aufzudecken. Die Logik ist ganz einfach. Echte Zahlen folgen der Benford-Verteilung, gefälschte Daten weichen (möglicherweise) davon ab. Weitere Tests kann man entwickeln, indem man die zweite und dritte Ziffer heranzieht und darüber hinaus die bedingten Verteilungen (z. B. die Verteilung der zweiten Ziffer unter der Voraussetzung, dass die erste Ziffer n ist) inspiziert.

Die Untersuchung betrügerischer Manipulationen in der Buchhaltung von Unternehmen, von Steuererklärungen usw. durch Nigrini (1997 und 1999) zeigt, dass der Benford-Test zumindest in diesem Bereich Anhaltspunkte liefert, um gefälschte Daten herauszufiltern. Bei den gefälschten Daten kommen bestimmte Ziffern viel häufiger vor als bei den echten Angaben (Abbildung 6).

Die ersten Ziffern der Angaben aus 169'662 Steuererklärungen, die von M. Nigrini ausgewertet wurden, stimmen sehr gut mit Benfords Gesetz überein. Dies

Abbildung 6: Nigrinis Analyse von Steuerbetrug mittels der Benford-Verteilung



Grafik modifiziert nach Hill, 1998. Daten von Nigrini, 1997 und 1999.

war bei betrügerischen Geschäftsdaten nicht der Fall.[9] Ebenso stimmen die ersten Ziffern der Angaben von 743 Studenten, die gebeten wurden, sechsstellige Zufallszahlen aufzuschreiben, nicht mit der Benford-Verteilung überein (nach Hill, 1998).

Nach diesem ermutigenden Schritt hat Nigrini auch die Steuererklärung von Ex-US-Präsident Bill Clinton mit Benfords Gesetz inspiziert. Der Test war negativ. Es gab keine Hinweise auf auffällige Ziffernhäufigkeiten. Für seinen Nachfolger liegt noch kein Testergebnis vor.

Nun wird man natürlich einwenden, dass Fälscher sich schnell auf den Benford-Test einstellen werden. Das ist aber kaum möglich. Es ist ungeheuer schwierig, Daten konsistent zu fälschen und dann noch Benford-konform zu frisieren, insbesondere, wenn auch noch die zweite und dritte Ziffer berücksichtigt werden muss.

9 Man kann sich allerdings fragen, ob die betrügerischen Steuerdaten wirklich typisch sind. Sie wurden als Illustration der Benford-Methode immer wieder auch in grossen Publikumszeitschriften abgedruckt, z. B. im Wall Street Journal, in der New York Times und im Spiegel. Merkwürdig ist, dass die Ziffern 1, 3 und 9 überhaupt nicht auftreten. Auf diesen Punkt hat mich Peter Preisendörfer hingewiesen.

Was bei der Entdeckung gefälschter Steuererklärungen und Bilanzfälschungen hilfreich ist, könnte sich auch zur Diagnose gefälschter wissenschaftlicher Daten als nützlich erweisen. Könnte der Benford-Test tatsächlich Hinweise zur Aufdeckung von Fälschungen liefern?

In vielen empirischen Studien werden z. B. Tabellen mit zahlreichen Regressionskoeffizienten berichtet. Um eine Art Referenzstichprobe zu gewinnen, habe ich eine Auszählung der ersten Ziffer von tausend geschätzten Regressionskoeffizienten aus Artikeln des American Journal of Sociology veranlasst. Es ist auch statistisch ganz interessant, dass die Häufigkeitsverteilung relativ gut mit der Benford-Verteilung übereinstimmt.[10] Gerät eine Veröffentlichung unter Fälschungsverdacht, könnte man den Benford-Test auf die Ziffern geeigneter Statistiken anwenden. Bevor das gemacht wird, sind aber weitere Tests erforderlich.

Eine Reihe von Experimenten wurde dazu am Institut für Soziologie in Bern mit Studierenden durchgeführt. In einem dieser Experimente sollten die Versuchspersonen Zufallszahlen generieren, und zwar eine sechsstellige Ziffernfolge. Vergleicht man den Anteil von Personen, die nur verschiedene Ziffern angeben mit dem erwarteten Anteil von Ziffernfolgen ohne Wiederholung, findet man eine erhebliche Differenz (39 Prozent beobachtet versus 11 Prozent bei echten Zufallszahlen). In zwei weiteren Experimenten wurden die Teilnehmer und Teilnehmerinnen an einem Statistik-Seminar in Soziologie sowie Studierende der Ökonometrie gebeten zu tun, was sie sonst nie tun dürfen, nämlich Tabellen mit Regressionskoeffizienten passend zu einer vorgegebenen Hypothese zu erfinden. Erste Indizien deuten darauf hin, dass nicht die erste, sondern die Inspektion der zweiten, dritten und weiterer Ziffern möglicherweise Aufschluss über konstruierte Daten liefern kann. Bezüglich der Fabrikation von Zufallszahlen wird ein ähnliches Resultat zur Bedeutung der nachfolgenden Ziffern von Mosimann et al. (1995) berichtet. Bevor solche Verfahren eingesetzt werden, sind aber weitere Untersuchungen erforderlich.[11]

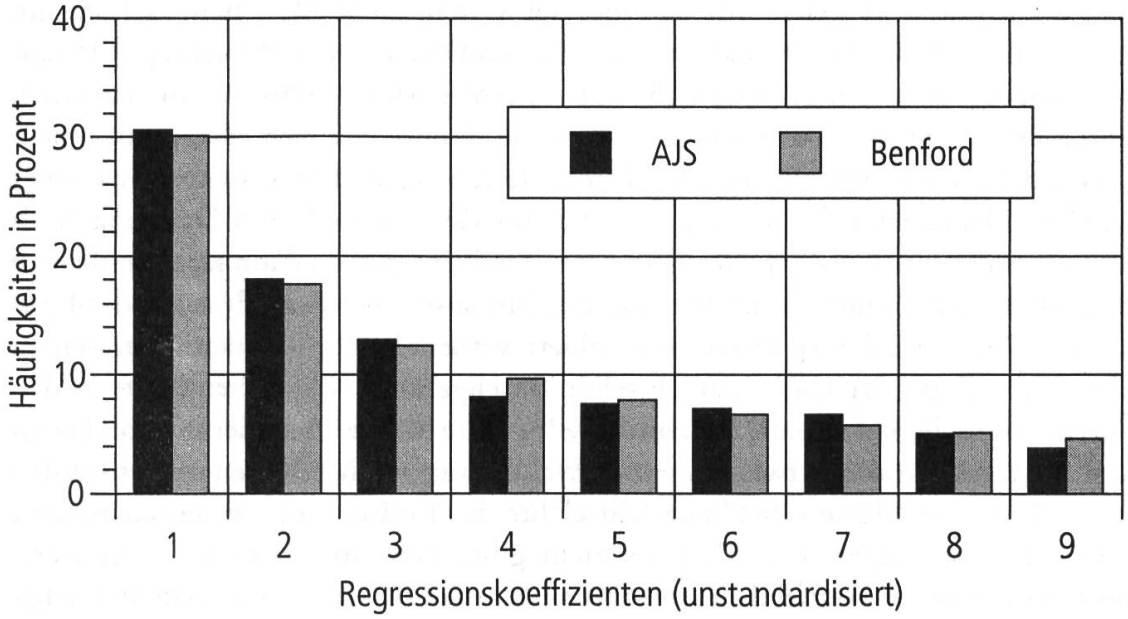
6 Was tun?

Es gibt viele Möglichkeiten, die Wahrscheinlichkeit von Wissenschaftsbetrug zu verringern. Ich möchte hier nur zwei Gesichtspunkte ansprechen, die sich vorwie-

10 Es wurden nicht nur OLS-Regressionskoeffizienten berücksichtigt, sondern z. B. auch Werte aus logistischen Regressionen. Eine zweite Stichprobe nur von OLS-Werten ergab eine signifikante Abweichung von der Benford-Verteilung, weil die Ziffer 5 etwas häufiger als erwartet vorkam. Bei den übrigen (ersten) Ziffern war die Übereinstimmung mit der Benford-Verteilung aber relativ gut. Bei der Verteilung der zweiten Ziffer zeigte sich keine signifikante Abweichung von der Benford-Verteilung. Die Ergebnisse (Diekmann, 2004b) findet man im Internet unter der folgenden Web-Adresse: http://www.socio.ethz.ch/de/diekmann/unpub.html.

11 Zu den Ergebnissen der vorliegenden Experimente siehe die Web-Adresse in der vorhergehenden Fussnote.

Abbildung 7: Verteilung der ersten Ziffer von unstandardisierten Regressionskoeffizienten



Stichprobe aus Tabellen von Artikeln, die im Zeitraum von Januar 1996 (Vol 101) bis Mai 1997 (Vol 102) im American Journal of Sociology publiziert wurden (N = 1000).

gend auf die Sozialwissenschaften beziehen, aber eventuell auch für die Naturwissenschaften von Belang sind. Dabei geht es nicht nur um Fälschungen, sondern auch um fahrlässige Irrtümer und Tricks bei der Datenerhebung und besonders der statistischen Analyse der Daten, die – wie erwähnt – in den Sozialwissenschaften, aber auch in der biomedizinischen Forschung weithin verbreitet sind.

Erstens müssen Herausgeber von Zeitschriften in viel stärkerem Masse eine Diskussions-, Kritik- und Replikationskultur entwickeln. Viele Herausgeber mögen es nicht, wenn bereits in ihrer Zeitschrift publizierte Beiträge kritisiert oder bei Replikationen massive Fehler entdeckt werden. Herausgeber lieben zudem stromlinienförmige Artikel mit einem schönen roten Faden, in dem sich die einzelnen Puzzleteile von Theorie und empirischen Fakten schnörkellos, prägnant und konsistent zusammenfügen, was das empirische Material in ehrlichen Manuskripten häufig aber gar nicht zulässt. Es ist auch bekannt, dass nicht-signifikante Befunde ungern gesehen werden. Damit entsteht eine Erwartungshaltung, der sich karriereorientierte Autoren nur schwer entziehen können. Die Neigung ist gross, Artikel zu glätten, gegenteilige Befunde wegzulassen und Effekte durch geeignete Verfahren oder grafische Darstellungen stärker hervortreten zu lassen. Das ist aber nicht nur die Schuld von Autoren, sondern auch eine Folge der Herausgeberpolitik von

Fachzeitschriften. Ehrliche Manuskripte haben eben in der Regel weniger gute Annahmechancen.[12]

Der zweite Aspekt beinhaltet einen konkreten Vorschlag. Zunächst gehe ich davon aus, was gute Zeitschriften – aber bei weitem nicht alle – heute schon tun, dass Autoren von Publikationen mit empirischen Daten eine Archivierungs-, Dokumentations- und Weitergabepflicht haben. Am besten verlangt man, natürlich unter Beachtung des Datenschutzes, dass die dokumentierten Daten auf einem Server der Fachzeitschrift für jeden Interessierten zugänglich ins Internet gestellt werden. Damit sind die Voraussetzungen für die Kontrolle des Datenmaterials und für eine Reanalyse der Daten geschaffen. Im Prinzip kann dann jeder publizierte Beitrag zumindest mittels interner Replikation, also einer Reanalyse auf der Basis der vom Autor erhobenen Daten, kontrolliert werden. Solche Replikationen werden ohne äussere Anreize aber ziemlich selten durchgeführt. Wir haben heute in den Sozialwissenschaften öffentlich finanzierte Institute für die Archivierung von Daten und Literatur, für die Entwicklung und Evaluierung von Methoden und es werden in wachsendem Masse erhebliche Mittel für die Evaluierung wissenschaftlicher Einrichtungen aufgewendet. Was es nicht gibt, dabei nicht einmal teuer wäre, aber gleichwohl einen erheblichen positiven Effekt auf die Qualität wissenschaftlicher Ergebnisse hätte, ist eine Institution, deren Aufgabe die Replikation publizierter wissenschaftlicher Arbeiten wäre. Genauer bestünde die Aufgabe darin, eine Stichprobe von z. B. einem Prozent aller in den Fachzeitschriften publizierten Arbeiten der entsprechenden Disziplin zu ziehen und die Daten zu reanalysieren. Die Ergebnisse würden in der Zeitschrift berichtet, in welcher der geprüfte Artikel erschienen ist. Allein die Tatsache, dass jede Forschungsgruppe mit einer solchen Kontrolle rechnen müsste, dürfte die Sorgfalt bei der Datenauswertung ebenso beeinflussen wie betrügerische Manipulationen erschwert würden.

Auf Kooperation angewiesene Systeme können je nach Sensitivität ein mehr oder minder grosses Ausmass an Nicht-Kooperation oder betrügerischen Handlungen verkraften. Finanzminister z. B. können mit zehn Prozent Steuerhinterziehung ganz gut leben, während zehn Prozent gefälschte Banknoten vermutlich eine Währung vernichten würden. Ebenso hat das System Wissenschaft wohl eher eine geringe Fehlertoleranz. Kein System aber kann dauerhafte Stabilität erlangen, ohne institutionelle Vorkehrungen zur Eindämmung betrügerischer Aktionen. Auf Auktionen im Internet wie z. B. bei eBay werden von anonymen Akteuren einmalige Geschäfte abgeschlossen, bei denen beide Parteien, Käufer und Verkäufer, normalerweise einen starken Anreiz zum Betrug hätten. Eine ganz einfache, aber raffinierte Institution, nämlich die Möglichkeit zur Bewertung der Verkäufer, trägt wesentlich dazu bei, dass Betrug in Internet-Auktionen sehr selten ist. Wie wir kürzlich in einer Studie nachgewiesen haben, erzielen Verkäufer mit vielen positiven Bewertungen, also hoher Reputation, auch höhere Auktionspreise

12 Siehe auch Frey (2003) zu Opportunismus fördernden Effekten des Begutachtungssystems.

(Diekmann und Wyder, 2002). Es lohnt sich also, in Reputation zu investieren. Das ist aber nur möglich, indem man sich kooperativ verhält. In einem solchen System haben auch betrügerische Händler ein Interesse kooperativ zu sein, weil nur diese Strategie für sie von Vorteil ist. Gute Institutionen zeichnen sich dadurch aus, dass sie auch bei bösen Leuten funktionieren. Das eBay-Reputationssystem ist aber leider nicht auf den Wissenschaftsmarkt übertragbar. Ein Konsument kann in der Regel sofort nach Lieferung erkennen, ob die ersteigerte Ware der Beschreibung entspricht. Der Konsument eines wissenschaftlichen Artikels kann dagegen nicht wissen, ob die Daten im Artikel gefälscht sind. Bei der Fälschung von Daten in der wissenschaftlichen Forschung handelt es sich zudem häufig um «Delikte ohne Opfer». Mangelnde direkte Betroffenheit und asymmetrische Information hat zur Folge, dass auch kritische Leserinnen und Leser im Normalfall die verborgenen Mängel nicht erkennen können. Deshalb wird eine Evaluationsinstanz benötigt, die natürlich keine neue Bürokratie oder Behörde darstellen soll. Die Kontrolle durch Replikation kann durchaus im Rahmen der wissenschaftlichen Selbstkontrolle erfolgen, ähnlich wie dies bei Evaluationen von Instituten oder Disziplinen heute praktiziert wird.

Unser gegenwärtiges Wissenschaftssystem kann man auch mit einer anderen Metapher beschreiben. Stellen sie sich einen öffentlichen Verkehrsverbund mit Bussen und Strassenbahnen vor, für die man zwar einen Fahrpreis entrichten muss, aber niemand Schwarzfahrer kontrolliert. Weil Verkehrsbetriebe wissen, was passieren würde, bezahlen sie Kontrolleure, die etwa ein Prozent der Fahrgäste überprüfen. Diese einfache Logik könnte auch im Wissenschaftsbereich durchaus von Nutzen sein.

7 Literatur

Alt, Christian (1991), Stichprobe und Repräsentativität, in: H. Bertram (Hrsg.), *Die Familie in Westdeutschland. Stabilität und Wandel familialer Lebensformen,* DJI: Familien-Survey 1, Opladen: Leske & Budrich, S. 497–531.

Babbage, Charles ([1830] 1969), *Reflections on the Decline of Science in England, and on Some of its Causes,* Westmead: Gregg Intern. Publishers.

Beasley, Malcolm R.; Supriyo Datta, Herwig Kogelnik, Herbert Kroemer und Don Monroe (2002), *Report of the Investigation Committee on the Possibility of Scientific Misconduct in the Work of Hendrik Schön and Coauthors,* Murray Hill, New Jersey: Bell Labs.

Beck-Bornholt, Hans-Peter und Hans-Herrmann Dubben (1997), *Der Hund, der Eier legt. Erkennen von Fehlinformationen durch Querdenken,* Reinbek: Rowohlt.

Blohm, Michael; Janet Harkness, Sabine Klein und Evi Scholz (2003), *Konzeption und Durchführung der «Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften» (ALLBUS) 2002. ZUMA-Methodenbericht 2003/12,* Mannheim: Zentrum für Umfragen, Methoden und Analysen.

Broad, William und Nicolas Wade (1984), *Betrug und Fälschung in der Wissenschaft,* Basel: Birkhäuser.

Dewald, William G.; Jerry G. Thursby und Richard G. Anderson (1986), Replication in Empirical Economics: The Journal of Money, Credit and Banking Project, *American Economic Review,* 76, 587–603.

DFG-Kommission «Selbstkontrolle in der Wissenschaft» (1998), *Vorschläge zur Sicherung guter wissenschaftlicher Praxis,* Bonn: Deutsche Forschungsgemeinschaft (DFG).

Diekmann, Andreas (2002), *Diagnose von Fehlerquellen und methodische Qualität in der sozialwissenschaftlichen Forschung,* Wien: Österreichische Akademie der Wissenschaften. Institut für Technikfolgenabschätzung (ITA).

Diekmann, Andreas (2003), *Methoden der empirischen Sozialforschung, 10. Aufl.,* Reinbek: Rowohlt.

Diekmann, Andreas (2004a), *«Betrug und Täuschung in der Wissenschaft». Datenfälschung, Diagnoseverfahren, Konsequenzen. Jahrbuch 2003 der Deutschen Akademie der Naturforscher Leopoldina,* Halle.

Diekmann, Andreas (2004b), *Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data,* Zürich: Mimeo.

Diekmann, Andreas und David Wyder (2002), Vertrauen und Reputation bei Internet-Auktionen, *Kölner Zeitschrift für Soziologie und Sozialpsychologie,* 54, 674–693.

Duerr, Hans-Peter, Hrsg.(1987), *Authentizität und Betrug in der Ethnologie,* Frankfurt a.M.: Suhrkamp.

DiTrocchio, Federico (1994), *Der grosse Schwindel. Betrug und Fälschung in der Wissenschaft,* Frankfurt a.M. und New York: Campus.

Feyerabend, Paul F. (2001), *Wider den Methodenzwang,* Frankfurt a. M.: Suhrkamp.

Finetti, Marco und Armin Himmelrath (1999), *Der Sündenfall. Betrug und Fälschung in der deutschen Wissenschaft,* Stuttgart: Raabe.

Fox, James Alan und Paul E. Tracy (1986), *Randomized Response. A Method for Sensitive Surveys,* London: Sage.

Fröhlich, Gerhard (2001), Betrug und Täuschung in den Sozial- und Kulturwissenschaften, in: T. Hug, Hrsg., *Wie kommt die Wissenschaft zu ihrem Wissen?,* Hohengehren: Schneider Verlag.

Frey, Bruno S. (2003), Publishing as Prostitution. Choosing Between One's Own Ideas and Academic Success, *Public Choice,* 116, 205–223.

Gould, Stephen J. (1988), *Der falsch vermessene Mensch,* Frankfurt a. M.: Suhrkamp.

Hill, Theodore P. (1998), The First Digit Phenomenon, in: *American Scientist,* 86, 358–363.

Hughson, Roy V. und Philip M. Kohn (1980), Ethics, *Chemical Engineering,* September 22, 132–147.

Hurt, Richard D. und Channing R. Robertson (1998), Prying Open the Door to the Tobacco Industry's Secrets About Nicotine, *Journal of the American Medical Association,* 280, 1173–1181.

List, John A.; Charles D. Bailey, Patricia J. Euzent und Thomas L. Martin (2001), Academic Economists Behaving Badly? A Survey on Three Areas of Unethical Behavior, *Economic Inquiry, 39,* 162–170.

Mosimann, James.E.; Claire V. Wiseman und Ruth E. Edelman (1995), Data Fabrication: Can People Generate Random Digits?, in: *Accountability in Research,* 4, 31–55.

Mayntz, Renate (1999), *Betrug in der Wissenschaft. Randerscheinung oder wachsendes Problem? Working paper,* Köln: Max-Planck-Institut für Gesellschaftswissenschaften.

Merton, Robert K. (1957), *Social Theory and Social Structure,* Ill: Glencoe.

Nigrini, Mark J. (1999), I've Got Your Number. How a Mathematical Phenomenon Can Help CPAs Uncover Fraud and Other Irregularities, *Journal of Accountancy,* 79–83.

Nigrini, Mark J. (2000), *Digital Analysis Using Benford's Law,* Vancouver: Global Audit Publications.

Rapp, Ulf et al. (2000), *Task Force F. H. Abschlussbericht,* Würzburg: Institut für Medizinische Strahlenkunde und Zellforschung der Universität Würzburg.

Reuband, Karl-Heinz (1990), Interviews, die keine sind. «Erfolge» und «Misserfolge» beim Fälschen von Interviews, *Kölner Zeitschrift für Soziologie und Sozialpsychologie,* 42, 706–733.

Ronzheimer, M. (1998), Eisberg in Sicht? Symposium über Fälschung in der wissenschaft gibt Entwarnung. *BerliNews,* Dezember 1998 (www.berlinews.de/archiv/242.shtml).

Schnell, Rainer (1991), Der Einfluss gefälschter Interviews auf Survey-Ergebnisse, *Zeitschrift für Soziologie,* 20, 25–35.

Schraepler, Joerg-Peter und Gert G. Wagner (2003), *Identification and Characteristics of Faked Interviews in Surveys. An Analysis by Means of Genuine Fakes in the Rraw Data of SOEP. Unveröff. Manuskript,* Bochum und Berlin: Ruhr-Universität Bochum und Deutsches Institut für Wirtschaftsforschung Berlin.

Stegemann-Boehl, S. (1994), *Fehlverhalten von Forschern. Eine Untersuchung am Beispiel der biomedizinischen Forschung im Rechtsvergleich USA-Deutschland.*

Stewart, Walter W. und Ned Feder (1987), The Integrity of the Scientific Literature, *Nature,* 325, 207–214.

Trute, Hans-Heinrich (2003), *Vortrag des Sprechers des Ombudsmanns der DFG vor der Mitgliederversammlung der DFG in Würzburg am 2. Juli 2003.*

Westfall, Richard S. (1973), Newton and the Fudge Factor, *Science,* 179, 751–758.

Warner, S. L. (1965), Randomized Response. A Survey Technique for Eliminating Evasive Answer Bias, *Journal of the American Statistical Association,* 60, 63–69.