

Das Problem des Betrugs in der Parapsychologie¹

CHRIS A. ROE²

Zusammenfassung – In Psychologielehrbüchern wird dazu geneigt, eine Assoziation der Parapsychologie zur Möglichkeit des Betrugs durch Versuchsleiter herzustellen, indem typischerweise zwei markante Fälle mit Beteiligung von S.G. Soal und Walter Levy angeführt werden. Damit wird impliziert, dass Betrug in der Parapsychologie häufiger vorkommt als in anderen Bereichen der Sozialwissenschaften; dies soll als Ursache für überzufällige Trefferquoten bei PSI-Experimenten in Frage kommen. In diesem Artikel unterziehe ich diese beiden und weitere Fälle, bei denen betrügerische Aktivitäten im Spiel gewesen waren, einer Prüfung und vergleiche sie mit jüngsten Betrugsfällen in anderen Wissenschaftsbereichen, insbesondere der Psychologie. Ich komme zu dem Schluss, dass in beiden Fällen unbestreitbare Beweise für Betrug vorliegen, doch scheinen sie typisch für die Forschung insgesamt zu sein, anstatt etwas Spezifisches über das Themengebiet Parapsychologie auszusagen. Tatsächlich ist die transparente und öffentliche Art und Weise, in der mit ihnen umgegangen wurde, eigentlich recht ungewöhnlich; ähnliche Vorfälle aus anderen Forschungsbereichen werden in der Regel diskret behandelt, damit der Ruf der Gastinstitution nicht beeinträchtigt wird. Anonyme Umfragedaten weisen auf fragwürdige Forschungspraktiken wie Mehrfachanalysen und selektive Berichterstattung hin, die weitaus häufiger vorkommen, als man bisher annahm. Abschließend stelle ich fest, dass die Faktoren, die die Wahrscheinlichkeit von Betrug erhöhen, wie z.B. materieller Gewinn, wenn man mit dem Betrug durchkommt, oder die Wahrscheinlichkeit der Entdeckung und deren Folgen, in der Parapsychologie nicht zu finden sind.

-
- 1 Dies ist eine Übersetzung des – leicht veränderten – Artikels “The problem of fraud in parapsychology”, der im englischen Original in der Ausgabe 1/2016 (S. 8–17) der Zeitschrift *Mindfield* erschienen ist. Die Redaktion bedankt sich bei den Herausgebern von *Mindfield* und bei Chris Roe für die Druckgenehmigung der deutschen Übersetzung.
 - 2 Chris A. Roe ist Professor für Psychologie an der University of Northampton und Direktor des dort ansässigen Centre for Psychology & Social Sciences. Roe ist Perrott-Warrick Senior Researcher (Trinity College, Cambridge) und Gastprofessor an der Bucks New University. Er ist Präsident der *Society for Psychical Research*, ehemaliger Vorsitzender der *British Psychological Society Transpersonal Psychology Section*, ehemaliger Präsident der *Parapsychological Association* und internationaler Partner der *Parapsychology Foundation* für England. Seine Forschungsinteressen liegen im Verständnis der Natur anomaler Erfahrungen und umfassen Forschungen zur Phänomenologie paranormaler Erfahrungen, insbesondere im Hinblick auf das Wohlbefinden, die Psychologie des paranormalen Glaubens und der Täuschung sowie experimentelle Ansätze zur Untersuchung außersinnlicher Wahrnehmung und Psychokinese, insbesondere im Zusammenhang mit relevanten psychologischen Faktoren.

Schlüsselbegriffe: Betrug durch den Versuchsleiter – fragwürdige Forschungspraktiken – Replikationsproblem

The Problem of Fraud in Parapsychology

Abstract – Mainstream psychology textbooks tend to associate parapsychology with the possibility of experimenter fraud, typically citing two high profile cases involving S. G. Soal and Walter Levy. The implication is that fraud is more prevalent in parapsychology than other areas of the social sciences and may account for above chance scoring in psi experiments. In this article I review these two cases and others that have been identified as involving fraudulent activity and compare them with recent cases of fraud in other areas of science, particularly psychology. I conclude that both cases provide incontrovertible evidence of cheating, but they seem very typical for the sector rather than signalling something distinct about the subject area. Indeed, the transparent and public manner in which they were managed is actually quite unusual; similar incidents from other research areas tend to be dealt with discretely so that the host institution's reputation is not tarnished. Anonymous survey data are presented that suggest questionable research practices such as multiple analyses and selective reporting are much more common than previously realized. In conclusion I note that the factors that increase the likelihood of fraud, such as material rewards for getting away with cheating, likelihood of detection and consequences if detected, are not found in parapsychology.

Keywords: experimenter fraud – questionable research practices – replication problems

Oh, what a tangled web we weave, when first we practise to deceive

Walter Scott, *Marmion* (1808)

Einleitung

In einem kürzlich erschienenen Artikel (Roe, 2016) habe ich untersucht, wie Parapsychologie in akademischen Mainstream-Lehrbüchern dargestellt wird. Denn dies ist eines der wichtigsten Mittel, mit denen die nächste Generation von Akademikern über die Grenzen der legitimen wissenschaftlichen Forschung aufgeklärt wird. In Anlehnung an Kuhn (1962) habe ich argumentiert, dass es der wissenschaftlichen Bildung weniger darum geht, den Studenten zu befähigen, ein unabhängiger Denker und innovativer Forscher zu werden, als vielmehr um das Vertrautmachen mit einer Reihe von vereinbarten Praktiken und Denkweisen und deren Verinnerlichung, die es ermöglichen, die technische Arbeit der Forschung effizienter zu gestalten. Dieser Prozess der Sozialisation informiert den Wissenschaftler über die Fragen, die er vernünftigerweise über die Natur stellen kann, und über die Techniken, mit denen er versu-

chen kann, sie zu beantworten. Anstatt Verzerrungen zu beseitigen und eine wirklich objektive Wissenschaft zu erzeugen, wirkt ein solches Dogma (wie Kuhn es, bewusst an die Bilder der religiösen Indoktrination erinnernd, benennt) auf eine Standardisierung dieser Vorurteile hin, so dass eine klare Unterscheidung getroffen werden kann zwischen dem Legitimen und dem Illegitimen, dem Sinn und dem Unsinn. Es war daher enttäuschend (wenngleich nicht ganz unerwartet) festzustellen, dass die Parapsychologie, falls sie nicht vollständig unbeachtet bleibt, in solchen Lehrbüchern dadurch gekennzeichnet wird, dass sie nicht den anerkannten Standards der wissenschaftlichen Praxis entspreche. Dies insbesondere, weil sie übermäßig unter Schwierigkeiten leide, die zentralen Phänomene unter strengen experimentellen Bedingungen zu replizieren, und weil sie ernsthafte Probleme mit Betrug durch die Versuchsleiter habe, weswegen alle verbleibenden positiven Ergebnisse auf solches Fehlverhalten zurückgeführt werden könnten.

Zu diesem letzten Punkt zitiert Gross (2010, 85) beispielsweise Colman (1987), der die Geschichte der Parapsychologie als „verunstaltet durch zahlreiche Betrugsfälle unter Beteiligung einiger der ‚angesehensten Wissenschaftler‘“ beschreibt. Aber solche vernichtenden Bewertungen sind nicht auf eingefleischte Skeptiker beschränkt. Douglas Stokes (2015) behauptet, dass die Beweislage der parapsychologischen Forschung „dem Muster entspricht, das zu erwarten wäre, wenn eine kleine Minderheit von Psi-Forschern betrügerisch agiert hätte“ (S. 42), und für James Kennedy (2014) sieht die Situation noch schlimmer aus: „Fehlverhalten von Versuchsleitern ist in der Parapsychologie mehrfach aufgetreten und stellt eine ständige Bedrohung dar. Sie beeinträchtigt die wissenschaftliche Akzeptanz des Feldes und behindert den Fortschritt, indem sie Ressourcen in falsche Hypothesen umleitet“ (S. 9). In diesem Artikel möchte ich einen genaueren Blick darauf werfen, wie verbreitet Betrug in der Parapsychologie tatsächlich war und wie es sich mit anderen wissenschaftlichen Disziplinen verhält, um beurteilen zu können, ob die oben genannten Behauptungen substantiell sind.

Betrugsfälle in der Parapsychologie

Um seine Behauptung zu untermauern, verweist Kennedy unheilvoll, aber stumpfsinnig auf 17 Fälle von Betrug in der Parapsychologie,³ von denen er 15 Rhines Aufsatz „Security Versus Deception in Parapsychology“ (1974a) entnommen hat. In diesem Artikel verweist Rhine in der Tat auf „ein Dutzend Fälle, um das in den 1940er und 1950er Jahren herrschende Problem der Unzuverlässigkeit der Versuchsleiter ziemlich typisch zu veranschaulichen“ (d. h. im Rahmen der Popularisierung seiner Monographie *Extrasensory Perception*, die die an der Duke Univer-

3 Interessanterweise erwähnt er nicht die vier im gleichen Artikel beschriebenen Fälle von Kritikern (S. 102–104), deren Verhalten als Beispiele von Datenmissmanagement angesehen werden könnte.

sity entwickelten Forschungsmethoden beschrieb, mit denen Replikationsversuche ohne Spezialausrüstung oder umfangreiche Ausbildung in der Forschungsmethodik ermöglicht werden sollten). Es scheint, als ob das Ziel der Monographie erfolgreich war. Denn diese Fälle klingen wie Berichte über Experimente, die das *Journal of Parapsychology* von Personen erhalten hat, die weder Akademiker („sieben hatten keinen Dokortitel“) noch Teil der parapsychologischen Community waren („mehrere waren Personen mit offensichtlichen Fähigkeiten, die sich aber an Orten [teilweise im Ausland] befanden, wo parapsychologische Forschung extrem schwer zu bewältigen, aber nicht annähernd so schwer zu fälschen war“). Es gibt keine Hinweise darauf, dass eine dieser Personen weiterhin Experimente durchführte oder dass eine ihrer Arbeiten veröffentlicht wurde. Es erscheint deshalb merkwürdig, dass Kennedy sie so darstellt, als wären sie typisch oder repräsentativ für die parapsychologische Community insgesamt. Rhine verweist (S. 105) auf vier weitere Fälle, an denen Personen beteiligt waren, die „alle besser für die Psi-Forschung qualifiziert waren als ... das ‚Dutzend‘. Sie alle kannten die Regeln und Normen, die im Laufe der Jahre entwickelt wurden“, und diese sind von größerer Bedeutung. Leider werden alle Fälle allgemein beschrieben, und es fehlen Informationen, die es ermöglichen könnten, diejenigen zu identifizieren, die der „Versuchsleiterunzuverlässigkeit“ verdächtigt werden. Ihre Nichteinhaltung allgemein anerkannter Sicherheitsstandards wird beschrieben, aber es wird nicht ganz klar, ob es tatsächlich Fehlverhalten darstellt. Beispielsweise beinhaltet der erste Fall einen Vergleich der Leistung von Teilnehmern an Psi- und Nicht-Psi-Aufgaben, bei denen zwei Experimentatoren für verschiedene Tests verantwortlich waren. Anstatt sicherzustellen, dass die Tests unabhängig und „verblindet“ hinsichtlich der Ergebnisse anderer Tests ausgewertet wurden, tauschten die Experimentatoren tatsächlich Informationen aus, wenn die Teilnehmer besonders gut abschnitten. Dies könnte sicherlich zu einer auf der Erwartungshaltung beruhenden Verzerrung (Bias) bei der Bewertung der Leistung führen, die zwar jede Korrelation der Ergebnisse zwischen den Aufgaben überhöhen könnte, aber verblasst im Vergleich zu den später in diesem Artikel diskutierten Arten von Betrug. Die Beispiele 2 und 3 liefern stärkere Indizien für Versuchsleiterbetrug, bei denen Effekte nur dann auftraten, wenn ein (bestimmter) Experimentator unkontrollierten Zugriff auf Rohdatensätze hatte. Wichtig ist dabei, dass Kollegen Verdacht bezüglich der Arbeit schöpften und nichts davon publiziert wurde, als die Forscher sie zurückzogen. Das letzte Beispiel, das Rhine in sein Buch aufgenommen hat, ist eine Studie, die die von Pratt und Woodruff beschriebene Screened Touch Matching-Methode anwendet. Sie wurde nämlich von Mark Hansel kritisiert, weil sie Betrug durch Ausnutzung unzureichender Matching-Möglichkeiten zulässt, und deshalb aufgenommen, weil „Tricks eine denkbare Möglichkeit waren“ und nicht, weil jemand *in flagranti* erwischt worden wäre. Zusammengefasst stellen diese 16 Fälle (akzeptierend, dass Kennedy nur die ersten 15 übernimmt) kein zwingendes Argument für Kennedys Behauptung dar, dass Fehlverhalten von Experimentatoren in der Parapsychologie üblich ist. Er äußert auch Bedenken gegenüber

anderen Forschern in Rhines Labor, aber diese sind wenig mehr als eine Unterstellung auf der Basis von Klatsch, und ich werde hier keine Zeit dafür opfern.

Am wichtigsten sind zwei besondere Fälle, auf die Kennedy (2014) und Stokes (2015) verweisen und die allgemein als kalkulierter und systematischer Betrug durch den Experimentator akzeptiert werden. Der erste betrifft Walter J. Levy Jr., einen Absolventen der Medizinischen Fakultät, der zum Institut für Parapsychologie von J. B. Rhine kam und so produktiv und hoch angesehen war, dass man ihn zum Direktor des Instituts ernannte; es wurde allgemein erwartet, dass er Rhine nach dessen bevorstehender Pensionierung ablösen würde (Stokes, 2015). Levy interessierte sich besonders für Psi bei Tieren (animal psi) und hatte geniale Experimente entwickelt, die die Psi-Fähigkeiten an Wüstenrennmäusen, Ratten und Kükenembryonen so testeten, dass eine erfolgreiche Psi-Leistung die Grundbedürfnisse der Tiere befriedigen konnte (Vermeidung von Schmerzen, Steigerung der Lusterlebnisse, Aufrechterhaltung der optimalen Körpertemperatur usw.). Eine der großen Stärken von Levys Experimenten war ihre Automatisierung – einmal eingerichtet, konnte das Gerät unabhängig laufen, das Verhalten von Zufallsgeneratoren (Random Event Generators, REGs) überwachen, die die Tiere beeinflussen mussten, um das gewünschte Ergebnis zu erzielen, und eine physikalische Aufzeichnung der Ergebnisse zur Analyse erstellen. Drei von Levys Forscherkollegen (Kennedy, Jim Davis und Jerry Levin) wurden schließlich misstrauisch aufgrund der Menge der Zeit, die Levy in der Nähe der Geräte zu verbringen schien, während die Experimente im Gange waren. Sie verkabelten heimlich den Computer, so dass er eine zweite Aufzeichnung des REG-Outputs produzierte, und zu ihrer Bestürzung entdeckten sie, dass der REG-Output völlig zufällig war, während Levys offizielle Aufzeichnung eine Abweichung in die vorhergesagte Richtung zeigte (das heißt in diesem Fall, dass die neurologischen Genusszentren der Ratten stärker stimuliert wurden, als man zufällig erwarten würde). Rhine wurden die Beweise vorgelegt und er konfrontierte Levy damit. Dieser gab zu, Daten in der konfirmatorischen Studie gefälscht zu haben, doch er bestand darauf, dass seine ursprüngliche Forschung solide war und die Fälschung erst begonnen hätte, als der ursprüngliche Effekt nicht mehr repliziert werden konnte (Kennedy, 1975). Diese Forschung war noch nicht veröffentlicht worden (da Rhine eine Replikation für notwendig hielt, um im Fall einer Bestätigung die Ergebnisse als beweiskräftig ansehen zu können). Levy verteidigte seine anderen, veröffentlichten Arbeiten und wies darauf hin, dass sie sowohl am Institut für Parapsychologie als auch anderswo unabhängig repliziert worden seien. Dennoch riet Rhine (1974b) in der nächsten Ausgabe des *Journal of Parapsychology* wise: „Obwohl seine einzige bekannte Verletzung nur eine seiner vielen experimentellen Zugänge betrifft, wirft sie unvermeidlich einen begründeten Zweifel an seiner gesamten Arbeit, auf die er in den fünf Jahren seiner parapsychologischen Tätigkeit alleine und gemeinsam mit anderen durchgeführt hat“ (S. 220). Rhine war zuversichtlich, dass unabhängige Replikationen langfristig zwischen soliden und unsoliden Befunden unterscheiden würden. Er schrieb sofort an alle, von denen er

glaubte, dass sie auf Levys Arbeit bei der Planung und Durchführung ihrer eigenen Studien zurückgreifen wollten, um sie über die Fragwürdigkeit der Daten zu informieren, und stellte sicher, dass dieser Betrugsfall schnell und öffentlich behandelt wurde. Fünfzehn Monate später gab Rhine (1975) ein Update, das andere verdächtige Praktiken von Levy aufdeckte. Dies führte ihn zu dem eher widerwilligen Schluss, dass, trotz doppelter Verblindung und eines Multi-Experimentier-Designs in einigen Fällen, in keiner Studie die Möglichkeit des Betrugs vollständig ausgeschlossen werden konnte, so dass keine andere Wahl blieb, als alles abzuschreiben.

Der zweite allgemein akzeptierte Fall betrifft Samuel G. Soal, einen Mathematikdozenten am Queen Mary College (Teil der University of London), der einer der Hauptvertreter von ESP-Tests mit Forced-choice-Paradigma in Großbritannien war; aber das konsequente Scheitern dieser Tests machten ihn zu einem der stärksten Kritiker von Rhine (Beloff, 1993). Das heißt, bis er (auf Anraten von Whately Carington) seine Daten reanalysiert hatte, um nach Verschiebungseffekten (displacement effects) zu suchen, also nach Fällen, in denen die Wahl des Teilnehmers nicht mit dem Target-Symbol übereinstimmte, sondern mit dem vorhergehenden oder nachfolgenden Symbol. Zwei sehr erfolgreiche Teilnehmer, Basil Shackleton und Gloria Stewart wurden zu weiteren Tests eingeladen, was zu einer stetigen Folge von signifikanten Ergebnissen führte. Soal war von Beruf Mathematiker, hat aber seinen D. Sc. (doctor of science) für parapsychologische Forschung erhalten. Obwohl sein Ruf zu seinen Lebzeiten intakt blieb, kam Verdacht auf, aber dieser wurde schnell durch Androhung von Klagen niedergeschlagen (Beloff, 1993: 147). In einem ausführlichen Bericht haben West und Markwick (2018) beschrieben, dass einer der Ankläger Mrs Gretl Albert war. Sie war einer der erfolgreichen *agents*⁴ von Basil Shackleton. Sie vertraute ihrer Freundin Mollie Goldney (Soals Co-Versuchsleiterin) an, dass sie nach dem Test gesehen hatte, wie er Zahlen auf einem Ergebnisprotokoll veränderte (die Ergebnisse wurden in Form der Zahlen 1 bis 5 und nicht als Symbole aufgezeichnet). Goldney bat um Einsicht in die Datenblätter, konnte aber keine Anzeichen einer Manipulation feststellen. Die Art der Änderungen wurde nicht offenkundig durch eine einfache Prüfung mit einer Lupe; sie erforderten eine statistische Analyse. Scott und Haskell (1974) spekulierten, dass es, wenn Soal die Einsen auf der Liste (target 1) in Vieren oder Fünfen (target 4 oder target 5) umwandelte, so dass die *targets* mit den Teilnehmeraussagen übereinstimmten, einen Überschuss an Treffern bei Versuchen geben sollte, bei denen das *target* anscheinend 4 und 5 war, und ein Defizit von *target 1s*,

4 Bei den *agents* handelte es sich um Personen, die als „Sender“ ein Bild betrachten und es mental an Shackleton übermitteln sollten, der in einem anderen Raum saß. Der Versuchsleiter selbst saß mit dem *agent* in einem Raum und zeigte ihm anhand einer Liste von Nummern (1 bis 5, entsprechend den fünf im Versuch verwendeten Bildern; diese Liste bildet gleichzeitig das Protokoll, in das die Ergebnisse – „hit“ oder „miss“ – eingetragen wurden), welches Bild der *agent* jeweils betrachten und „senden“ sollte (Anm. des Übers.).

bei denen die Teilnehmer 4 oder 5 geraten hatten.⁵ Beide Effekte wurden so dramatisch bestätigt, dass sie nahelegen, dass bei einer Wahl von 4 oder 5 die meisten *target 1s* verändert wurden, um falsche Treffer zu erzeugen. Es gab jedoch insgesamt keine zu geringe Anzahl von *target 1s* oder zu hohe Anzahl von *target 4s* und *5s*, was Barrington (persönliche Kommunikation, zitiert in West & Markwick, 2018) als Hinweis darauf interpretierte, dass die Protokollblätter schon mit Blick auf Manipulationen vorbereitet worden waren, indem sie mit zu vielen 1s und zu wenigen 4s oder 5s versehen waren. Scott und Haskell erkennen zwar an, dass sie damit nicht die überzufälligen Trefferraten in den meisten der Sitzungen erklären können, die Soal mit Shackleton durchführte, hielten es aber für wahrscheinlicher, dass andere Fälschungsmethoden angewendet wurden, als dass die Ergebnisse eine Mischung aus Echtem und Betrügerischem darstellten. Beweise für solche anderen Methoden kamen von Markwick (1978), der herausfand, dass Soals vorbereitete Listen von Zufallszahlen (angeblich aus veröffentlichten Quellen) Wiederholungen von Sequenzen mit bis zu 25 Ziffern enthielten, manchmal in umgekehrter Reihenfolge. Diese wiederholten Sequenzen enthielten eine gelegentliche zusätzliche Ziffer, und 75% von ihnen ergaben Treffer, was darauf hindeuten könnte, dass Platzhalter-Ziffern (z. B. „1“) eingegeben wurden mit der Absicht, sie später an die Wahl des „Empfängers“ anzupassen. Abgesehen vom allgemeinen Verdacht bezüglich der von Soal gemeldeten Erfolgsraten (laut Stokes [2015] hatte J. B. Rhine lange Zeit vermutet, dass seine Forschung betrügerisch war) sind andere Aspekte seines Verhaltens verwirrend. West und Markwick (2018: 162) kommen zu dem Schluss, dass „unsere Aufdeckungen seiner weiteren Täuschungen die Annahmen einiger Skeptiker stärken könnte, dass alle scheinbar überzeugenden Behauptungen des Paranormalen betrügerisch sein müssen. Das ist nicht unsere Meinung. Soal war außergewöhnlich in seiner Geheimniskrämerei, seinem Widerstand gegen Einflussnahme von außen und seiner mangelnden Bereitschaft, seine Probanden von anderen Experimentatoren testen zu lassen.“

Fälle von Betrug in anderen Disziplinen

Wie steht nun dieses Bild von Betrug in der Parapsychologie im Vergleich zu anderen Wissenschaften? Rhine selbst hatte den Eindruck, dass die Parapsychologie hinterherhinkt: „Die meisten anderen Wissenschaftszweige sind bereits so weit gereift, dass das Problem des Betrugs durch Versuchsleiter keine große Sorge bereitet. Dies liegt zum Teil daran, dass vorsätzlicher Betrug zu schnell erkannt und aufgedeckt würde“ (1974a: 112). Dieses Vertrauen in die wechselseitige Kontrolle in der gängigen wissenschaftlichen Praxis ist durchaus

5 Für das Verständnis dieses Vorgehens ist es wichtig zu beachten, dass im englischsprachigen Raum im Unterschied zu Deutschland die „1“ nur mit einem senkrechten Strich dargestellt wird. Dieser einfache senkrechte Strich kann leicht nachträglich in eine „4“ oder „5“ umgewandelt werden, worauf Scott und Haskell ihre Betrugshypothese aufbauen (Anm. d. Übers.).

typisch. Braud und Wade (1982: 11) geben eine Aussage von Philip Handler, dem Präsidenten der National Academy of Sciences, vor dem US House Committee on Science and Technology wieder, der das Problem des wissenschaftlichen Betrugs als von der Presse als „stark übertrieben“ bezeichnete, und selbst wenn es sich ereignet, „geschieht dies innerhalb eines Systems, das in einem effektiven demokratischen und selbstkorrigierenden Modus funktioniert“. Aber dieses System hat sich als unvollkommen erwiesen und die Sammlung von Fällen wissenschaftlichen Fehlverhaltens ist stetig angewachsen, so dass die National Science Foundation zwischen drei Formen unterscheidet: Herstellung, Fälschung und Plagiat (vgl. Gross, 2016: 694). Plagiate sind definiert als die Aneignung von Ideen, Verfahren, Ergebnissen oder Worten einer anderen Person, ohne angemessene Angabe der Quelle bzw. des Urhebers. Unter Fälschung versteht man die Manipulation von Forschungsmaterialien, -geräten oder -prozessen oder die Veränderung von Daten oder Ergebnissen. Mit Herstellung ist das Erfinden von Daten oder Ergebnissen gemeint. Um diese drei Formen zu veranschaulichen und ihr allgegenwärtiges Auftreten zu belegen, möchte ich kurz auf wichtige Persönlichkeiten der Wissenschaftsgeschichte eingehen, die heute allgemein als Betrüger gelten. Diese Auflistung ist alles andere als vollständig (für weitere Beispiele siehe Braud & Wade, 1982; Grant, 2007; Judson, 2004; Kohn, 1986).

Der früheste bekannte Fall von Plagiat ist vielleicht der von Claudius Ptolemäus, dessen *Almagest* astronomische Beobachtungen vorstellt, die die Grundlage für ein mathematisches Modell zur Beschreibung der Bewegungen von Himmelskörpern um die Erde bilden, und das enorm einflussreich war, bis es durch das von Kopernikus und anderen vorgeschlagene heliozentrische Modell ersetzt wurde. Ptolemäus behauptete, diese Beobachtungen selbst in Alexandria/Ägypten gemacht zu haben. Spätere Rückrechnungen aus den aktuellen Positionen der Planeten deuten jedoch darauf hin, dass viele dieser Beobachtungen selbst für damalige Standards sehr schlecht waren und viel genauer denjenigen entsprachen, die Ptolemäus' Vorgänger Hipparchus etwa 278 Jahre zuvor aus Rhodos hätte beobachten können (Newton, 1977). Ptolemäus' Beobachtungen beinhalten seltsame Lücken – von den 1.025 Sternen, die er dokumentierte, stammt keiner aus dem Fünf-Grad-Abschnitt, der von Alexandria aus sichtbar, aber von Rhodos aus unsichtbar ist. Wie Grant (2007: 20) zusammenfasst: „Anstatt hinauszugehen und Beobachtungen zu machen, scheint es, dass Ptolemäus seine Zeit in der Bibliothek in Alexandria verbrachte, um viele von Hipparchus' Ergebnissen abzukupfern und sie als seine eigenen zu beanspruchen“.

Isaac Newton ist vielleicht die bedeutendste Person, die der Fälschung beschuldigt wird. Er war eine aufbrausende Persönlichkeit, die regelmäßig mit Zeitgenossen gestritten hat. Sein größter Gegner war wahrscheinlich Leibniz, dessen Naturphilosophie im Widerspruch zu Newtons Gravitationstheorie und Bewegungsgesetzen stand, wie sie in seinen *Philosophiae Naturalis Principia Mathematica* beschrieben sind. Leibniz' Einfluss im kontinentalen Europa bedeutete, dass Newtons Theorie dort auf eine lauwarmer Reaktion traf. In den folgenden Aus-

gaben wurde Newtons Position durch die Anpassung der Daten verstärkt, einschließlich seiner Berechnungen der Schallgeschwindigkeit, der Präzession der Äquinoktien und der Messung der Gezeiten, so dass sie genau mit seiner Theorie übereinstimmten (Braud & Wade, 1982). Die von Newton berichteten Messungen werden mit sechs signifikanten Stellen angegeben, eine Präzision, die selbst noch heute fast unmöglich ist (Westfall, 1973).

Dem Augustinermönch Gregor Mendel wird zugeschrieben, dass er die Vererbung von Merkmalen über Generationen von Erbsenpflanzen in Proportionen beobachtet hat, die auf eine Art diskreter Übertragung hindeuteten. Damit legte er den Grundstein für die Genforschung. Die Proportionen, die Mendel berichtete, entsprachen jedoch so genau der theoretischen Erwartung, dass sie R. A. Fisher, dem bedeutenden Statistiker, der unter anderem für die Analyse der Varianz (ANOVA) verantwortlich war, verdächtig vorkamen. Fisher untersuchte die Methoden und Daten von Mendel eingehend und fand heraus, dass die Daten zu gut waren, um wahr zu sein, und schloss daraus auf eher freundliche Weise, dass die Assistenten von Mendel die Zahlen möglicherweise den Erwartungen entsprechend angepasst hätten. Andere haben unbewusste Fehler oder Datenselektion (vgl. Judson, 2004: 52–58), oder, weniger freundlich, Fälschung angenommen (Anonymous, 1972).

Inzwischen ist man der Ansicht, dass Galileo Galilei die Ergebnisse vieler seiner Experimente gefälscht hat, was gewissermaßen eine Ironie darstellt angesichts des exemplarischen Empirikers, der experimentelle Beobachtungen gegenüber ästhetischen oder theoretischen Anliegen den Vorzug gab. Erzählungen über seine Versuche, die Wirkung der Schwerkraft zu testen, indem er Objekte vom Schiefen Turm von Pisa fallen ließ, gelten als apokryph, wurden aber nicht von Galileo selbst, sondern später erfunden (Grant, 2007: 20). Andere Experimente, von denen er behauptete, er habe sie „fast hundert Mal“ mit konsistenten Ergebnissen repliziert, hätten den berichteten Grad an Homogenität mit den damals verfügbaren Materialien nicht erreichen können – wie Zeitgenossen wie Père Marin Mersenne erfahren mussten, die seine Ergebnisse nicht replizieren konnten (Braud & Wade, 1982: 26).

Betrug in der Psychologie

Ein prominenter und weitreichender Betrugsfall aus der Psychologie betrifft Diederik Stapel. Er war Professor für Sozialpsychologie und Dekan der School of Social and Behavioral Sciences an der Universität Tilburg. Er hatte eine erfolgreiche Karriere hinter sich, erhielt 2,2 Millionen Euro an Zuschüssen von der Netherlands Organisation for Scientific Research (Enserink, 2012b) und produzierte 124 Artikel in Fachzeitschriften, die insgesamt 1.756 Mal zitiert wurden (Budd, 2013). Ein solcher Artikel, der in der Fachzeitschrift *Science* veröffentlicht wurde, berichtete von einem Experiment am Bahnhof in Utrecht, das zeigte, dass eine vermüllte

Umgebung tendenziell rassistische Neigungen bei Einzelpersonen hervorruft (Stapel & Lindenberg, 2011). Das Reinigungspersonal war kurz vor den Sommerferien im Streik gewesen. Dies ermöglichte, die Reaktionen von Personen, die den Bahnhof während des Streiks aufsuchten, als der Bahnsteig ungereinigt blieb, mit denen zu vergleichen, die dort waren, nachdem die Reinigungskräfte zurückgekehrt waren. Er hatte angenommen, dass die geäußerten Einstellungen in der ehemaligen, weniger ordentlichen Umgebung stereotyper wären („Brasilianer sind sexy, Briten sind höflich, New Yorker sind aufdringlich“). Auf ausgeklügelte Weise entwarf er ein Instrument zur Erfassung des Verhaltens, indem er die überwiegend weißen Befragten einlud, sich auf einen von sechs in einer Reihe angeordneten Sitzen zu setzen; an einem Ende der Reihe saß eine schwarze Person, die anscheinend ebenfalls bereits an der Studie teilnahm. Stapel berichtete, dass die Teilnehmer in der „unordentlichen“ Versuchsbedingung tatsächlich mehr stereotype Ansichten zum Ausdruck brachten und interessanterweise weiter weg von der bereits sitzenden schwarzen Person saßen, wenn ihre Umgebung unordentlicher war, was indirekt auf erhöhten Rassismus hindeutet. Das Problem bei diesem eleganten Ergebnis ist, dass Stapel von der Existenz dieses Phänomens als wahren Merkmal der realen Welt so überzeugt war, dass er es nicht für notwendig hielt, die Daten tatsächlich zu sammeln! Stattdessen erfand er zu Hause Daten, um damit ein saubereres Ergebnis zu erzielen. Er behauptete, dies sei eine Folge seiner frühen Erfahrungen mit den Herausgebern von Zeitschriften, die seine realen experimentellen Daten zu kompliziert fanden, mit Zusammenhängen, die zu chaotisch waren, und die ihn oft baten, Elemente wegzulassen und die Dinge zu vereinfachen, bevor sie veröffentlicht würden – es sei einfacher, sicherzustellen, dass die Daten ordentlich und konsistent sind, indem man sie erfindet (Bhattacharjee, 2013).

Schnell war er in eine ausgeklügelte Farce verwickelt, in der er mit Forschungsassistenten oder Doktoranden bei der Entwicklung von Forschungsmaterialien wie Fragebögen und maßgeschneidertem Equipment zusammenarbeitete, aber dann (erstaunlicherweise) darauf bestand, die Studien allein durchzuführen bzw. die Materialien zu den „Kontakten“ von ihm in Schulen und Hochschulen zu bringen, um sie dort aufzubereiten. Dies machte es ihm möglich, das Experiment zu Hause zu simulieren, um einen plausiblen Benchmark-Score zu erhalten und dann Datensätze um diesen Wert herum zu generieren, die eine eindeutige, aber dennoch glaubwürdige Bestätigung seiner Hypothese ergeben würden. Danach würde er die Datensätze bzw. Analyseergebnisse den Kollegen zurückbringen und mit ihnen bei der Ausarbeitung zusammenarbeiten. Stapel mag mit seinen Erfindungen so lange durchgekommen sein, weil er sicherstellte, dass seine Ergebnisse den allgemeinen Erwartungen entsprachen. „Ich habe keine seltsamen Dinge getan, ich habe nie gesagt, dass wir ein Experiment machen sollen, um zu zeigen, dass die Erde flach ist [...]. Ich habe immer gecheckt [...], dass das Experiment vernünftig ist, dass es an bestehende Forschung anschließt, dass es nur diesen zusätzlichen Schritt darstellt, auf den alle warten“ (Bhattacharjee, 2013).

Letztendlich wurde sein Betrug aufgedeckt, als seine Mitarbeiter nach möglichen internen Analysen (z. B. Geschlechtsunterschieden) fragten, an die Stapel nicht gedacht und die er daher nicht erfunden hatte, oder sie baten um die Rohdaten für explorative Analysen, aber man sagte ihnen, dass diese aus Mangel an Lagerplatz vernichtet worden seien. Verdachtsmomente ergaben sich auch durch Stapels nahezu fehlerfreie Signifikanz. Ein Kollege war erstaunt, wie großartig die Daten unabhängig vom Experiment aussahen: „Ich weiß nicht, ob ich jemals gesehen habe, dass eine Studie gescheitert ist, was sehr ungewöhnlich ist [...]. Selbst die besten Leute haben nach meiner Erfahrung Studien, die konstant scheitern. Normalerweise funktioniert die Hälfte nicht“ (Bhattacharjee, 2013). Eine Untersuchung durch die Universitäten, die ihn beschäftigt hatten, ergab einen Abschlussbericht, nach dem Stapel in mindestens 55 seiner Arbeiten sowie in 10 Doktorarbeiten seiner Studenten Betrug begangen hatte (Levelt, 2012). Bis 2014 wurden 58 seiner veröffentlichten Arbeiten zurückgezogen.⁶

Während Stapel als „vielleicht der größte Betrüger in der akademischen Wissenschaft“ bezeichnet werden kann (Bhattacharjee, 2013), ist er unter Psychologen keineswegs einmalig: Mangel an Platz hindert mich daran, gut dokumentierte Fälle zu diskutieren, unter anderem um Sir Cyril Burt, der Datensätze (und Co-Autoren) erfunden hat, um die Rolle der genetischen Vererbung für Persönlichkeit und Intelligenz zu bekräftigen (Tucker, 1997), Marc Hauser, der Daten fabrizierte und seine Doktoranden unter Druck setzte, seine bevorzugten Schlussfolgerungen zu ziehen (Wade, 2010), Dirk Smeesters, der zwei vielbeachtete Aufsätze zurückgezogen hatte, als man deren Daten als zu gut, um wahr zu sein, befunden hatte (Enserink, 2012a), und Karen Ruggiero, die zugab, fünf Experimente erfunden zu haben, die in zwei Artikeln veröffentlicht wurden, sowie die Doktorarbeit, die in einem dritten erschienen ist (Price, 2010).

Diese detaillierten Exposés können den Eindruck erwecken, dass es sich um sehr ungewöhnliche Verirrungen handelt, die von Personen begangen werden, deren Verhalten sicher als pathologisch angesehen werden kann (vielleicht verursacht durch Stress oder Überarbeitung). Aber wir müssen misstrauisch bleiben, da dies einfach eine Verteidigungsstrategie darstellen könnte, die es uns ermöglicht, den Übeltäter von normalen Forschern zu distanzieren und sein Verhalten von der normalen Praxis zu unterscheiden. Braud and Wade (1982: 20) beschreiben den Beginn von Betrug als oft aus kleineren Schritten bestehend; „diejenigen, die wissenschaftliche Daten fälschen, beginnen und haben wahrscheinlich Erfolg mit dem viel kleineren Delikt, bestehende Ergebnisse zu verbessern. Geringfügige und scheinbar triviale Fälle von Datenmanipulation – wie z. B. das ‚Schönen‘ von Ergebnissen, die nur ein wenig knackiger oder klarer erscheinen, als sie wirklich sind, oder die Auswahl der ‚besten‘ Daten für die Veröffentlichung

6 Damit ist er immer noch nur Vierter in der „Rangliste“ der Retraction Watch, die vom japanischen Anästhesisten Yoshitaka Fujii mit 183 Zurücknahmen angeführt wird (<http://retractionwatch.com/the-retraction-watch-leaderboard/>).

und das Ignorieren derjenigen, die nicht in das Konzept passen – sind in der Wissenschaft wahrscheinlich alles andere als ungewöhnlich. Aber es ist nur ein gradueller Unterschied zwischen dem ‚Zubereiten‘ der Daten („cooking“ the data) und dem Sich-aus-den-Fingern-Saugen eines ganzen Experiments“. Dies spiegelt sich in Stapels eigener Darstellung wider. Zu Beginn seiner Biographie *Faking Science: A True Story of Academic Fraud* schreibt er (S. iii): „Es ging mir gut, aber dann wurde ich ungeduldig, überambitioniert, leichtsinnig. Ich wollte immer schneller und besser und höher und intelligenter werden, die ganze Zeit. Ich dachte, es würde helfen, wenn ich nur diese eine winzige kleine Abkürzung nehmen würde, aber dann befand ich mich immer öfter auf der völlig falschen Spur, und am Ende war ich überhaupt nicht mehr auf der Straße. Ich verließ die Straße, wo ich hätte geradeaus fahren sollen, und machte meinen eigenen, spektakulären, zerstörerischen, fatalen Unfall.“

Neuere Erkenntnisse aus der Psychologie deuten darauf hin, dass diese eher kleinen Überschreitungen relativ häufig sind. John, Loewenstein und Prelec (2012) befragten über 2.000 Psychologen zu ihrer Mitwirkung an fragwürdigen Forschungspraktiken. Beunruhigenderweise fanden sie heraus, dass einige Praktiken weit häufiger als erwartet auftraten, wie z. B. multiple Analysen mit selektiver Berichterstattung (78%) und optionales Stopping (also Abbruch des Experiments – Anm. d. Übers.), sobald das Signifanzniveau erreicht war (36%), aber auch Praktiken, die näher am Ende der Skala des bewussten Betrugs lagen, wie z. B. der Ausschluss von Daten, nachdem ihre Auswirkungen auf die Analyse bekannt waren (62%), und die Fälschung von Daten (9%). Über den Bereich der Psychologie hinausgehend, präsentierte Fanelli (2009) eine Meta-Analyse von Umfragedaten über wissenschaftliches Fehlverhalten, die einen gepoolten gewichteten Mittelwert⁷ von 1,97% von Forschern ergab, die bekannten, schon einmal Forschungsdaten hergestellt oder gefälscht zu haben; auf die Frage, ob sie persönliche Kenntnisse über einen Kollegen hätten, der Forschungsdaten verändert, fabriziert oder gefälscht hat, stieg diese Zahl jedoch auf 14,12%; und wenn Fehlverhalten umfassender definiert wurde, zum Beispiel als „experimentelle Mängel, Mängel in den Berichten, Falschdarstellung von Daten, Datenfälschung“, ist die Zahl weiter auf 46,24% gestiegen (S. 7). Als die American Association for the Advancement of Science eine Stichprobe ihrer Mitglieder befragte, fanden sie heraus, dass 27% der Befragten glaubten, in den letzten 10 Jahren auf fingierte, gefälschte oder plagiierte Forschung gestoßen zu sein, wobei durchschnittlich 2,5 Beispiele angegeben wurden (Titus, Wells & Rhoades, 2008). Die Häufigkeit ist bei einer Beschränkung auf die biomedizinischen Wissenschaften bezeichnenderweise noch höher (Ranstam et al., 2000; Roberts & St John, 2014; Wells, 2008).

Ein Indikator für Betrug ist die Quote der Aufsätze, die nachträglich aus den Zeitschriften zurückgezogen werden. Steen, Casadevall und Fang (2013) fanden 2.047 zurückgezogene Arti-

7 Mittelwert aus den Mittelwerten verschiedener beteiligter Stichproben, die je nach Stichprobenumfang gewichtet werden (Anm. d. Übers.).

kel, die in PubMed indexiert waren, wobei die Zahl in den letzten zehn Jahren sowohl aufgrund von Betrug als auch von „Fehlern“ (einschließlich Plagiaten) stark angestiegen ist. Dies konnte nicht nur auf die höhere Anzahl an Publikationen insgesamt zurückgeführt werden, sondern es handelte sich auch um einen prozentualen Anstieg (wobei allerdings Gross [2016] vor einer Überinterpretation dieser Zahlen warnt, da erhöhte Rückzugsraten auch die größere Aufmerksamkeit für Fehlverhalten oder die größere Bereitschaft, Redakteure und Verleger auf solche Vorfälle aufmerksam zu machen, widerspiegeln könnten). Fang, Steen und Casadevall (2012) fassten nach, indem sie sekundäre Quellen zur Ermittlung der Widerrufsgründe hinzuzogen, wenn in der Widerrufserklärung keiner erwähnt worden war (was oft der Fall war). Sie stellten fest, dass 21,3% der Widerrufe auf Fehler und 67,4% auf Fehlverhalten zurückzuführen waren, einschließlich Betrugs oder vermuteten Betrugs (43,4%), doppelter Veröffentlichung (14,2%) und Plagiaten (9,8%). Es ist nicht ungewöhnlich bei Zurückziehungen, dass sie mehrdeutig und mit wenig Getöse angekündigt werden, so dass sie möglicherweise unbemerkt bleiben. Zurückgezogene Papiere werden auch nach ihren Widerrufen weiterhin zitiert (vgl. Budd, 2013), darunter von 18% der Autoren von zurückgezogenen Papieren selbst, wobei weniger als 5% angeben, dass die Papiere zurückgezogen wurden (Gross, 2016). In jüngster Zeit wurden Webseiten wie Retraction Watch (<http://retractionwatch.com>) eingerichtet mit dem Ziel, unzuverlässige Forschung bekannt zu machen. Ihr Einfluss ist schwer abzuschätzen, aber sie können es erleichtern, zweifelhafte Praktiken herauszustellen. Der größte Teil wissenschaftlichen Fehlverhaltens wird von Laborkollegen des Übeltäters entdeckt, einschließlich ihrer Vorgesetzten und Studenten, die hinreichend eng zusammenarbeiten, um Seltsamkeiten im Verhalten oder in den Daten feststellen zu können (Shamoo & Resnik 2003). Aber die Folgen für solche „Whistleblower“ können schwerwiegend sein. Lubalin, Ardini und Matheson (1995) stellten fest, dass 47 von 68 der von ihnen befragten Beschwerdeführer mindestens eine negative Konsequenz zu erleiden hatten, wie z. B. dass sie unter Druck gesetzt wurden, ihre Behauptung zurückzuziehen, Verfemung durch Kollegen oder eine Verringerung der Unterstützung ihrer Forschung, und 56% glaubten, dass „Whistleblowing“ den Beschwerdeführer stigmatisiert. Zurückziehungswebseiten bieten die Möglichkeit, Bedenken anonym vorzubringen und so den Beschwerdeführer zu schützen, und können sicherstellen, dass Behauptungen über bestimmte Forscher umfassender diskutiert und behandelt werden und nicht unter den Teppich gekehrt werden. Zur Veranschaulichung: Im Zeitraum Juli bis Dezember 2012 veröffentlichte die Blog-Site sciencefraud.org 274 anonyme E-Mails von Biowissenschaftlern, die ein Fehlverhalten in publizierten Studien geltend machen. Im Januar 2013 zwangen rechtliche Drohungen zur Schließung dieser Webseite, doch wurden weitere 233 anonyme E-Mails eingereicht, die nicht veröffentlicht werden konnten. Das Schicksal der in diesen E-Mails genannten Aufsätze wurde von Brookes (2014) analysiert. Er fand keine anfänglichen Unterschiede in den Merkmalen der „öffentlichen“ und „privaten“ Fälle, aber die Anzahl der Widerrufe war bei den „öffentlichen“

Fällen um 650% und die Rate der veröffentlichten Korrekturen oder anderen Errata um 770% höher. Insgesamt wurden bei 23% der öffentlich diskutierten Aufsätze irgendwelche Korrekturmaßnahmen ergriffen, bei den nicht veröffentlichten Papieren waren es jedoch nur 3,1%.

Schlussfolgerungen

Abschließend möchte ich sagen, dass die Parapsychologie zwar zwei sehr gut begründete Betrugsfälle hat, diese jedoch sehr normal für diesen Bereich zu sein scheinen, anstatt etwas Besonderes über den Themenbereich zu anzuzeigen (Levy erinnert sehr an Betrüger wie den Cornell-Krebsforscher Mark Spector, der Laborleiter werden sollte, und Soals Fall scheint dem von Cyril Burt sehr ähnlich zu sein). Angesichts der von mir hier angesprochenen jüngsten Enthüllungen über Fehlverhalten in einer Reihe von Disziplinen scheint die Betrugshäufigkeit in der Parapsychologie im Vergleich zum Bereich der Forschung generell nicht sehr hoch zu sein (obwohl ein Pro-Kopf-Vergleich möglicherweise etwas anderes nahelegen könnte). Es erscheint mir daher völlig ungerechtfertigt, wenn Kommentatoren unterstellen, dass die Parapsychologie ein besonderes Problem mit dem Fehlverhalten von Versuchsleitern hat.

Tatsächlich gibt es Grund zu der Annahme, dass die Parapsychologie sogar viel *weniger* anfällig für Betrug ist als andere Forschungsbereiche. Obwohl es noch keine systematischen empirischen Studien über die Merkmale von Tätern wissenschaftlichen Fehlverhaltens gibt, hat Gross (2016) den typischen betrügerischen Wissenschaftler (und sein Umfeld) als „einen klugen und ehrgeizigen jungen Mann beschrieben, der in einer Eliteeinrichtung in einem sich schnell verändernden und hart umkämpften Zweig der modernen Biologie oder Medizin arbeitet, dessen Ergebnisse wichtige theoretische, klinische oder finanzielle Auswirkungen haben. Er wurde von einer hochrangigen und angesehenen Persönlichkeit betreut und unterstützt, die oft Koautor vieler seiner Arbeiten ist, aber vielleicht nicht eng an der Forschung beteiligt war.“ Und Braud und Wade (1982: 86) kommen aufgrund ihrer umfangreichen Primärforschung zu dem Schluss, dass „die Kriminalitätsrate in der Wissenschaft von drei Hauptfaktoren beeinflusst wird: die Belohnungen, die wahrgenommenen Chancen, erwischt zu werden, und die persönlichen ethischen Standards des Wissenschaftlers“. (Ich möchte noch die zu erwartenden Folgen des Erwischtwerdens hinzufügen, da es eine Reihe von Fällen gibt, in denen diejenigen, die Fehlverhalten zugegeben haben, ihre Institution auf diskrete Weise verlassen durften, um ihren Ruf nicht zu schädigen oder in langwierige und kostspielige Gerichts- und behördliche Verfahren verwickelt zu werden.) Mit anderen Worten, Betrug wird dort häufiger vorkommen, wo er wahrscheinlich lukrativ ist, wo die eigene Forschung relativ unbeaufsichtigt (oder sogar unbemerkt) verlaufen kann und wo mit den eigenen Handlungen, selbst wenn sie entdeckt werden, eher still und nicht öffentlich umgegangen wird. Ich würde behaupten, dass keine dieser Bedingungen für die Parapsychologie zutrifft, die stark unterfinanziert ist, bei positiven

Befunden einer sehr starken Kontrolle unterliegt und bei der Betrugsaufdeckung eine schnelle und öffentliche Enthüllung nachweisen kann.

Das ist kein Grund zur Selbstzufriedenheit, und es könnte noch viel mehr getan werden, um zu begründen, warum parapsychologische Daten ernst genommen werden sollten. Betrug ist ohne Zugriff auf Rohdaten nur sehr schwer zu erkennen. Kennedy (2014) stellt fest, dass die Betrüger oft zögern, Rohdaten für eine erneute Analyse zur Verfügung zu stellen (was sicherlich bei Soal der Fall war), und dies umso mehr, seit die stochastischen Eigenschaften von „echten“ Daten besser verstanden werden (z. B. Benfords Gesetz, das die Häufigkeitsverteilung der vorangestellten Ziffern in vielen natürlich vorkommenden Datensätzen beschreibt – siehe Benford, 1938; Miller, 2016). Der Fall scheint mir zwingend für die Einrichtung einer offenen Datenhaltung zu sprechen, damit die Beweise, auf die wissenschaftliche Behauptungen aufgebaut werden, von jedem Interessierten überprüft werden können; dies scheint eine natürliche Ergänzung der Design-Registrierung und Datenbanken für unveröffentlichte Forschung zu sein, die bereits verfügbar sind oder sich in der Entwicklung befinden.

(aus dem Englischen von Gerhard Mayer)

Literatur

- Anonymous (1972). Peas on Earth. *Horticultural Science*, 72, 438. Reproduced at http://rowdysites.msudenver.edu/~churchcy/BIO3600/bio3600_misc/Peas%20on%20Earth.htm
- Beloff, J. (1993). *Parapsychology: A concise history*. London: Athlone Press.
- Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4), 551–572.
- Bhattacharjeethe, Y. (2013, 28. April). Mind of a Con Man. *New York Times*. Retrieved from <https://www.nytimes.com/2013/04/28/magazine/diederik-stapels-audacious-academic-fraud.html>
- Broad, W., & Wade, N. (1982). *Betrayers of the truth: Fraud and deceit in the halls of science*. London: Century Publishing.
- Brookes, P.S. (2014). Internet publicity of data problems in the bioscience literature correlates with enhanced corrective action. *PeerJ*, 2, e313. doi 10.7717/peerj.313
- Budd, J.M. (2013). The Stapel Case: An object lesson in research integrity and its lapses. *Synesis: A Journal of Science, Technology, Ethics, and Policy*, G47–G53.
- Enserink, M. (2012a). *Rotterdam marketing psychologist resigns after university investigates his data*. June 25, 2012. Retrieved from <http://www.sciencemag.org/news/2012/06/rotterdam-marketing-psychologist-resigns-after-university-investigates-his-data>

- Enserink, M. (2012b). *Diederik Stapel under investigation by Dutch prosecutors. Oct. 2, 2012*. Retrieved from <http://www.sciencemag.org/news/2012/10/diederik-stapel-under-investigation-dutch-prosecutors>
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, 4(5), e5738. doi:10.1371/journal.pone.0005738
- Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *PNAS*, 109, 17028–17033.
- Goodstein, D. (2000). In defence of Robert Andrews Millikan. *Engineering & Science*, 4, 30–38. Retrieved from <http://calteches.library.caltech.edu/4014/1/Millikan.pdf>
- Grant, J. (2007). *Corrupted science: Fraud, ideology and politics in science*. Wisley, Surrey: AAPPL Artists' and Photographers' Press Ltd.
- Gross, C. (2016). Scientific misconduct. *Annual Review of Psychology*, 67, 693–711.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532.
- Judson, H. F. (2004). *The great betrayal: Fraud in science*. London: Harcourt, Inc.
- Kennedy, J. E. (1975). *Summary of rat implementation work*. Unpublished document, retrieved from <http://jeksite.org/psi/frat.pdf>.
- Kennedy, J. E. (2014). *Experimenter misconduct in parapsychology: Analysis manipulation and fraud*. Published at <http://jeksite.org/psi/misconduct.pdf> and <http://jeksite.org/psi/misconduct.htm>
- Kohn, A. (1986). *False prophets: Fraud and error in science and medicine*. Oxford: Basil Blackwell.
- Kuhn, T. (1962). *Structure of scientific revolutions*. Chicago: University of Chicago Press.
- Lubalin, J. S., Ardini, M. E., & Matheson, J. L. (1995). *Consequences of whistleblowing for the whistleblower in misconduct in science cases*. Washington, DC: Res. Triangle Inst. Available at <https://ori.hhs.gov/sites/default/files/final.pdf>
- Markwick, B. (1978). The Soal-Goldney experiments with Basil Shackleton: New evidence of data manipulation. *Proceedings of the Society for Psychical Research*, 56, 250–277.
- Miller, S. J. (2016) (Ed.). *Benford's Law: Theory and applications*. Princeton: Princeton University Press.
- Newton, R. R. (1977). *The crime of Claudius Ptolemy*. Baltimore: Johns Hopkins University Press.
- Price, M. (2010). Sins against science. *Monitor on Psychology*, July/August 2010. Retrieved from <http://www.apa.org/monitor/2010/07-08/misconduct.aspx>
- Ranstam, J., Buyse, M., George, S. L., Evans, S., Geller, N. L., et al. (2000). Fraud in medical research: An international survey of biostatisticians. *Controlled Clinical Trials*, 21, 415–27.
- Rhine, J. B. (1974a). Security versus deception in parapsychology. *Journal of Parapsychology*, 38, 99–121.
- Rhine, J. B. (1974b). A new case of experimenter unreliability. *Journal of Parapsychology*, 38, 215–225.
- Rhine, J. B. (1975). A second report on a case of experimenter fraud. *Journal of Parapsychology*, 38, 306–325.

- Roberts, D. L., & St John, F. A. (2014). Estimating the prevalence of researcher misconduct: A study of UK academics within biological sciences. *PeerJ*, 2, e562. doi 10.7717/peerj.562
- Roe, C. A. (2016). Oh, what a tangled web we weave, when first we practise to deceive: The problem of fraud in parapsychology. *Mindfield*, 8(1), 8–17.
- Shamoo, A. E., & Resnik, D. B. (2003). *Responsible conduct of research*. New York: Oxford Univ. Press.
- Stapel, D. (2012). *Faking science: A true story of academic fraud*. Translation 2014 by Nicholas J.L. Brown, available under a creative commons licence from <http://nick.brown.free.fr/stapel/FakingScience-20161115.pdf>
- Stapel, D. A., & Lindenberg, S. (2011). Coping with chaos: How disordered contexts promote stereotyping and discrimination. *Science*, 332(6026), 251–253.
- Steen, R. G., Casadevall, A., & Fang, F. C. (2013). Why has the number of scientific retractions increased? *PLoS ONE*, 8, e68397. doi:10.1371/journal.pone.0068397
- Stokes, D. (2015). The case against psi. In E. Cardena, J. Palmer, & D. Marcusson-Clavertz (Eds.), *Parapsychology: A handbook for the 21st century* (pp. 42–48). Jefferson, NC: McFarland & Co.
- Titus, S. L., Wells, J. A., & Rhoades, L. J. (2008). Repairing research integrity. *Nature*, 19(453), 980–82.
- Tucker, W. H. (1997). Re-reconsidering Burt: Beyond a reasonable doubt. *Journal of the History of the Behavioral Sciences*, 33(2), 145–162.
- Wade, N. (2010, August 27). Harvard researcher may have fabricated data. *New York Times*. Retrieved from <http://www.nytimes.com/2010/08/28/science/28harvard.html>
- Wells, J. A. (2008). *Final report: Observing and reporting suspected misconduct in biomedical research*. Washington, DC: Gallup Org. Retrieved from http://ori.hhs.gov/sites/default/files/gallup_finalreport.pdf
- West, D. J., & Markwick, B. (2018). Dr Soal: A psychic enigma. *Proceedings of the Society for Psychical Research*, 60(224), 1–172.
- Westfall, R. S. (1973). Newton and the fudge factor. *Science*, 179(4075), 751–758.