

## Empirische Arbeiten als Zufallsbefunde

Martin Erik Horn

bbw Hochschule Berlin-Brandenburg, Leibnizstr. 11 – 13, 10625 Berlin  
mail@martin-erik-horn.de

### Kurzfassung

Untersuchungen zeigen, dass bis zu 90 % aller empirischen Arbeiten im Bereich der Biomedizin fehlerbehaftet und eventuelle Schlussfolgerungen dieser Studien wissenschaftlich wertlos sind. Die Ergebnisse dieser Arbeiten stellen somit größtenteils lediglich Zufallsbefunde dar.

Es steht zu befürchten, dass die Fehler, die im Bereich der empirischen biomedizinischen Forschung beschrieben werden, in ähnlicher Art und Weise auch im Bereich der empirischen physikdidaktischen Forschung auftreten. Und solange physikdidaktische Forschungsarbeiten nicht standardmäßig reproduziert werden, kann die Hypothese, dass empirische physikdidaktische Arbeiten in ähnlicher Größenordnung fehlerbehaftet und wissenschaftlich wertlos sind, nicht zurückgewiesen werden.

### 1. Ein Scheitern der Empirie

Was kann es Schlimmeres für einen Wissenschaftler oder eine Wissenschaftlerin geben, als sein oder ihr Lebenswert zerstört zu sehen – zerstört nicht von Neuerem, das das Ältere präziser und genauer fasst, sondern zerstört von der Einsicht, dass das bisher Geleistete schlichtweg wertlos und falsch ist?

Diese Erfahrung musste einer der renommiertesten Biomediziner und Medizinstatistiker Deutschlands, Prof. Beck-Bornholdt machen, der nach 35-jähriger Lehr- und Forschungstätigkeit rückblickend schreibt: „Obwohl meine Forschungsergebnisse meist in angesehenen internationalen Fachzeitschriften erschienen, weiß ich heute, dass leider fast alles, was ich so mühsam erforscht habe, falsch ist“ [1].

Wollen wir als empirisch arbeitende und empirisch forschende Physikdidaktikerinnen und Physikdidaktiker riskieren, dass wir am Ende unseres Berufslebens ähnliche Schlussfolgerung zu ziehen haben?

Es ist zu befürchten, dass genau dies passieren könnte. Diese bange Vorahnung ergriff mich zum ersten Mal vor knapp zehn Jahren, als ein Kollege von Prof. Beck-Bornholdt in einem Interview [2] feststellte: „90 % aller Studien sind falsch.“ Er bezieht sich damit hauptsächlich auf empirische Studien seines Forschungsgebietes – der biomedizinischen Grundlagenforschung.

Doch das methodische Vorgehen bei der Erhebung und Auswertung statistischer Daten ist in der biomedizinischen Grundlagenforschung nicht anders als in der physikdidaktischen Forschung. Die methodischen Fehler, die dort beschrieben werden, dürften in ähnlicher Art und Weise auch im Bereich der empirischen physikdidaktischen Forschung auftreten.

### 2. Ein Scheitern der Physikdidaktik

Wir begehen einen enormen wissenschaftlichen Selbstbetrug, wenn wir die Problematik eines Scheiterns der Empirie nicht diskutieren. Seit zehn Jahren warte ich darauf, dass auf Tagungen zur Mathematik- [3] oder zur Physikdidaktik eine solche Diskussion geführt wird. Doch leider war mein Warten bisher vergeblich.

Das Scheitern der Empirie wird damit auch zu einem Scheitern der Physik- und Mathematikdidaktiken.

Das Versagen von uns als physikdidaktisch wirkender wissenschaftlicher Community ist enorm. Es hat sogar den Anschein, dass wissenschaftlich interessierte Laien durch populärwissenschaftliche Zeitungs- und Zeitschriftenbeiträge [4], [5], [6] sich bei weitem intensiver mit dieser Thematik auseinandersetzen und weit tiefergründiger über die Problematik



**90 %  
ALLER STUDIEN  
SIND FALSCH.**

Abb.1: Hans-Hermann Dubben in [2].

empirischer Studien nachdenken und streiten als wir, vermeintliche Fachleute auf dem Gebiet empirischer Forschung, dies tun. Selbst die Werbung (siehe Abb. 2) greift diesen Sachverhalt engagierter und pointierter auf als wir.

**100 JAHRE LANG  
DACHTE MAN, SPINAT  
HÄTTE VIEL EISEN.**

**BIS JEMAND EINE  
ZWEITE MESSUNG  
MACHTE.**

Abb.2: Zitat einer Opel-Werbekampagne [7], [8].

### 3. Das größte Problem

Fälschungen und die bewusste Manipulation von Forschungsergebnissen sind ein gravierendes Problem – sowohl in der Fachphysik [9], [10] wie auch in allen anderen empirisch forschenden Wissenschaftsgebieten.

Die Dunkelziffer ist erheblich. Schätzungen reichen von Größenordnungen zwischen 1% und 2 % aller wissenschaftlichen Arbeiten [10] bis hin zu der Feststellung: „Aus anonymen Umfragen von Kollegen in den Lebenswissenschaften ergibt sich, dass in etwa einem Viertel aller Forschungsprojekte vermutlich nicht sauber gearbeitet wird“ [11]. Gemeint ist hier ganz offensichtlich nicht ein versehentliches, unbewusstes unsauberes Arbeiten, das den Kolleginnen und Kollegen nicht bekannt war, sondern wissenschaftlich unsauberes Arbeiten, das von den Betroffenen in anonymen Umfragen eingeräumt werden konnte und von dem sie somit wissen mussten.

Diesen Forscherinnen und Forschern war während der Durchführung der Forschungsprojekte eindeutig bewusst, dass sie unwissenschaftlich – und damit vielleicht größenordnungsmäßig zurückhaltender aber in der Tendenz vergleichbar zu Jan Hendrik Schön [9] – agierten.

Gerade in der Medizin scheint ein solches Vorgehen gang und gäbe und zur Sicherung wirtschaftlicher Interessen strukturell verankert zu sein<sup>1</sup>. Pharmafirmen verzerren oder verschweigen in der öffentlichen

<sup>1</sup> Bei Krankheiten mit unklarer Forschungslage und einem rein auf Vermarktung ausgerichteten wirtschaftlichem Umfeld scheinen Zeitungsschilderungen über Einzelfallerlebnisse wie [12] glaubwürdiger als jede publizierte wissenschaftliche Studie: „Ein MS-Statistiker gewährt ihm Einblick in die Datensätze, unter der Auflage, dass er nicht schreibt, was er liest. Der Statistiker fragt: Spritzen Sie sich eines dieser Medikamente? Nein, sagt Sven Böttcher. Sehr gut, sagt der Mann. Sven Böttcher sammelt die Fakten. Tausend bis 2 000 Euro pro Monat kosten MS-Medikamente. 9,6 Milliarden verdient die Pharmaindustrie damit 2011. „Obwohl niemand weiß, woher die Krankheit kommt, wie sie verläuft, was hilft. Obwohl langfristige Studien zeigen, dass die Medikamente keinen Unterschied machen“ [12].

Darstellung offenkundig bewusst und zielgerichtet wesentliche Erkenntnisse ihrer Studien [13] oder outsourcen die Erstellung dieser Studien unter massiver direkter oder indirekter finanzieller Anreizgebung in wirtschaftlich schwache Drittländer [14] mit der Garantie, dass diese dann die gewünschten Ergebnisse liefern [15].

Für die direkt Betroffenen sind diese ethisch verwerflichen und unwissenschaftlichen Auswüchse der Wissenschaft mit tragischen persönlichen Konsequenzen verbunden – bis hin zu Todesfällen oder massiven Nebenwirkungen, die durch solche massiv manipulierten medizinischen Studien [14] nicht aufgeklärt, sondern verschleiert werden.

Wir als Wissenschaftscommunity haben ein erhebliches Glaubwürdigkeitsproblem, denn die genannten 25 % – ein Viertel aller Forschungsprojekte [11] – stellt eine enorme Anzahl bewusst unsauberer und damit manipulierter Studien dar. Es tangiert uns alle auch wirtschaftlich ganz direkt, wenn politische Entscheidungsträger, die über die finanzielle Verteilung der Mittel im Hochschulbereich zu entscheiden haben, wie in [11] mit solchen Zahlen argumentieren.

Dennoch ist eine hohe Täuschungs- und Betrugsrate empirischer Studien nicht unser eigentliches und nicht unser größtes Problem. Denn es kommt für uns alle noch weit schlimmer: So fasst die Wissenschaftsredaktion von ZEIT Online beim Vergleich von bewusst manipulierten und unbewusst fehlerbehafteten Studien überblicksweise zusammen: „Im Geschäft der Erkenntnisgewinnung läuft zu viel schief. Das größte Problem sind dabei nicht etwa Plagiatoren und bewusste Datenfälschung, sondern das Grundgerüst der Forschung selbst“ [5].

**DAS GRÖSSTE PROBLEM ... (IST) DAS GRUNDGERÜST DER FORSCHUNG SELBST.**

Abb.3: Schlussfolgerung auf ZEIT Online [5].

In erster Linie zerstört nicht die beschämende, direkte Manipulation von Forschungsergebnissen die Glaubwürdigkeit der empirisch arbeitenden Wissenschaftler, sondern es scheint überhaupt keine Glaubwürdigkeit vorhanden zu sein, die zerstört werden könnte: Das Grundgerüst der Forschung ist so brüchig, so unwissenschaftlich, dass selbst korrekt und gewissenhaft arbeitende Wissenschaftler unwissen-

schaftliche Ergebnisse erhalten – ganz so, wie von Prof. Beck-Bornholdt eingangs beschrieben.

#### 4. Relation von Referenzwissenschaft und Didaktik im Bereich der Physik

Im Bereich der Physikdidaktik wird die Problematik der Fundierung unserer Forschungsbemühungen beim Vergleich mit der Referenzwissenschaft unserer Disziplin – der Physik – sichtbar.

Physik und Astronomie sind Jahrtausende alte Wissenschaften, deren Ursprünge in den Kulturen Chinas, Indiens, des Zweistromlands und Griechenlands zu finden sind. Mit Aristoteles und seiner Fundierung des antiken, griechischen Weltbildes der Physik und Astronomie schuf die Menschheit einen für über zweitausend Jahre gültigen, fest geschlossenen und in sich (nahezu) konsistenten konzeptuellen Rahmen, der die mit den Mitteln ihrer Zeit experimentell zugänglichen und empirisch überprüfbaren Phänomene zuverlässig beschrieb.

Dieses äußerst erfolgreiche, stabile und langlebige Konzept von Aristoteles und seinen Zeitgenossen, die wissenschaftlich genauso solide und fundiert analysieren, logisch schlussfolgern und kognitiv anspruchsvoll argumentieren konnten wie heute wir<sup>2</sup>, war in der Tat konzeptionell so festgefügt, dass es nicht in einem evolutionären Übergang weiterentwickelt, sondern nur in einem revolutionären Umbruch aus der Welt geschafft und ersetzt werden konnte.

Diese vor 500 Jahren durch Kopernikus, Galilei, Kepler und dann Newton initiierte wissenschaftliche Umgestaltung unseres Weltbildes, die wir heute als ‚Kopernikanische Wende‘ bezeichnen, ist eine Grunderfahrung in der Rezeption unserer Referenzwissenschaft, der Physik.

Wesentlicher konzeptueller Bestandteil der Kopernikanischen Wende war die Erkenntnis Galileis, dass naturwissenschaftliche Sachverhalte in der Sprache der Mathematik zu verfassen sind. Das führt zu einem paradoxen Zweiklang: Zum einen gelingt durch den Rückgriff auf mathematische Strukturen und Erkenntnisse, die, falls einmal bewiesen, immer gültig sind, eine extrem solide Fundierung der wissenschaftlichen Sachverhalte, wie sie nur eine logisch-axiomatisch strukturierte Geisteswissenschaft zu leisten vermag.

Das paradoxe Gegenstück, gewissermaßen die andere Seite der gleichen Medaille, führt jedoch zu einem Dahinschmelzen der als solide geglaubten empirischen Fundierung. Carl Friedrich von Weizsäcker analysiert dies mit deutlicher Klarheit: „Die neuzeitliche Naturwissenschaft hat ihren eigenen histori-

schen Mythos. Es ist der Mythos von Galilei, (... denn) die Hauptschwäche des Aristoteles war, dass er zu empirisch war. Deshalb brachte er es nicht zu einer mathematischen Theorie der Natur. Galilei tat seinen großen Schritt, indem er wagte, die Welt zu beschreiben, wie wir sie nicht erfahren. Er stellte Gesetze auf, die in der Form, in der er sie aussprach, niemals in der wirklichen Erfahrung gelten und die darum niemals durch irgendeine einzelne Beobachtung bestätigt werden können“ [17], [18, S. 95/96].

Die Kopernikanische Wende umfasst somit gerade nicht nur inhaltliche Aspekte wie die Ersetzung der geozentrischen Weltansicht durch eine heliozentrische, sondern auch wesentliche methodische Aspekte, die sich insbesondere bezüglich der Wissensakquisition und Wissensgenese zeigen.

**DIE HAUPTSCHWÄCHE  
DES ARISTOTELES  
WAR, DASS ER ZU  
EMPIRISCH WAR.**

Abb.4: Carl Friedrich von Weizsäcker in [17].

Das Scheitern der vor-kopernikanischen Naturwissenschaft liegt nach Weizsäckers Analyse gerade darin begründet, dass sie zu starr an sichtbaren Oberflächenstrukturen verhaftet war, aber konzeptuell weder tiefer gehen konnte noch wollte: Die vor-kopernikanische Physik war zu empirisch.

Diese Analyse lässt sich auf die historisch betrachtet noch verhältnismäßig junge Disziplin der Physikdidaktik übertragen. Es benötigt seine Zeit, bis eine neu begründete Wissenschaft wie die Physikdidaktik ihre inneren Strukturen ordnet und zu sich findet.

Ebenso wie die Fachphysik hat die Physikdidaktik, wie andere Didaktiken auch, einen deutlichen empirischen Anteil. Und so lautet die Kernthese dieses Beitrags: In der Physikdidaktik besteht unsere heutige Hauptschwäche darin, zu empirisch oder zumindest der Empirie gegenüber zu unkritisch zu sein. Die Physikdidaktik hat ihre Kopernikanische Wende mithin noch vor sich.

**DIE HAUPTSCHWÄCHE  
DER HEUTIGEN PHY-  
SIKDIDAKTIK BESTEHT  
DARIN, ZU EMPIRISCH  
ZU SEIN.**

Abb.5: Zentrale Kernthese dieses Beitrags.

<sup>2</sup> Hardy formulierte dies einmal so: „The Greeks first spoke a language which modern mathematicians can understand; as Littlewood said to me once, they are not clever schoolboys or ‘scholarship candidates’, but ‘Fellows of another college’“ [16, S. 81].

Diese Kopernikanische Wende der Physikdidaktik (und anderer Didaktiken) wird zwangsläufig im Befund des vorigen Abschnitts ihren Ausgang nehmen: Das Grundgerüst der empirischen Forschung selbst ist das größte Problem (siehe Abb. 3). Unser derzeitiges Vorgehen ist wissenschaftlich problematisch und kann zu fehlerhaften Schlussfolgerungen führen.

**DIE PHYSIKDIDAKTIK  
HAT IHRE  
KOPERNIKANISCHE  
WENDE NOCH VOR  
SICH.**

**Abb.6:** Eine Schlussfolgerung aus der Kernthese dieses Beitrags.

Dieses fehlerbehaftete Vorgehen kann auf zwei Ebenen diskutiert werden: Zum einen auf grundlegender Ebene durch Analyse der in der Empirie verwendeten Vorgehensweisen, was im folgenden Abschnitt geschehen soll. Und zum zweiten anhand diverser Beispiele, die in den dann nachfolgenden Abschnitten analysiert werden.

### 5. Grundsätzliches: Das Papstproblem

Das grundsätzliche Problem des empirischen Forschens besteht darin, dass prinzipiell nie die Richtigkeit oder Unrichtigkeit einer Hypothese durch Tests, die auf Wahrscheinlichkeitsaussagen beruhen, belegt bzw. bewiesen werden kann.

Selbstverständlich war dieser einfache Sachverhalt auch den Gründungsvätern der modernen Wahrscheinlichkeitstheorie bekannt: „No test based upon a theory of probability can by itself provide any valuable evidence of the truth or falsehood of that hypothesis,“ schrieben Neyman und Pearson 1933 [19, S. 291], [20, S. 7 & S. 193].

**NO TEST BASED UPON A  
THEORY OF PROBABILITY  
CAN BY ITSELF PROVIDE  
ANY VALUABLE EVIDENCE  
OF THE TRUTH OR  
FALSEHOOD OF THAT  
HYPOTHESIS.**

**Abb.7:** Neyman und Pearson, zitiert nach [19, S. 291].

Das einzige, das man tun kann, besteht darin, sicher-

zustellen, dass man auf längere Sicht mit seinen Hypothesen nicht allzu oft falsch liegt: „Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour with regard to them in following which we insure that, in the long run of experience, we shall not be too often wrong“ [19, S. 291]. Unter vielen wahrscheinlich korrekten Hypothesen rutschen auch ein paar falsche durch.

In einer geradezu kafkaesken Umdeutung dieser Aussage von Neyman und Pearson besteht die heutige Forschungspraxis oft darin, viele wahrscheinlich falsche Hypothesen aufzustellen, unter denen ein paar korrekte erhofft werden. Dubben beschreibt dieses Vorgehen in einem Zeitungsinterview [2] recht plastisch: „Statt genau zu definieren, was Sie untersuchen wollen, sagen Sie pauschal: Ich schaue mal, ob Medikament A besser ist als Medikament B. Dann können Sie an Ihren Patienten eine Reihe von Parametern messen: Blutdruck, Cholesterinspiegel, Lungenvolumen und vieles andere.

Und am Schluss machen Sie eine statistische Erhebung für alle diese Endpunkte. Wenn Sie 100 Parameter untersuchen, werden Sie im Mittel fünf Ergebnisse erzielen, bei denen Ihr Medikament zufällig besser abschneidet.

(...) Ich habe gerade erst eine Untersuchung in die Hand bekommen, die international publiziert wurde. Da haben Forscher untersucht, welchen Einfluss das Alter des Fahrers auf einen Verkehrsunfall hat, ob er Alkohol getrunken hat, wie das Wetter war, welche Farbe das Auto hatte und zig andere Dinge. Natürlich haben sie ein statistisch signifikantes Ergebnis gefunden: Silberfarbene Autos haben weniger Unfälle“ [2].

Das moderne Data Mining als Artefakt der vom Internet erzeugten Datenflut wird diese Art der Forschung sicher noch befördern. Hübsche Beispiele für signifikante und hochsignifikante, mit ziemlicher Sicherheit aber wohl doch zufällige Zusammenhänge finden sich in [21], [22].

Ärgerlicherweise baut die Logik wissenschaftlicher Erkenntnisgewinnung durch empirische Erhebungen

**DER ÜBERGANG VON ABSOLUTER  
SICHERHEIT ZU EINER  
WAHRSCHEINLICHKEITSAUSSAGE  
FÜHRT ZU EINER FALSCHEN  
SCHLUSSFOLGERUNG.**

**UNGLÜCKLICHERWEISE IST DIES ABER  
FORMAL EXAKT DAS VERFAHREN,  
DAS BEIM STATISTISCHEN TEST VON  
HYPOTHESEN EINGESETZT WIRD.**

**Abb.8:** Analyse der von uns allen genutzten statistischen Testverfahren in [23, S. 211]. Die Ablehnung von Nullhypothesen basiert auf einen logischen Fehler.

auf einer solchen kafkaesken Umdeutung (siehe Abb. 8) auf.

Die Grundlagen logischen Schließens bei sicheren Aussagen, die einer Irrtumswahrscheinlichkeit von  $p = 0$  unterliegen und somit immer gültig sind, gehen auf Aristoteles zurück. Folgt aus einer solchen sicheren Aussage A eine Aussage B, so kann mit absoluter Sicherheit, also ebenfalls mit einer Irrtumswahrscheinlichkeit von  $p = 0$ , geschlossen werden, dass aus der gegenteiligen Aussage  $\bar{B}$  (Negation von B) die zu A gegenteilige Aussage  $\bar{A}$  (Negation von A) folgt.

Wird somit das Gegenteil von B gemessen, so ist die Aussage A mit Sicherheit falsch.

$$A \rightarrow B \quad \Rightarrow \quad \bar{B} \rightarrow \bar{A} \quad \{1\}$$

$$N \rightarrow \bar{M} \quad \Rightarrow \quad M \rightarrow \bar{N} \quad \{2\}$$

Oder aber aus einer fiktiven, absolut sicheren Nullhypothese N folge, dass die Messwerte M nicht eintreffen und somit nur  $\bar{M}$  als Negation von M gemessen werden darf. Wird dennoch ein Messwert M gemessen, so ist diese Nullhypothese zwangsläufig nicht wahr (siehe Gleichung 2).

Diese Logik des Schließens gilt jedoch nur für absolut sichere Zusammenhänge bei Irrtumswahrscheinlichkeiten von  $p = 0$ . Sobald eine Wahrscheinlichkeitsaussage vorliegt, führt diese Logik in die Irre, selbst wenn die als Irrtumswahrscheinlichkeit bezeichnete Messgröße bei nur 5 % – oder bei einem noch geringeren Wert – liegt und somit einen signifikanten Zusammenhang vermuten lässt.

Als prägnantes Beispiel für die Fehlerhaftigkeit dieser in der Wissenschaft standardmäßig bei Überprüfen von Nullhypothesen eingesetzten Logik führen Beck-Bornholdt und Dubben das Papstproblem [23, S. 200-213], [24] an. Mit einer Irrtumswahrscheinlichkeit von 1 : 8 000 000 000 ist ein zufällig ausgesuchter Mensch nicht Papst, woraus bei einer

**WIR KOMMEN SOMIT ZU DEM SCHLUSS, DASS DAS ALLSEITS ANERKANNTE UND VERBREITETE WISSENSCHAFTLICHE VERFAHREN ZUM VERGLEICH DER WIRKSAMKEIT ZWEIER BEHANDLUNGEN, DIE SOGENANNT ABLEHNUNG DER NULLHYPOTHESE, UNGÜLTIG IST.**

Abb.9: Schlussfolgerung der beiden führenden Medizinstatistiker Deutschlands in [23, S. 212].

zufälligen Messung des Papstes zwangsläufig zu folgern ist, dass der Papst ein Außerirdischer sein muss.

Mit der gleichen Logik kann statistisch höchst signifikant (mit 1 : 4 000 000 000) belegt werden, dass der Papst kein Mann, sondern eine Frau sein muss. Und in der gleichen Art der Un-Logik belegen wir bei der Beurteilung von Nullhypothesen deren Unrichtigkeit (siehe Abbildung 10).

Deshalb ziehen Beck-Bornholdt und Dubben die für unseren empirischen Wissenschaftsalltag verheerende Schlussfolgerung, „dass das allseits anerkannte und verbreitete wissenschaftliche Verfahren zum

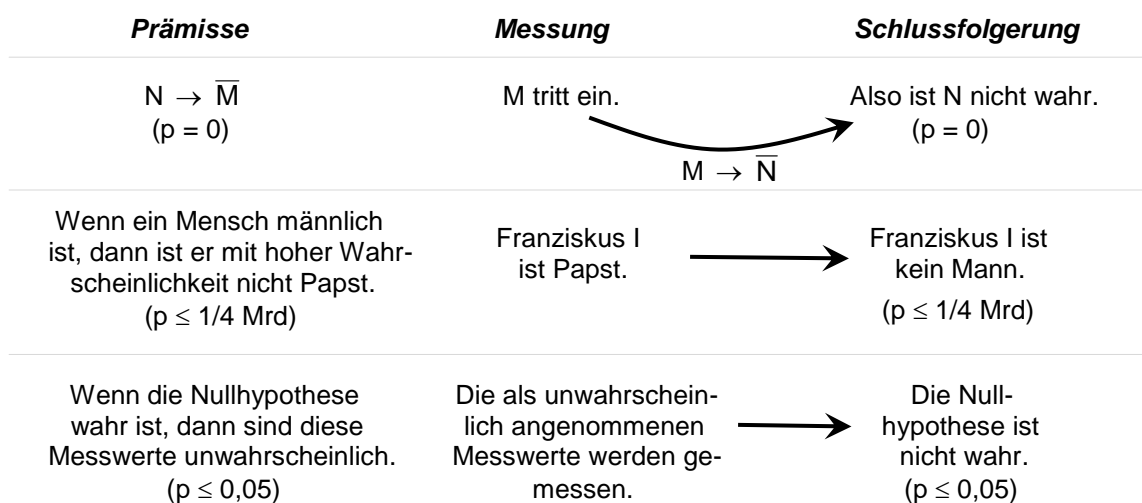


Abb.10: Die Logik wissenschaftlichen Schließens entspricht der Logik des Papstproblems. Darstellung in Anlehnung an Beck-Bornholdt und Dubben in [23, S. 213].

Vergleich der Wirksamkeit zweier Behandlungen, die sogenannte Ablehnung der Nullhypothese, ungültig ist“ [23, S. 212], siehe auch Abbildung 9.

Das ist nichts mehr und nichts weniger als die Bankrotterklärung für unseren empirisch ausgerichteten Wissenschaftsbetrieb. Wir arbeiten nicht auf einem unsicheren und schwankenden Fundament. Wir arbeiten ganz ohne Fundament – im statistisch luftleeren Raum.

Sollten Beck-Bornholdt und Dubben recht haben, wovon ich nach einigem Nachdenken sehr überzeugt bin, dann besteht unsere empirisch ausgerichtete Arbeit derzeit aus nichts anderem als einem „intelligenten Raten“, was keine schlechten Ergebnisse bringen muss, denn dann „sind Studien und Statistik ohnehin überflüssig<sup>3</sup>, wenn die Experten so gut sind, wie sie glauben“ [20, S. 222].

»DANN IST DOCH DIESE GANZE  
IRRTUMSWAHRSCHEINLICHKEITS-  
RECHNEREI MIT P-WERTEN UND  
ALLEM DRUMHERUM IN SO  
GENANNTEN WISSENSCHAFTLICHEN  
FACHZEITSCHRIFTEN NUR EIN  
**FAULES VOODOO-RITUAL?**  
DAS KANN DOCH NICHT SEIN!  
WARUM MACHEN ES DANN ALLE?«  
»VIELLEICHT, WEIL ES BEQUEM IST  
UND DIE KARRIERE FÖRDERT?«

Abb.11: Bettina und ihr Cousin Matthias im Gespräch [20, S. 168].

## 6. Beispielfelder für ein Versagen der Empirie

Doch leider sind wir und unsere Kolleginnen und Kollegen aus anderen Fachgebieten als Experten offenkundig bei weitem nicht so gut. Es fällt in der Physikdidaktik (und anderen Didaktiken) nur nicht so auf wie in anderen Fachgebieten.

Dies wird vor allem auf die unterschiedliche gesellschaftliche Relevanz und wirtschaftliche Bedeutung der einzelnen Forschungsdisziplinen zurückzuführen sein: In der Medizin geht es um Leben und Tod. In der Didaktiken geht es um Lernen oder Nicht-Lernen – verglichen mit der Medizin ein aus gesell-

<sup>3</sup> Das ist von den Autoren ernst gemeint, denn an anderer Stelle erläutert einer ihrer Protagonisten: „Wenn die Experten so gut sind, wie sie glauben, dann sind Studien und Statistik völlig überflüssig. Man muss nur sechs Koryphäen zusammentrommeln und diese sechsköpfige Sphinx befragen. – Sind die Experten nicht so gut und ihre Wahrscheinlichkeiten für gute Ideen obendrein nicht genau bestimmbar, dann sind Studien und Statistik ebenfalls überflüssig, da die Irrtumswahrscheinlichkeit sehr hoch und nicht messbar ist“ [20, S. 199].

schaftlicher Sicht also vergleichsweise irrelevantes Problem.

Fehlerhafte Studien in der Medizin führen dazu, dass Kranke falsch behandelt und vorzeitig sterben werden. Damit betreffen sie einen Sachverhalt, der uns alle früher oder später betrifft, was für das Physiklernen nicht zutrifft – und spätestens in der Oberstufe ohne größere negative Konsequenzen von Schülerinnen und Schülern mithilfe einer ausweichenden Kurswahl erledigt werden kann.

Deshalb sind die öffentlich geführten Debatten unter dem Stichwort: „Rettet die Wissenschaft“ [4], [5], [6] und etliche populärwissenschaftliche Bücher zu diesem Themenbereich [20], [23], [25] auch sehr medizinlastig. Und sie zeigen deutlich, dass die klinischen Studien, die auf der in Abschnitt 5 diskutierten Methodik aufbauen, oft versagen. Dieses

**NUR JEDES FÜNFTE  
NEUE MEDIKAMENT  
BRINGT MEHR NUTZEN.**

Abb.12: Spiegel Online-Schlagzeile über den Nutzen der 2011 in Deutschland neu zugelassenen Medikamente [26].

Versagen bleibt langfristig den direkt Betroffenen – also den Patienten sowie den die Gesundheitsleistungen finanzierenden Krankenkassen und staatlichen Körperschaften – nicht verborgen.

Die Reaktion darauf kann nun darin bestehen, sich entweder Gedanken über die Verbesserung dieser Praxis zu machen und alternative wissenschaftliche Herangehensweisen zu diskutieren, wie dies Beck-Bornholdt und Dubben in [20, Kap. 22, S 241-253] tun.

Oder aber, man belässt es bei dieser wissenschaftlichen Praxis eines „faulen Voodoo-Rituals“ [20, S. 168] und versucht anschließend, in einem korrigierenden weiteren Schritt die medizinisch unzuverlässigen Studien und Medikamente auszusortieren.

In Deutschland haben sich die politisch Verantwortlichen für die zweite Strategie entschieden und stellen dabei – ganz wie Prof. Zöllner in [11] – die wirtschaftliche Kosten-Nutzen-Relation in den Vordergrund der Korrekturbemühungen.

So werden seit Inkrafttreten des Arzneimittelmarktneuordnungsgesetzes (AMNOG) im Januar 2011 alle neu zugelassenen Medikamente auf ihren tatsächlichen wirtschaftlich-therapeutischen Wert im medizinischen Versorgungsalltag (Effectiveness) [28, S. 13] untersucht, um in Verhandlungen mit den Pharmafirmen zu einer angemessenen Preisfestsetzung zu gelangen.

**ABOUT 1 IN 6 NEW PRODUCTS OR INDICATIONS WERE RATED “NOT ACCEPTABLE“ (...) THE DRUG SHOULD SIMPLY NOT BE USED IN THE INDICATION IN QUESTION.**

**Abb.13:** Einschätzung der 2011 in Frankreich neu zugelassenen Medikamente [29, S. 107].

Dabei zeigt sich, dass unter den in Deutschland im Jahr 2011 neu zugelassenen 73 Medikamenten nur 14 Medikamente (also knapp 20 %) [26] bzw. unter 20 untersuchten neu zugelassenen Wirkstoffen nur 3 Wirkstoffe (also 15 %) [27], [28] einen entsprechenden beträchtlichen Zusatznutzen aufwiesen.

80 % aller Medikamente bzw. 85 % der hier untersuchten Wirkstoffe sind somit therapeutisch von nur sehr geringem Nutzen oder gar nutzlos, obwohl zuvor in klinischen Studien ihre pharmazeutische Qualität und Wirksamkeit (Efficacy) [28, S. 12] statistisch signifikant nachgewiesen wurde.

Dabei sind diese Befunde [28] naturgemäß noch moderat austariert. Krankenkassen und Pharmaindustrie gelangen bei ihren Preisfestsetzungen immer zu einem positiven Preis. Negative Preise für Medikamente, die einen höheren therapeutischen Schaden als Nutzen verursachen, sind in diesem System nicht vorgesehen.

Internationale Studien zeigen allerdings, dass der Anteil der neuen Medikamente, die einen therapeutischen Rückschritt darstellen, in etwa genauso hoch ist wie der Anteil der Medikamente, die einen therapeutischen Fortschritt ermöglichen [28, S. 15], [29], [30]. Unter den in Frankreich neu zugelassenen 92 Medikamenten des Jahres 2011 [29] bzw. 82 Medikamenten des Jahres 2012 [30] wies kein Medikament einen beträchtlichen Zusatznutzen auf. Nur drei bzw. vier zeigten immerhin einen Vorteil im therapeutischen Ansatz. Jedoch befinden sich unter diesen Medikamenten 16 (also ca. 17 %) im Jahr 2011 [29] und 15 (ca. 18 %) im Jahr 2012 [30], die als „nicht akzeptabel“ eingeschätzt wurden, da ihr therapeutischer Schaden den therapeutischen Nutzen übersteigt.

Auch diese Medikamente hatten vor der Zulassung alle Phasen klinischer Studien erfolgreich durchlaufen, in denen Verträglichkeit und Sicherheit des Medikaments (Phase 1), Dosisfindung und Überprüfung von Therapiekonzepten (Phase 2) sowie der Wirkungsnachweis (Phase 3) untersucht und überprüft wurde.

Dieses Versagen der klinischen Studien ist dabei kein Charakteristikum der klinischen Praxis, sondern stellt schlichtweg eine natürliche Fortsetzung der wissenschaftlichen Praxis in der biotechnologischen Grundlagenforschung dar.

So gelang es Forschern der Biotechnologie-Firma Amgen nur in 10 %, Forschern von Bayer Health Care nur in 25 %, die Ergebnisse onkologischer Studien aus dem Universitätsbereich in ihren Laboren zu reproduzieren [31]. 75 % und 90 % dieser Studien sind somit Zufallsbefunde – ein für die universitäre Forschung katastrophaler Befund.

Dieser Befund ist jedoch nicht auf die medizinische und biotechnologische Grundlagenforschung beschränkt. In zahlreichen weiteren Disziplinen zeigt sich eine ähnlich hohe Versagensrate in der empirischen Forschung.

Insbesondere in der Psychologie, deren wissenschaftliche Herangehensweise sich deutlich mit der der Didaktiken überschneidet, erschrecken zwei Erkenntnisse.

Zum einen zeigt sich, dass Fragebogenstudien oft wertlos sind. Zwar lassen sich mit Hilfe von Fragebögen sehr zeitgünstig hohe Fallzahlen generieren, die dann statistisch beeindruckte, hochsignifikante Ergebnisse liefern. Doch es macht eben einen Unterschied, ob tatsächliches Verhalten von einem neutralen Dritten (Wissenschaftler) beobachtet oder per Fragebogen vom Probanden selbst beschrieben wird.

„Auf dem Papier ist jeder mutig“, fasst dies die ZEIT [32] aphoristisch zusammen. Gleiches gilt für die Erforschung motivationaler Prozesse und anderer psychologischer Variablen bei Lernvorgängen zur Physik: Auf dem Papier ist jeder so, wie es gerade sozial erwünscht ist – und eben nicht so, wie er wirklich ist. Fazit: „Wie Menschen handeln, hat also oft nicht viel mit dem zu tun, was sie sagen“ [32].

Zum zweiten zeigt sich jedoch, dass auch dann, wenn in der experimentellen Psychologie das tatsächliche Verhalten in realen Situationen untersucht wird, sich die Ergebnisse oft nicht reproduzieren lassen. Die Situation scheint mittlerweile dramatisch zu sein, da selbst klassische Lehrbuchversuche und eine erhebliche Anzahl grundlegender Studien der Sozialpsychologie nicht eindeutig reproduziert werden konnten [4], [33], [34].

**WIR KOMMEN UNSERER VERANTWORTUNG GEGENÜBER DER ÖFFENTLICHKEIT NICHT NACH.**

**Abb.14:** Der Psychologe Robert Caldini über die Forschung mit Hilfe standardisierter Fragebögen [32].

**I SEE A TRAIN WRECK  
LOOMING. ... THE FIRST  
MISSION ... WOULD BE TO  
ORGANIZE AN EFFORT TO  
EXAMINE THE REPLICABIL-  
ITY OF PRIMING RESULTS.**

**Abb.15:** Der Psychologe und Nobelpreisträger Daniel Kahneman in einem Email-Appell an seine Kolleginnen und Kollegen [35].

### 7. Eine erste Schlussfolgerung

Als einen möglichen Lösungsansatz, um die „doubts about the integrity of psychological research“ [35] zu zerstreuen, schlug der Psychologe und Nobelpreisträger Daniel Kahneman 2012 in einem eindringlichen Appell an seine Fachkolleginnen und Fachkollegen vor, zentrale Versuche der Sozialpsychologie und insbesondere der Priming-Forschung zu reproduzieren [35].

Gerade auch aufgrund dieses offenen Email-Appells wurde mittlerweile ein solches „Reproducibility Project“ [36] in der Psychologie gestartet. Und Kahneman weist zurecht darauf hin, dass „success (say, replication of four of the five positive priming results) would immediately rehabilitate the field“ [35].

Ob aber eine solche Rehabilitation der empirischen sozialpsychologischen Forschung und damit indirekt der methodischen Ansätze der empirischen Forschung auch anderer Fachgebiete gelingt, ist zweifelhaft. Erste Ergebnisse belegen – ganz in Analogie zu den Befunden von Begley und Ellis [31] aus der Biotechnologie – eine hohe Rate nicht-reproduzierbarer Studien.

Aber auch ein solches Resultat ist ein Befund, vor dem wir die Augen nicht verschließen dürfen. Wissenschaft ist in erster Linie Wahrheitssuche – und

immer dann, wenn dieser Grundsatz verletzt wird, verliert wissenschaftliches Handeln seine Berechtigung.

Der erster Schritt in diese Richtung einer empirischen Wissenschaft echter Wahrheitssuche besteht naturgemäß darin, sich die Problematik erst einmal einzugestehen: „To deal effectively with the doubts you should acknowledge their existence and confront them straight on“ [35].

Leider scheint die Physikdidaktik derzeit noch nicht einmal so weit, diesen ersten Schritt zu gehen. Deshalb sollten wir uns deutlich vor Augen stellen, dass folgende Schlussfolgerung zwingend gilt: Solange physikdidaktische Forschungsarbeiten nicht standardmäßig reproduziert werden, kann die Hypothese, dass empirische physikdidaktische Arbeiten in ähnlicher Größenordnung wie in anderen Forschungsfeldern fehlerbehaftet und als Zufallsbefunde wissenschaftlich wertlos sind, nicht zurückgewiesen werden.

Lügen und Täuschen sind in erster Linie Selbstbetrug. Erfolgreiche Lügnerinnen und Lügner belügen sicherlich mit erheblichem Erfolg ihre Umgebung. Aber das ist nicht ihr eigentliches Erfolgsrezept. Sie sind vor allem deshalb erfolgreich im Lügen und Täuschen, weil es ihnen gelingt, sich selbst zu belügen und zu täuschen.

Nicht anders ist es in der Forschung. Beck-Bornholdt und Dubben schreiben dazu: „Unter den gegenwärtigen Bedingungen der Forschungsförderung wächst die Bereitschaft, mit Tricks und Täuschungsmanövern zum Ziel zu kommen. Wir haben den Eindruck, dass dies auch unbewusst geschehen kann, denn die uns persönlich bekannten Kollegen sind im Allgemeinen von ihren Ergebnissen überzeugt“ [23, S. 214/215].

Ignorieren wir die Befunde anderer Disziplinen, die in der methodischen Herangehensweise unserer Disziplin – der Physikdidaktik – vergleichbar sind, und ignorieren wir die daraus zu ziehenden Schlussfolgerungen (siehe Abbildung 16), so stellt dies einen wissenschaftlichen Selbstbetrug dar.

**SOLANGE PHYSIKDIDAKTISCHE FORSCHUNGSARBEITEN  
NICHT STANDARDMÄßIG REPRODUZIERT WERDEN,  
KANN DIE HYPOTHESE, DASS EIN ERHEBLICHER ANTEIL  
ALLER EMPIRISCHEN PHYSIKDIDAKTISCHEN ARBEITEN  
FEHLERBEHAFTET UND ALS ZUFALLSBEFUNDE WISSEN-  
SCHAFTLICH WERTLOS SIND, NICHT ZURÜCKGEWIESEN  
WERDEN.**

**Abb.16:** Wissenschaftliches Handeln beruht auf der Reproduzierbarkeit wissenschaftlicher Ergebnisse.



### 8. Eine zweite Schlussfolgerung

Die erste Schlussfolgerung ist zwingend. Ohne gesicherte Reproduktion unserer Ergebnisse sind diese wissenschaftlich wertlos.

Es gibt allerdings noch eine weitere mögliche Konsequenz, die aus dem Versagen der empirischen Forschung gezogen werden sollte. Diese zweite Schlussfolgerung ist nicht unbedingt zwingend, aber hilfreich, um die gravierenden Glaubwürdigkeitsprobleme der Empirie abzumildern.

Sie lautet schlicht: Reduzieren wir die empirischen Anteile unserer Forschungsdisziplin so, dass sie sich im Gleichgewicht mit anderen Forschungsausrichtungen befinden. Derzeit scheint zahlreichen Kolleginnen und Kollegen der Anteil empirischer Forschung im Bereich der Physik- und Mathematikdidaktik zu dominierend (siehe Abbildung 18) und teilweise geradezu erdrückend zu sein. Die lebhaft und engagierte Diskussion nach diesem Kurzvortrag auf der DPG-Frühjahrstagung in Frankfurt sowie einem ähnlichen Vortrag auf der diesjährigen GDM-Jahrestagung [3] spiegelt die Kritik an dieser Dominanz recht deutlich wieder.

### 9. WEITERFÜHRENDE ANMERKUNGEN

Die in den vorangehenden Abschnitten aufgezeigten grundlegenden Probleme (Abschnitt 5) und die Befunde aus Nachbardisziplinen (Abschnitte 1 und 6) rechtfertigen für sich alleine schon die Schlussfolgerungen der Abbildungen 5 und 16.

Dennoch lohnt ein tieferer Blick in die Problematik statistischer Studien, denn unser Forschungsalltag wird so stark von statistischen Anwendungen bestimmt, dass eine Analyse, weshalb zwei der führenden deutschen Medizinstatistiker zu der Überzeugung kommen, „dass die Statistik, zumindest so, wie sie heute angewandt wird, eine Sackgasse für die Forschung ist“ [20, S. 255], nur hilfreich sein kann.

**UNTER DEN GEGENWÄRTIGEN  
BEDINGUNGEN DER  
FORSCHUNGSFÖRDERUNG  
WÄCHST DIE BEREITSCHAFT,  
MIT TRICKS UND  
TÄUSCHUNGSMANÖVERN ZUM  
ZIEL ZU KOMMEN.**

**WIR HABEN DEN EINDRUCK,  
DASS DIES AUCH UNBEWUSST  
GESCHEHEN KANN,**

**DENN DIE UNS PERSÖNLICH  
BEKANNTEN KOLLEGEN SIND  
IM ALLGEMEINEN VON IHREN  
ERGEBNISSEN ÜBERZEUGT.**

Abb.17: Beck-Bornholdt und Dubben in [23, S. 214/215].

Deshalb erfolgt in den abschließenden Kapiteln eine kurze Diskussion von Problemen des statistischen Alltags, die uns alle direkt betreffen.

#### 9.1. Texanischer Schütze

Wie gehen wir mit dem Phänomen um, dass Hypothesen gelegentlich erst nach Studienabschluss angefügt oder während des Studienverlaufs nachträglich modifiziert werden?

Ohne Formulierung einer Hypothese vor Erhebung statistischer Daten können prinzipiell keine Kausalbeziehungen belegt werden. Werden statistische Daten ohne vorherige Hypothesenbildung gesamt-

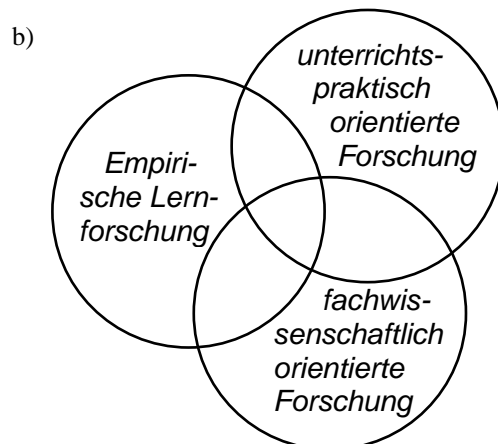
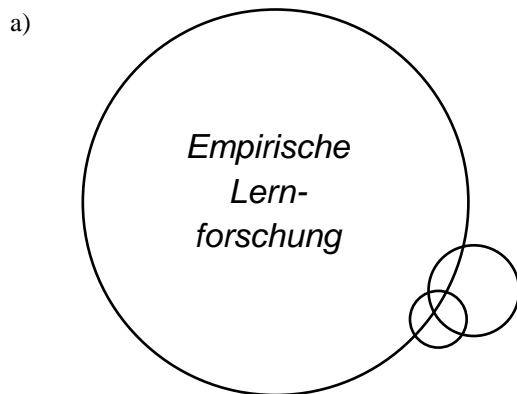


Abb.18: Struktur der physikdidaktischen und mathematikdidaktischen Forschungsausrichtung in Deutschland  
a) derzeitige Ausprägung – b) Gleichgewicht der Forschungsrichtungen.

melt, so handelt es sich hierbei lediglich um eine Beschreibung einer nur für diese eine Stichprobe und nur zu diesem einen Zeitpunkt gültigen Momentaufnahme. Schlussfolgerungen und Verallgemeinerungen für andere Lerngruppen oder für das Verhalten der gleichen Lerngruppe zu anderen Zeitpunkten können nicht gezogen werden.

**WIR SIND DER  
ÜBERZEUGUNG,  
DASS DIE STATISTIK,  
ZUMINDEST SO, WIE SIE  
HEUTE ANGEWANDT WIRD,  
EINE SACKGASSE  
FÜR DIE FORSCHUNG IST.**

Abb.19: Beck-Bornholdt und Dubben in [20, S. 255].

Der gleiche Sachverhalt trifft zu, wenn Daten zuerst gesammelt und dann auf Grundlage dieser Daten Hypothesen formuliert oder Hypothesen angepasst werden. Trotz der Konsequenz, dass dieses Vorgehen zur Wertlosigkeit wissenschaftlicher Schlussfolgerungen führt, kann ein solches illegitimes Verhalten im wissenschaftlichen Alltag beobachtet werden, wie Forscher des Reproducibility Projects berichten [4, S. 34] und Beck-Bornholdt und Dubben in ihren Büchern [20], [23], [25] plastisch beschreiben.

Auch uns als Bildungsforscher betrifft dieses Problem. So stellten sich bisher alle wissenschaftlichen Behauptungen, einen Zusammenhang zwischen Genvarianten und Bildungserfolg [37] ziehen zu können, als Folge einer texanischen Schießerei heraus.

Fazit: Wissenschaftliche Forschung ohne vorherige Hypothesenbildung ist Erbsenzählerei auf hohem Niveau und bringt uns nur dann weiter, wenn man außer der genauen Anzahl an Erbsen zu einem einzigen Zeitpunkt nichts anderes wissen will.

Und auch für die reichliche Ansammlung gezählter Erbsen in der Physikdidaktik gilt der schon klassische Ausspruch von Poincaré: „Science is built up of facts, as a house is built of stones; but an accumulation of facts is no more a science than a heap of stones is a house“ [38, S. 141].

## 9.2. Simpsons Paradoxon

Jede Studie ist multizentrisch. Wie gehen wir mit der Tatsache um, dass prinzipiell nie alle Variablen bekannt sind?

Dazu ein Beispiel: Zwischen Tierschützern und mit Tierversuchen arbeitenden Biologen gibt es erhebliche Auseinandersetzungen nicht nur um die prinzipielle Notwendigkeit oder Nicht-Notwendigkeit von Tierversuchen, sondern auch darüber, wie die Tiere

bei als notwendig erachteten Tierversuchen gehalten werden.

Unter dem Stichwort des „Environmental Enrichments“ wird diskutiert, ob beispielsweise für Labormäuse über die Grundausrüstung (Käfigmindestgröße, Zugang zu Futter und Wasser sowie Einstreu auf dem Boden) hinaus das Käfigleben durch weiteres Zubehör (Laufrad, Nagestange, Unterschlupf, etc.) angenehmer gestaltet werden sollte [39].

Das Problem allerdings ist, dass jede Änderung der Laborumgebung eine Auswirkung auf die Versuchsergebnisse haben kann. Deshalb wird für Tierversuche gefordert, „dass alle Faktoren, die einen Tierversuch beeinflussen können, strikt standardisiert werden“ [39]. Doch wenn schon onkologische Versuche mit Tieren in der Tumorforschung durch die Änderung weniger Parameter so verzerrt werden und so stark streuen, dass die Versuchsergebnisse nicht mehr verlässlich reproduziert werden können, wie sollen dann erst Untersuchungen zum Physiklernen von Menschen verlässlich interpretiert werden können?

Jeder Mensch lebt anderes, jeder Lernende hat eine andere Vorgeschichte. Beim Umgang mit Menschen variieren ja nicht nur wie beim „Environmental Enrichment“ im Tierversuch wenige Parameter, sondern eine Unzahl von Einflussgrößen, die zum einen nicht kontrolliert werden können und zum zweiten nicht einmal alle bekannt sind.

In der Lernforschung sind prinzipiell nie alle Variablen zugänglich. Dies hat einen grundlegenden Einfluss auf die eindeutige Interpretation unserer Ergebnisse – sie wird dadurch unmöglich.

Beispielsweise könnte die Untersuchung der Wirksamkeit zweier verschiedener Unterrichtsmethoden A und B bei 100 Schülerinnen und Schülern die in Abbildung 20 gezeigten Ergebnisse liefern.

Unterrichtsmethode	A	B
Schüler	50	50
im Nachtest nicht erfolgreich	30	20
im Nachtest erfolgreich	20 (40 %)	30 (60 %)

Abb.20: Veranschaulichung von Simpsons Paradoxon am Beispiel der Wirksamkeit zweier verschiedener Unterrichtsmethoden (akkumulierte Daten).

Diese Daten erwecken den Anschein, dass Unterrichtsmethode B erfolgreicher als Unterrichtsmethode A ist. Unterrichtsmethode B scheint Unterrichtsmethode A somit überlegen. Wir wissen es aber nicht, denn wir kennen nicht alle Variablen – beispielsweise

se den Fernsehkonsum der Schülerinnen und Schüler am Abend vor dem Unterricht. Hätten wir diesen Fernsehkonsum als zusätzliche Variable erhoben, könnte mit den gleichen akkumulierten Zahlen die in Abbildung 21 dargestellte Situation möglich sein.

Die Gutachter des British Medical Journal fanden in einem Text, den sie mit Vorwarnung von ihrer Herausgeberin erhielten, durchschnittlich nur ein Viertel der absichtlich eingebauten Fehler [4]. Das Peer Review-Verfahren reagiert also nicht sehr sensitiv

Unterrichtsmethode	S. haben am Abend zuvor keinen Horrorfilm gesehen		S. haben am Abend zuvor Horrorfilm gesehen		Gesamtbilanz	
	A	B	A	B	A	B
Schüler	10	40	40	10	50	50
im Nachtest nicht erfolgreich	2	12	28	8	30	20
im Nachtest erfolgreich	8 (80 %)	28 (70 %)	12 (30 %)	2 (20 %)	20 (40 %)	30 (60 %)

**Abb.21:** Veranschaulichung von Simpsons Paradoxon am Beispiel der Wirksamkeit zweier verschiedener Unterrichtsmethoden (Daten der Untergruppen).

Die Daten der Untergruppen von Abbildung 21 erwecken den Anschein, dass nun Unterrichtsmethode A erfolgreicher als Unterrichtsmethode B ist. Unterrichtsmethode B scheint Unterrichtsmethode A unterlegen zu sein, da die Erfolgsrate im Nachtest von Schülerinnen und Schülern der Lerngruppe B in beiden Untergruppen niedriger liegt als von Schülern der Lerngruppe A.

Aber letztendlich muss auch dies nicht das endgültige Resultat sein – weitere unbekannte Variablen wurden nicht berücksichtigt. Sie könnten das Resultat von Abbildung 21 auch wieder auf den Kopf stellen.

Als Fazit bleibt die einzige Gewissheit, dass wir uns nie wirklich sicher sein können, ob die von uns empirisch aufgefundenen Gesetzmäßigkeiten bzw. die von uns beschriebenen kausalen Beziehungen oder letztendlich doch das Gegenteil dieser Gesetzmäßigkeiten bzw. Beziehungen stimmt.

### 9.3. Eleganter Unsinn

Wie gehen wir mit dem Versagen des Peer Review-Systems bei Veröffentlichungen um?

Nicht erst seit der Veröffentlichung des „parodistischen Artikels“ von Alan Sokal über den „Weg einer transformativen Hermeneutik der Quantenschwerkraft“ [40] wissen wir, dass es relativ leicht ist, inhaltsleere oder sogar absurd falsche Beiträge in Fachzeitschriften zu veröffentlichen.

Aber angesichts einer immer größeren Publikationsflut und einer sich im Internet rasant ausbreitenden Welle neu gegründeter elektronischer Zeitschriften hat sich diese Problematik in den vergangenen Jahren weiter verschärft.

auf fehlerhafte Texte. Ein „erkennbar fingiertes“ Manuskript des Biologen John Bohannon über ein Krebsmittel wurde sogar von 150 Zeitschriften akzeptiert [4].

Arbeitsüberlastung und Nicht-Bezahlung der Gutachtertätigkeit mögen viel zu dieser Misere beitragen. Dennoch reicht das Problem tiefer. So zeigt die Analyse des Falls Schön, dass der Peer Review-Prozess der Fachzeitschriften – zumindest in diesem Fall – nicht Teil der Lösung, sondern eindeutig Teil des Problems war.

Der Wissenschaftsjournalist Eugenie Samuel Reich schreibt dazu: „Not only is there no guarantee that a thorough review process will detect a false claim, but even more disturbingly, a thorough review may do little more than reveal to authors what changes

**ALS DIE HERAUSGEBERIN DES RENOMMIERTEN BRITISH MEDICAL JOURNAL IHRE EIGENEN GUTACHTER MIT EINEM FEHLERGESPICKTEN ARTIKEL TESTETE, FANDEN DIESE – TROTZ VORWARNUNG – DURCHSCHNITTlich NUR WENIGER ALS JEDEN VIERTEN FEHLER.**

**Abb.22:** Das Peer Review-Verfahren reagiert nicht sehr sensitiv auf Fehler, Zitat aus [4, S. 33].

they need to make in order to turn a false claim into a more plausible scam“ [9, S. 122]. Es ist verstörend, aber ohne die hilfreichen Hinweise der Reviewer hätte es Jan Hendrik Schön bei seinen Fälschungen weit schwerer gehabt.

**DAS VERTRAUEN DARIN,  
DASS SICH DIE WISSEN-  
SCHAFT SELBST KORRI-  
GIERT, IST EXZESSIV  
UNGERECHTFERTIGT.**

Abb.23: Der Psychologe Frank Renkewitz in [4, S. 34].

#### 9.4. ESTRO-Umfrage

Wie gehen wir mit der ungenügenden Vermittlung statistischer Grundkenntnisse im Studium um? Und wie gehen wir mit den teilweise „erschütternd“ [23, S. 17] mangelhaften statistischen Kenntnissen unserer Fachkolleginnen und Fachkollegen um?

Auf einer Klausurtagung der Europäischen Gesellschaft für Radioonkologie (ESTRO) konfrontierten Beck-Bornholdt und Dubben ein Drittel der Teilnehmer mit einer grundlegenden und relativ einfachen Aufgabe [23, S. 16] zur Abschätzung einer Erkrankungshäufigkeit, die mit Hilfe bedingter Wahrscheinlichkeiten gelöst werden konnte.

Von den 15 Befragten, allesamt Experten auf dem Gebiet der Radioonkologie, darunter fünf Spezialisten für prädiktive Tests, konnte nur einer die korrekte Antwort nennen. Elf Befragte lagen mit ihrer Antwort vollständig daneben. Drei nannten ein falsches Resultat, das sich aber immerhin größenordnungsmäßig in der Nähe der korrekten Antwort befand [23, S. 17].

Es sind also nicht nur Studierende, die Probleme mit relativ elementaren Grundlagen der Wahrscheinlichkeitsrechnung und der Testtheorie haben. Auch eine nicht zu vernachlässigende Anzahl an Experten dürfte erhebliche Wissenslücken aufweisen. Und es ist anzunehmen, dass dieses Ergebnis aus der Radioonkologie auch auf die Situation in den Didaktiken extrapoliert werden kann.

Ersetzen Sie also bitte im folgenden Zitat das Wort ‚Schule‘ durch das Wort ‚Hochschule‘: „Wenn ich als Historiker Studien empirischer Bildungsforscher lese, habe ich ein Grundverständnis für deren mathematische Methoden. Und ich kenne meine Grenzen. Auch dazu ist die Schule da: einem zu zeigen, was man nicht kann, die eigene Inkompetenz zu erfahren“ [41].

#### 9.5. Scheitern muss erlaubt sein

Nicht alle neuen Ideen sind besser als die Ideen unserer Vorgängerinnen und Vorgänger, aber 100 % aller Studien scheinen genau dies zu belegen. Sind wir in der Physikdidaktik wirklich so gut?

In der Physikdidaktik gehen wir mit einem enormen Selbstbewusstsein ans Werk: Nahezu alles, was wir so Neues machen, wird in den empirischen Studien, die wir veröffentlichen, als besser befunden als das Ältere, das schon vorhanden ist.

Zum Teil mag dies am „Publication Bias“ liegen, der Tatsache, dass in den empirisch forschenden Wissenschaften positive Befunde und Wirkungsbeziehungen weit bereitwilliger veröffentlicht werden als negative oder neutrale Befunde, wenn sich das Neue als schlechter oder höchstens genauso gut wie das Bisherige herausstellt [42].

Doch es gibt in der Physikdidaktik einen zweiten Aspekt, der zu berücksichtigen ist: Es ist nahezu unmöglich, mit der Erforschung einer Unterrichtsidee oder eines physikdidaktischen Ansatzes zu promovieren, der sich nach langjährigem Bemühen dann doch als schlechter wirksam und insgesamt den Lernenden unzugänglicher herausstellt als bereits vorhandene Unterrichtsideen oder bereits existierende physikdidaktische Ansätze.

In der Physikdidaktik wird von einer dissertationswürdigen Arbeit uneingeschränkt erwartet, dass nicht nur etwas Neues untersucht wurde, sondern auch etwas Besseres. Untersucht eine bedauernswerte Promovendin oder ein bedauernswerter Promovend einen neuen physikdidaktischen Unterrichtsansatz, der sich in mühsam erhobenen Vergleichsstudien nach eventuell langjähriger schulischer Erprobung als weniger wirksam zeigt, so sind dieser Promovend und diese Promovendin dazu verdammt, so

**VON FÜNFZEHN BEFRAGTEN  
GAB NUR EINER DIE RICHTIGE  
ANTWORT ...  
EIN ERSCHÜTTERNDES ERGEB-  
NIS, WENN MAN BEDENKT, DASS  
DIE MEISTEN BEFRAGTEN AN  
EUROPÄISCHEN ODER AMERI-  
KANISCHEN HOCHSCHULEN  
LEHREN UND FÜNF ALS SPEZI-  
ALISTEN AUF DEM GEBIET DER  
PRÄDIKTIVEN TESTS GELTEN.**

Abb.24: Hans-Peter Beck-Bornholdt & Hans-Hermann Dubben in [23, S. 17].

lange weiter zu forschen, bis sie einen Befund vorlegen können, der die Überlegenheit ihres Ansatzes zeigt.

Da wir in der Physikdidaktik (wie in anderen empirischen arbeitenden Wissenschaften auch) mit statistisch fragwürdigen Techniken agieren (siehe Abschnitt 5), fällt dies in der Regel nicht schwer. Und langjährige schulische Erprobungen lassen sich notfalls ja auch durch einfache Laboruntersuchungen in geschütztem Rahmen ersetzen. Darüber hinaus helfen die in [2] beschriebenen Vorgehensweisen – bewusst oder unbewusst eingesetzt – auch in der Physikdidaktik aus der Klemme. Es ist also nicht verwunderlich, dass tatsächlich nur positive Befunde zu existieren scheinen.

Doch werfen wir einen Blick auf andere Gebiete: Welche Erfolgsquote haben die Besten ihres Faches in anderen Bereichen? Eine Produktentwicklerin der Fa. Katjes berichtet: „80 % aller neuen Süßwaren floppen“ [43]. Diese wohl zu den besten Süßwareningenieuren gehörenden Produktentwickler weisen eine Erfolgsquote von 20 % auf. Die restlichen 80 % der neu entwickelten Süßigkeiten sind aus Sicht der Verbraucher schlechter als das bereits vorhandene Süßwarenangebot.

Die Ausgaben für Forschung und Entwicklung der Fa. Vileda/Freudenberg belaufen sich derzeit auf ca. 220 Mio. Euro [44]. Das ist eine Menge Geld, und wohl nur die besten Reinigungsingenieure schaffen es in diese finanzstark ausgestatteten Labors. Sie bringen unablässig neue Produkte zur Marktreife. Das Ergebnis: „Bei der Flop-Rate gehen Experten von 90 % aus“ [44]. Nur 10 % aller neuen Reinigungsprodukte werden von den Verbrauchern in den Läden per Kaufentscheidung als tatsächlich besser eingestuft.

Die für Informatiker attraktivsten Stellen finden sich derzeit wohl bei finanzstarken Internetgiganten wie Facebook und Google, so dass dort die Besten ihres Faches hinstreben. Neue Ideen und Programmumsetzungen können relativ schnell in die Internetseiten ausgewählter User integriert und so erprobt werden. „Das Ergebnis aus Hunderttausenden solcher Tests: Rund 90 Prozent aller neuen Ideen sind schlechter als das, was schon da ist“ [45].

In nahezu allen Disziplinen, in denen neue Ideen

**RUND 90 % ALLER  
NEUEN IDEEN SIND  
SCHLECHTER ALS DAS,  
WAS SCHON DA IST.**

Abb.25: Das ZEIT-Magazin in seiner Neujahrsausgabe 2014 [45].

verhältnismäßig schnell und eindeutig getestet werden können, findet sich eine Misserfolgsrate zwischen 80 % und 90 %.

Ist es wirklich vorstellbar, dass wir in der Physikdidaktik so viel besser sind als in anderen Disziplinen? Wohl kaum – auch wenn die Erprobung neuer Ideen in der Physikdidaktik oft sehr lange dauert (sollte man diese auch tatsächlich im Unterrichtsalltag testen), so dürften doch auch unsere neuen Ideen in einer Größenordnung von 80 % bis 90 % schlechter als frühere Ideen sein.

Und hier schließt sich der Kreis zwischen den Abbildungen 25 und 1: Wenn 90 % aller neuen Ideen schlechter sind als das, was schon da ist, aber 100 % aller neuen Ideen in Studien als besser bewertet werden, als das, was schon da ist, dann müssen zwangsläufig 90 % aller Studien falsch sein.

**DENN DAS IST VIEL-  
LEICHT DAS GEMEINSTE  
AM SCHEITERN:  
ES LÄSST SICH SO  
SCHWER VOM ERFOLG  
UNTERSCHIEDEN.**

Abb.26: Das ZEIT-Magazin in seiner Neujahrsausgabe 2014 [45].

## 10. Literatur

- [1] Beck-Bornholdt (2014): Internet-Homepage unter der URL: [www.beck-bornholdt.de](http://www.beck-bornholdt.de) [30. Mai 2014].
- [2] Kosog, Simone (2006). „Ein kahler Kopf macht nicht reich“, Interview mit dem Medizinstatistiker Hans-Hermann Dubben. Der Tagesspiegel, 16. Januar 2006. Zugriff unter der URL: [www.tagesspiegel.de/zeitung/ein-kahler-kopf-macht-nicht-reich/675478.html](http://www.tagesspiegel.de/zeitung/ein-kahler-kopf-macht-nicht-reich/675478.html) [30. Mai 2014].
- [3] Horn, Martin Erik (2014): Plädoyer für eine Kopernikanische Wende in der Mathematikdidaktik. In: Roth, Jürgen; Ames, Judith (Hrsg.): Beiträge zum Mathematikunterricht 2014. Tagungsband der Gesellschaft für Didaktik der Mathematik, Band 1, WTM-Verlag, Münster, S. 563-566.
- [4] Schmitt, Stefan; Schramm, Stefanie (2013): Rettet die Wissenschaft! In: Die Zeit, Nr. 1, 69. Jg., 27. Dez. 2013, S. 33-35, siehe auch URL: [www.zeit.de/2014/01/wissenschaft-forschung-rettung](http://www.zeit.de/2014/01/wissenschaft-forschung-rettung) [02. Jan. 2014].
- [5] Zeit Online-Wissenschaftsredaktion (2013/2014): Forschung in der Krise. Übersichtskas-

- ten des Schwerpunktes „Rettet die Wissenschaft!“ [4], Zugriff unter der URL: [www.zeit.de/2014/01/wissenschaft-forschung-rettung](http://www.zeit.de/2014/01/wissenschaft-forschung-rettung) [02. Jan. 2014].
- [6] Kuhrt, Nicola (2014): Qualitätssicherung bei Studien: Rettet die Wissenschaft! In: Spiegel Online, Zugriff unter der URL: [www.spiegel.de/wissenschaft/medizin/qualitaet-der-forschung-rettet-die-wissenschaft-a-959629.html](http://www.spiegel.de/wissenschaft/medizin/qualitaet-der-forschung-rettet-die-wissenschaft-a-959629.html) [23. März 2014].
- [7] Labinski, Anne (2014): Umparken für Opel. In: Der Tagesspiegel, Online-Meldung vom 28. Feb. 2014, Zugriff unter der URL: [www.tagesspiegel.de/wirtschaft/werbekampagne-e-sucht-urheber-umparken-fuer-opel/9552780.html](http://www.tagesspiegel.de/wirtschaft/werbekampagne-e-sucht-urheber-umparken-fuer-opel/9552780.html) [30. Mai 2014].
- [8] Adam Opel GmbH (2014): Begleitende Internetseiten zur Werbekampagne „Umparken im Kopf“, Zugriff unter der URL: [www.umparkenimkopf.de](http://www.umparkenimkopf.de) [18. März 2014].
- [9] Reich, Eugenie Samuel (2009): Plastic Fantastic. How the Biggest Fraud in Physics Shook the Scientific World. Palgrave Macmillan, New York.
- [10] Al-Shamery, Katharina (2013): Die Säulen der Wissenschaft. Wie steht es zehn Jahre nach dem Fall Schön mit der guten wissenschaftlichen Praxis? In: Physik Journal, Nr. 2, Vol. 12, S. 3.
- [11] Spiewak, Martin (2013): „Die Folgekosten können hoch sein“, Interview mit dem früheren rheinland-pfälzischen Wissenschaftsminister und Berliner Bildungssenator Jürgen Zöllner. In: Die Zeit, Nr. 1, 69. Jg., 27. Dez. 2013, S. 34, siehe auch URL: [www.zeit.de/2014/01/wissenschaft-zoellner-forschung-transparenz](http://www.zeit.de/2014/01/wissenschaft-zoellner-forschung-transparenz) [02. Jan. 2014].
- [12] Novotny, Rudolf (2013): Ein unverhoffter Schub. In: Berliner Zeitung, 2. Sept. 2013. Zugriff unter der URL: [www.berliner-zeitung.de/panorama/multiple-sklerose-ein-unverhoffter-schub,10808334,24189632.html](http://www.berliner-zeitung.de/panorama/multiple-sklerose-ein-unverhoffter-schub,10808334,24189632.html) [30. Mai 2014].
- [13] Grabar, Edda; Schweitzer, Jan (2014): Getrübter Durchblick. In: Die Zeit, Nr. 25, 69. Jg., 12. Juni 2014, S. 34, siehe auch URL: [www.zeit.de/2014/25/pharmaindustrie-europaeische-arzneimittelbehoerde](http://www.zeit.de/2014/25/pharmaindustrie-europaeische-arzneimittelbehoerde) [14. Juni 2014].
- [14] Ahr, Nadine; Hawranek, Christiane (2014): Eine Überdosis Risiko. In: Die Zeit, Nr. 12, 69. Jg., 13. März 2014, S. 34, siehe auch URL: [www.zeit.de/2014/12/medikament-gilena-arzneimitteltest-russland](http://www.zeit.de/2014/12/medikament-gilena-arzneimitteltest-russland) [23. März 2014].
- [15] Vom Lehn, Brigitta (2014): Virologen befürchten Engpässe. In: Berliner Zeitung, Nr. 115, 70. Jg., 19. Mai 2014, S. 2.
- [16] Hardy, Godfrey Harold (1992): A Mathematician's Apology. With a foreword by C. P. Snow. Canto Edition, Cambridge University Press, Cambridge.
- [17] Von Weizsäcker, Carl Friedrich (1962). Kopernikus, Kepler, Galilei. In: Oehler, Klaus; Schaeffler, Richard (Hrsg.): Einsichten. Gerhard Krüger zum 60. Geburtstag. Klostermann Verlag, Frankfurt am Main, S. 376-394. Wiederabdruck in [18, S. 86-104].
- [18] Von Weizsäcker, Carl Friedrich (2002): Große Physiker. Von Aristoteles bis Werner Heisenberg. Deutscher Taschenbuch Verlag, München.
- [19] Neyman, Jerzy; Pearson, Egon S. (1933): On the Problem of the Most Efficient Tests of Statistical Hypotheses. In: Philosophical Transactions of the Royal Society of London, Series A, Vol. 231, S. 289-337.
- [20] Beck-Bornholdt, Hans-Peter; Dubben, Hans-Hermann (2003): Der Schein der Weisen. Irrtümer und Fehltritte im täglichen Denken. Zweite Auflage, rororo science # 61450, Rowohlt Taschenbuch Verlag, Reinbek bei Hamburg.
- [21] Böhm, Markus (2014): Was Nicolas Cage mit Stolperunfällen verbindet. In: Spiegel Online, Netzwelt, Zugriff unter der URL: [www.spiegel.de/netzwelt/web/spurious-correlations-korrelationen-vs-kausaler-zusammenhang-a-968848.html](http://www.spiegel.de/netzwelt/web/spurious-correlations-korrelationen-vs-kausaler-zusammenhang-a-968848.html) [12. Mai 2014].
- [22] Vigen, Tyler (2014): Spurious Correlations. Internet-Blog zu zufälligen Korrelationen statistischer Größen, Zugriff unter der URL: [www.tylervigen.com](http://www.tylervigen.com) [12. Mai 2014].
- [23] Beck-Bornholdt, Hans-Peter; Dubben, Hans-Hermann (1997): Der Hund, der Eier legt. Erkennen von Fehlinformation durch Querdenken. rororo science # 60359, Rowohlt Taschenbuch Verlag, Reinbek bei Hamburg.
- [24] Beck-Bornholdt, Hans-Peter; Dubben, Hans-Hermann (1996): Is the Pope an alien? In: Nature, Vol. 381, No. 6585, 27. Juni 1996, S. 730.
- [25] Dubben, Hans-Hermann; Beck-Bornholdt, Hans-Peter (2005): Mit an Wahrscheinlichkeit grenzender Sicherheit. Logisches Denken und Zufall. rororo science # 61902, Rowohlt Taschenbuch Verlag, Reinbek bei Hamburg.
- [26] cib; dpa (2014): G-BA-Bilanz: Nur jedes fünfte neue Medikament bringt mehr Nutzen. In: Spiegel Online, Zugriff unter der URL: [www.spiegel.de/gesundheit/diagnose/amnog-bilanz-g-ba-bescheinigt-nur-wenigen-medikamentenzusatznutzen-a-970976.html](http://www.spiegel.de/gesundheit/diagnose/amnog-bilanz-g-ba-bescheinigt-nur-wenigen-medikamentenzusatznutzen-a-970976.html) [22. Mai 2014].
- [27] irb; AFP; dpa (2014): TK-Report: Kaum ein neues Medikament bringt medizinischen Fortschritt. In: Spiegel Online, Zugriff unter der URL: [www.spiegel.de/gesundheit/diagnose/neue-medikamente-nur-die-wenigsten-bringen-mehr-nutzen-a-962198.html](http://www.spiegel.de/gesundheit/diagnose/neue-medikamente-nur-die-wenigsten-bringen-mehr-nutzen-a-962198.html) [2. April 2014].

- [28] Windt, Roland; Boeschen, Daniela; Glaeske, Gerd (2014): Innovationsreport 2014. Wissenschaftliche Studie zur Versorgung mit innovativen Arzneimitteln – Eine Analyse von Evidenz und Effizienz (Kurzfassung). Zentrum für Sozialpolitik der Universität Bremen ZeS, Bremen.
- [29] Prescrire-Redaktion (2012): New Drugs and Indications in 2011. France is better focused on patients' interests after the Mediator<sup>o</sup> scandal, but stagnation elsewhere. In: Prescrire International, No. 126, Vol. 21, S. 106-110.
- [30] Prescrire-Redaktion (2013): New Drugs and Indications in 2012. Sluggish progress, timid measures to protect patients. In Prescrire International, No. 137, Vol. 22, S. 105-107.
- [31] Begley, C. Glenn; Ellis, Lee M. (2012): Drug Development: Raise Standards for Preclinical Cancer Research. In: Nature, Vol. 483, No. 7391, 29. März 2012, S. 531-533.
- [32] Paulus, Jochen (2009): Auf dem Papier ist jeder mutig. In: Die Zeit, Nr. 40, 64. Jg., 24. September 2009, S. 45, siehe auch URL: [www.zeit.de/2009/40/PS-Studie](http://www.zeit.de/2009/40/PS-Studie) [30. Sept. 2009].
- [33] Schramm, Stefanie (2013): Ein einmaliges Ergebnis. In: Die Zeit, Nr. 22, 68. Jg., 23. Mai 2013, S. 38, siehe auch URL: [www.zeit.de/2013/22/sozialpsychologische-studien](http://www.zeit.de/2013/22/sozialpsychologische-studien) [29. Mai 2013].
- [34] Yong, Ed (2012): Nobel Laureate Challenges Psychologists to Clean Up Their Act. In: Nature, News vom 3. Oktober 2012, doi: 10.1038/nature.2012.11535, siehe URL: [www.nature.com/news/nobel-laureate-challenges-psychologists-to-clean-up-their-act-1.11535](http://www.nature.com/news/nobel-laureate-challenges-psychologists-to-clean-up-their-act-1.11535) [30. Mai 2014].
- [35] Kahneman, Daniel (2012): A Proposal to Deal With Questions About Priming Effects. Offener Email-Rundbrief vom 26. September 2012. Download unter „Supplementary information“ des Nature News-Beitrags [34].
- [36] Open Science Framework (2014): Reproducibility Project Psychology. Zugriff auf die Internetseiten des Projekts unter der URL: <https://osf.io/ezcu> [30. Mai 2014].
- [37] Schramm, Stefanie (2012): Die rechte Mischung. In: Die Zeit, Nr. 45, 67. Jg., 31. Oktober 2012, S. 40, siehe auch URL: [www.zeit.de/2012/45/Genoeconomics-Geno-Wirtschaft-Erfolg-Ashraf-Galor](http://www.zeit.de/2012/45/Genoeconomics-Geno-Wirtschaft-Erfolg-Ashraf-Galor) [30. Mai 2014].
- [38] Poincaré, Henri (1952): Science and Hypothesis. Dover Publications, New York.
- [39] Hamm, Magdalena (2009): Schöner wohnen für Labortiere. In: Die Zeit, Nr. 47, 64. Jg., 12. November 2009, S. 39, siehe auch URL: [www.zeit.de/2009/47/N-Labortiere](http://www.zeit.de/2009/47/N-Labortiere) [13. Nov. 2009].
- [40] Sokal, Alan; Bricmont, Jean (2001): Eleganter Unsinn. Wie die Denker der Postmoderne die Wissenschaften mißbrauchen. Deutscher Taschenbuch Verlag, München.
- [41] Kerstan, Thomas (2011): „Bildung ist, was übrig bleibt“, Interview mit dem Erziehungswissenschaftler Heinz-Elmar Tenorth. In: Die Zeit, Nr. 33, 66. Jg., 11. Aug. 2011, S. 30, siehe auch URL: [www.zeit.de/2011/33/Interview-Tenorth](http://www.zeit.de/2011/33/Interview-Tenorth) [15. Aug. 2011].
- [42] Beck-Bornholdt, Hans-Peter; Dubben, Hans-Hermann (2004): Unausgewogene Berichterstattung in der medizinischen Wissenschaft – Publication Bias. Universitätsklinikum Hamburg-Eppendorf UKE, Institut für Allgemeinmedizin, Hamburg.
- [43] Kobel, Jörg (2009): Dynastien in Deutschland – Katjes. Dokumentarfilm, WDR 2009, gesendet auf Eins Festival am 3. April 2014, 9:45 h – 10:30 h.
- [44] Dierig, Carsten (2014): Kampf dem Schmutz. Berliner Morgenpost, Nr. 99, 116. Jg., 10. April 2014, S. 9.
- [45] Kucklick, Christoph (2013): Schluss mit dem Scheitern! In: ZEIT-Magazin, Nr. 1, 69. Jg., 27. Dez. 2013, S. 14-22, siehe auch URL: [www.zeit.de/2014/01/scheitern-misserfolg](http://www.zeit.de/2014/01/scheitern-misserfolg) [02. Jan. 2014].