Anonyme Kritik: Peer Review auf dem Prüfstand der Wissenschaftsforschung

Gerhard Fröhlich, Linz

0 Schlagwort Evaluation

Die Evaluation wissenschaftlicher Leistungen ist zur Zeit Gegenstand vieler politischer Debatten, Gesetzesvorlagen, reißerischer Medienberichte. Bedenklicherweise stützen selbst höchste EntscheidungsträgerInnen ihre Urteile und Reformvorschläge zum Thema nicht auf dem bereits akkumulierten großen Fundus an Modellen und Befunden der theoretisch-empirischen Wissenschaftsforschung und Informationswissenschaft[1], sondern geben offensichtlich bloße private bzw. ideologische Meinungen von sich. Die Rahmenbedingungen der Wissenschaften sind also von wissenschaftlicher Analyse wie metawissenschaftlicher Reflexion ausgeblendet. Sollen die Wissenschaften somit auf irrationaler Grundlage betrieben werden?

Wissenschaftstheoretisch gesehen gehört Evaluation (im Sinne von Kritik, Bewertung von Theorien, Methoden, Befunden nach „wissenschaftsinternen" Kriterien) von vornherein zum wissenschaftlichen Alltagshandeln, ist konstitutiv für alle wissenschaftliche Methoden. Was viele WissenschaftlerInnen in letzter Zeit eher empört, ist der Versuch, von außen, etwa geleitet von Prinzipien der Marktforschung oder Controlling-Verfahren, die Parameter festzulegen.

Es geht um Gegen-Evaluation: Wer evaluiert die Evaluierer? Wer begutachtet die Gutachter? Wer misst die Messer? Neben der evaluativen Szientometrie (Stichwort: Impact Faktor; vgl. zur Kritik Fröhlich 1999a, 2002c) sind seit einiger Zeit auch die diversen Peer-Review-Verfahren in den Fokus kritischer Diskussionen, medialer Berichte und empirischer Untersuchungen geraten.

1 Kritik als Definitionsmerkmal von Wissenschaft und ihre Ohnmacht in der Alltagspraxis

Rücksichtslose" (Karl Popper) Kritik ist ein Kriterium für Wissenschaft. Doch dieser Forderung der normativen Wissenschaftstheorie wird in der wissenschaftlichen Öffentlichkeit eher selten nachgekommen, der „realexistierende" Wissenschaftsalltag sieht meist anders aus:

In der öffentlich zugänglichen wissenschaftlichen Kommunikation dominiert de facto laue bis ehrerbietige Höflichkeit: Komplimente zu verteilen ist relativ risikolos. Brisante Informationen erhält man über informelle Kanäle, Internet-Listen und Massenmedien. Wird Kritik veröffentlicht, wird sie von den Betroffenen kaum zur Kenntnis genommen. Selbst wenn Arbeiten als gefälscht zurückgezogen werden, hat dies kaum Folgen: Sie werden weiterhin kommentarlos zitiert.

Die öffentlich sichtbare Kritik bildet nur eine kleine Spitze des Eisberges kritischer Forschungskommunikation, die vor allem in zwei Kontexten stattfindet:
- in informellen Dyaden und Kleingruppen (z. B. in Form des Gegenlesens von Manuskripten im Freundeskreis),
- im Rahmen der Arkanpraxis der Zeitschriftenverlage, deren Herausgeber-Referee-Begutachteten-Interaktionen nach Modell konspirativer Organisationen funktionieren: die GutachterInnen sind für gewöhnlich anonym, kaum einer der Beteiligten weiß voneinander, und nur selten bekommen die Begutachteten die Gutachten überhaupt oder gar vollständig zu Gesicht.

KritikerInnen sehen durch die derzeit üblichen Geheimpraktiken die Offenheit der wissenschaftlichen Kommunikation bedroht (Jasanoff 1985, Munthe/Welin 1996). Als Folgen dieser Informationsvorenthaltung (vgl. Fröhlich 1998) wird der "context of discovery" zum Betriebsgeheimnis, werden Fehler und Irrwege vielfach und unnötigerweise wiederholt.

2 Peer Review und Abweisungsraten als Prestigeschmuck wissenschaftlicher Journale

In den Wissenschaften herrscht, ähnlich wie in der katholischen Kirche („nihil obstat"), offiziell Zensur. Allerdings ist sie eine relative: Niemand wird am Forschen gehindert - aus eigener Tasche finanziert, und in einem Kleinverlag vervielfältigt oder ins Web gestellt, ist sie prinzipiell jedem möglich. Doch von den Hardlinern der Evaluation werden nur Beiträge in *„peer reviewed" Journalen* als wissenschaftlich akzeptiert. Bücher oder detaillierte Forschungsberichte („Graue Literatur") seien nur von geringem Wert.

„Peer-Review" nennt sich das Kontrollsystem, der TÜV der Wissenschaften: GutachterInnen (wenn man etwas auf sich hält: *anonyme* GutachterInnen - ohne Angst vor Rache) sichern Qualität, verhindern Redundanz und Chaos - so die offiziöse Sicht. Die Referees entscheiden nicht nur über den Abdruck von Manuskripten mit, sondern auch bei Mittelverteilung, Kongresszulassung, Auswahl von Preisträgern, bei wissenschaftlichen Karrieren in allen Etappen.

Das Peer-Review-System als einheitliches Prüfsystem gibt es jedoch nicht - obwohl vielfach behauptet. Betont sei daher der Plural: Zu willkürlich unterschiedlich - und vor allem: zu unterschiedlich willkürlich - sind die Prozeduren und Betriebssitten beim „Refereeing" in einzelnen Disziplinen, bei einzelnen Kongressorganisatoren, Forschungsförderern, Journal- und Buchherausgebern. Selbst die Definitionen von „Peer" (eig.: Gleichaltriger, Gleichrangiger; meist gemeint: eine wissenschaftsinterne ExpertIn zum Thema) sind höchst vage und variabel.

Die Heterogenität, ja Unvergleichlichkeit der dutzenden Peer-Review-Varianten (inkl. blind, doppel- und dreifachblind)[2] hängt zusammen mit ihren - je nach dem Verhältnis von Angebot und Nachfrage - *unterschiedlichen Funktionen*:

- In Disziplinen mit *niedrigen Abweisungsraten* (rejection rates) zwischen 20-30% wie in der Physik sollen Gutachter Aufsätze zum Druck vorschlagen, daher wird oft das *serielle Verfahren* praktiziert: Nur ein Gutachter wird beigezogen, wenn dieser zustimmt, wird das Manuskript gedruckt, wenn dieser ablehnt, wird ein weiterer Gutachter beauftragt.

- In Disziplinen mit *hohen Abweisungsraten* (80-90%) wie in der Psychologie sollen Gutachter hingegen Argumente für die Ablehnung von Manuskripten liefern. Daher werden *hier parallele Verfahren* mit 2 oder 3 Gutachtern zugleich bevorzugt (je mehr Gutachter, desto mehr negative Voten; bei 4 Gutachtern fand sich in der Politikwissenschaft kein einziges Manuskript ohne mindestens eine Ablehnung). Bei Diskrepanz in der Beurteilung wird das Manuskript hier

meist abgewiesen, ansonsten zur Bearbeitung zurückgesendet (auch das stellt oft bloß eine höfliche Form der Abweisung dar, bzw. eine Möglichkeit zur Gesichtswahrung für jenen Herausgeber, der sich für das abgelehnte Manuskript stark gemacht hat).

Die Abweisungsraten sind z. T. von ganz banalen Faktoren abhängig. Auf die erwähnte Schere zwischen Angebot und Nachfrage finden sich unterschiedliche Antworten: Physikalische Zeitschriften reagierten stark expansiv, mit der Zunahme an Volumen und der Schaffung von Nebenreihen: 10 000 Seiten Jahresumfang sind hier keine Seltenheit. Im Vergleich dazu wirken die 200-300 Seiten pro Jahrgang psychologischer Journale recht bescheiden und - nicht zuletzt angesichts der kostengünstigen digitalen Möglichkeiten - eher wie künstliche Verknappung.

Der gängige Mythos lautet nun: je höher die Abweisungsrate, umso strenger sei das Journal, desto höhere wissenschaftliche Qualität hätten sohin Journal wie Publikation. Manche wissenschaftliche Zeitschriften geben daher schon Kürzungsforderungen oder die Verlegung in andere Journalsektionen, z.B. zu den Forschungsnotizen, als Abweisung aus. Im Konkurrenzkampf um AutorInnen, LeserInnen, Sponsoren fungieren sowohl solcherart *künstlich überhöhte Abweisungsraten* als auch der *„Peer-Review"-Stempel* nicht selten als *Prestigeschmuck*:

- Vielfach ist unklar, welche Zeitschriften überhaupt extern begutachtet werden und welche nicht - unterschiedliche offiziöse Zeitschriftenkataloge weisen nur teilweise Schnittmengen auf (vgl. Eldredge 1997).

- Meist bleibt gänzlich unklar, wie das Peer-Review konkret praktiziert wird, d.h. den Lesern werden Angaben dazu vorenthalten.

- Nicht wenige *behaupten*, auf Basis eines Peer-Review-Verfahren zu editieren, praktizieren es aber nicht oder kaum bzw. gewähren großzügige Ausnahmen, entweder informell oder auch formell geregelt (z. B. sind US-Akademiemitglieder von der Begutachtung in den akademieeigenen Publikationen ausgenommen).

- Viele Journale verhielten sich inkonsistent, so wurden nach eklatanten Fehlentscheidungen mitunter über Jahrzehnte bereits eingeführte Peer-Review-Prozeduren wieder ausgesetzt.

- Auch bei sauber begutachteten Journalen wird *nur ein Teil*, oft nur ein Bruchteil der Zeitschrift, *tatsächlich extern qualitativ evaluiert*. Doch alles dort sonstig Veröffentlichte schmückt sich mit diesem Nimbus, vgl. den US-Passivraucher-Skandal: Justiziare eines von hoher Klageforderung bedrohten US-Tabakkonzerns ließen unter dem Namen von Top-Medizinwissenschaftlern - gegen Entgelt für ebendiese versteht sich - Leserbriefe an Top-Medizin-Journale unterzeichnen, auf die sich die Konzernverteidigung vor Gericht dann erfolgreich berufen konnte. Da der Anteil „nicht zitierbarer" Beiträge den Impact Faktor eines Journals künstlich erhöht, werden solche Beiträge immer beliebter.[3]

- Viele Journale schmücken sich mit einem großen Pool an Gutachtern, de facto wird oft nachweislich nur ein winziger Bruchteil überhaupt eingesetzt: Einige wenige „old boys" bekommen fast alle Gutachten zugeschanzt und beherrschen so, da oft bei einer Vielzahl von Journalen gleichzeitig tätig, eine ganze Disziplin.

Allerdings gibt es etliche Möglichkeiten, Peer-Review-Verfahren im engeren Sinne zu *umgehen*. Zur Forschungsfinanzierung empfiehlt sich vor allem
- "*pork barrel*" (so werden politische Subventionen an Forschungseinrichtungen in den USA genannt, vgl. Savage 1999),
- Rüstungsforschung („classified", d. h. mit Geheimhaltungspflicht) und
- Industrieaufträge oder Zuwendungen privater Stifter. Gerade bei letzteren mangelt es z. T. an universalistischen wissenschaftlichen Kriterien, mitunter zählt bloße Gesinnung.[4]

Industrie- und Stiftungsgelder können in den privaten US-Universitäten, die uns laufend als Vorbild und Elite vorgehalten werden, auch als Ersatz für eine peer reviewed Literaturliste dienen: Das nicht nur in Stanford praktizierte Prinzip des *„salary splitting"* besagt, dass Professoren nicht nur Gelder für die Finanzierung ihrer MitarbeiterInnen, sondern auch des eigenen Postens mitbringen.[5] Auch bei uns gilt das Ausmaß akquirierter Drittmittel bereits als Kriterium für wissenschaftliche Leistung, zumindest von Instituten.

Für WissenschaftlerInnen mit nicht so üppiger finanzieller „Mitgift" lohnt es sich zumindest in den Naturwissenschaften, ihre Artikel auf Preprint-Server zu legen. Diese haben in der Physik bereits solche Reputation gewonnen, dass ihre mehrfach positiv zitierten Beiträge kaum mehr von Journalen abgelehnt werden können (während aufgrund der sogenannten „Ingelfinger-Rule" psychologische Journale bereits im WWW erschienene Artikel zur Veröffentlichung in der Regel ablehnen).

3 Schwere Leistungsmängel des Gutachter-"systems" im Konnex von Betrug und Täuschung

Das Vertrauen in die Leistungsfähigkeit der Peer-Review-Verfahren dürfte auch durch ihr offensichtliches Versagen im Konnex *von Plagiat, Betrug und Täuschung* geschwunden sein. Trotz plumper und offensichtlicher Manipulationen wurden bisher nur wenige der zahlreichen bekannt gewordenen Fälle devianten wissenschaftlichen Verhaltens durch anonyme Gutachter aufgedeckt. Wenn Gutachter ausnahmsweise eine Fälschung entdecken, wagen es die Herausgeber oft nicht einmal, den Fall publik zu machen bzw. zur Anzeige zu bringen, aufgrund der Angst vor eigener Rufschädigung und vor Klagedrohungen der Fälscher. Die bisher eruierten Fälle wurden in der Regel aufgrund von Interaktionsprozessen abseits der formalen Wissenschaftskommunikation enttarnt, vornehmlich über persönliche Denunziation bzw. über Selbstanzeigen von MitarbeiterInnen mit Gewissensbissen.

Das Verdienst der Aufdeckung wissenschaftlichen Betrugs und des Umschwungs des Verhaltens wissenschaftlicher Institutionen, aber auch wissenschaftlicher Gesellschaften (vgl. Fröhlich 2002a) gegenüber Plagiat und Fälschung ist zweifellos (neben zivilcouragierten WissenschaftlerInnen, die brisante Informationen weitergaben) zu Beginn *außer*wissenschaftlichen Instanzen (Presse, Politik) zuzuschreiben. Ohne außerwissenschaftliche Informationskanäle wäre wohl *keine einzige* der in „erster Generation" bekannt gewordenen Betrugsaffären je publik worden: Denn die wissenschaftlichen Institutionen (und die wissenschaftlichen Zeitschriften) waren anfangs nicht bereit, dieses Thema überhaupt nur zu erwähnen. Auch der bekannte Cyril-Burt-Skandal (Erfindung von Zwillingsstudien zur Intelligenz zwecks Stärkung der Vererbungslehre) konnte erst durch einen Presseartikel ins Rollen gebracht werden (vgl. Fröhlich 2001).

Auch bei den neueren bundesdeutschen Betrugsaffären aus Medizin und Biologie (Krebsforscherskandal Hermann/Brach/Mertelsmann; Max Planck Institut für Züchtungsforschung, Köln) passierten jeweils mehrere Dutzend gefälschte Publikationen anstandslos die Peer-Review-Prozeduren renommiertester Wissenschaftsjournale (inkl. *science* und *nature*), obwohl betrügerische Manipulationen z. T. ganz offensichtlich waren - so war z. B. ein und dieselbe Abbildung als Beleg für völlig unterschiedliche Messreihen aufgeführt worden, vgl. Finetti/Himmelrath 1999.

Einige Gutachter waren sogar selbst in Betrugsaffären verwickelt; Verfahren vor Kontrollkommissionen oder Gerichten sind anhängig. So missbrauchten die erwähnten Krebsforscher Friedhelm Hermann und Marion Brach auch ihre Stellung als Referees: Sie wiesen einen Projektantrag zurück und reichten ihn wortwörtlich unverändert bei derselben Stiftung selbst ein – mit Erfolg: die Kölner Thyssen-Stiftung bewilligte ihnen 260 000 DM Fördermittel (Finetti & Himmelrath 1999, 48). Generell ist anzunehmen, dass Gutachter sich vom Ideenfundus der Einreichenden anregen lassen und diese als 'eigene Ideen' nutzen. „Kryptamnesien", d. h. unbewusste Plagiate (nach Vergessen der Quelle wird Erinnertes als eigene neue Idee erfahren), sind bei höchst umtriebigen TopwissenschaftlerInnen und MultifunktionärInnen fast unvermeidlich.

4 Peer-Review-Forschung nach Forschungsdesigns

Auch zahlreiche negative Befunde empirischer Untersuchungen zum Peer-Review-"System" ramponieren dessen Ruf. Beispiele hierfür sollen im Folgenden in aller Kürze vorgestellt werden.

4.1 "Experimentelle" Peer-Review-Studien

In "experimentellen " Peer-Review-Studien werden z. B. in Manuskripte Fehler eingebaut oder (bei Konstanthaltung sonstigen Manuskriptteile) die Forschungsresultate systematisch variiert, um Referee-Vorurteile zu decouvrieren. Mitunter werden sogar bereits abgedruckte Manuskripte bei denselben Journalen wiedereingereicht :

- Peters/Ceci 1982 hatten zwölf psychologischen Fachzeitschriften je einen Aufsatz entnommen und dieses bereits veröffentlichte Manuskript (nach Änderung des Verfassernamens und dessen institutioneller Affiliation sowie kleinen 'kosmetischen' Veränderungen) bei denselben Zeitschriften erneut zur Publikation eingereicht. Nur 3 von 12 Herausgebern bzw. Redaktionskollegien - genauer: *nur 3 von 38 Herausgebern bzw. Referees - erkannten die* in ihren Journalen vor 1 1/2 bis 3 Jahren bereits publizierten *Aufsätze wieder*. Die übrigen erneut eingereichten Aufsätze wurden erneut der üblichen Bewertungsprozedur ausgesetzt. Das Ergebnis: *Nur ein einziger Beitrag wurde angenommen.* 8 von 9 bereits publizierten Beiträgen wurden aufgrund negativer Gutachten abgelehnt, v. a. mit dem Hinweis auf „schwerwiegende methodologische Mängel".

- Abramowitz et al. 1975 konstatieren einen eindeutigen *politischen Gutachter-Bias* unter PsychologInnen (n=322). Übereinstimmung in der politischen Einstellung zwischen eingereichtem Beitrag und dem Referee und dessen Fehlererkennung korrelieren negativ: Lehnt der Gutachter die politische Einstellung des Beitrags ab, findet er wesentlich mehr Fehler und kommt zu einem wesentlich negativeren Urteil als bei politischer Übereinstimmung.

- Baxt, W. G. et al. 1998 erstellten ein fiktives Manuskript (eine placebokontrollierte Doppelblind-Studie zu einer Migränetherapie), in das sie 10 schwerere and 13 leichtere *Fehler* plazierten. Die Studie wurde an alle Referees der Annals of Emergency Medicine (n=262) mit Ersuchen um Review übersendet. 78% der GutachterInnen sendeten eine Beurteilung. Jene 15 Referees, die das Manuskript zur Publikation empfahlen, identifizierten 17% der schwereren und 12% der leichteren Fehler. Die 117 ablehnenden GutachterInnen erkannten zwar mehr, aber auch nur 39% der schwereren und 25% der leichteren Fehler. Die 67 Gutachter, die eine Überarbeitung empfahlen, fanden 30% der schwereren und 22% der kleineren Fehler: „Sixty-eight percent of the reviewers did not realize that the conclusions of the work were not supported by the results... Peer reviewers in this study failed to identify two thirds of the major errors in such a manuscript." (ebd., 310)

4.2 Rekonstruktionen bzw. Replikationen von Begutachtungen

In rekonstruktiven Untersuchungen werden abgeschlossene Projekt- oder Manuskriptbegutachtungen neuerlich aufgerollt. Z. B. werden sowohl bewilligte / akzeptierte als auch abgelehnte Projektanträge oder Manuskripte ein zweites Mal von GutachterInnen aus demselben Referee-Pool begutachtet.

- Je 50 abgelehnte und akzeptierte Projektanträge an die National Science Foundation (NSF) wurden noch einmal von NSF-Gutachtern bewertet. Cole et al. 1981 kommen zum Ergebnis: Das Resultat der Gutachterbewertung und der Annahme oder Ablehnung des Antrags sei „zufällig".

- Eine statistisch aufwendige schwedische Untersuchung (Wenneras /Wold 1997) erbrachte den Nachweis für *Nepotismus und Sexismus* in den Vergabepraktiken des Swedish Medical Research Council (MCR) bei ihren „postdoctoral fellowhip applications". Den Zugang zu den Untersuchungsdaten, d. h. zu den Unterlagen der Einreichenden und ihrer Beurteilungen hatten die Forscherinnen aufgrund des schwedischen „Freedom of Press Act" gerichtlich erkämpft - der MRC hatte ihre Herausgabe verweigert (in fast allen anderen Staaten wäre eine solche Untersuchung mithin kaum möglich gewesen). Die ernüchternden Resultate*: Frauen mussten 2,5 mal so produktiv* sein wie Männer, um von den MRC-Referees eine gleich gute Bewertung wie ihre männlichen Kollegen zu erhalten. Auch ein starker Freundschaftsbonus wurde nachgewiesen. Die *persönliche Beziehung zu einem Kommissionsmitglied* erbrachte wesentliche Vorteile bei der Bewertung, d. h. eine enorme Ersparnis bei den nachzuweisenden Publikationen: „an applicant lacking personal ties with the reviewers needed to have 67 more impact points than an applicant of the same sex who was associated with one of the reviewers, to be perceived as equally competent. So, although MRC policy does not allow 'biased' reviewers to participate in the scoring of applicants they are associated with, this rule was insufficient, as the 'neutral' committee members compensated by raising their source when judging applications affiliated with one of their peers." (ebd., 342)

- Mehrere Studien berichten auch von einem *„language bias"* des Peer-Review-Systems (z. B. Herrera 1999), d. h. AutorInnen ohne Englisch als Muttersprache werden benachteiligt.

- Die Strenge der Referee-Urteile hängt nach der empirischen Untersuchung von Nylenna (1994) vom *Lebensalter* der Gutachter ab. Als Faustregel gelte: je jünger, desto rigider. Die Rigidität junger Gutachter wird gewöhnlich mit ihrer ehrgeizig-panischen Angst, einen Fehler zu übersehen, erklärt, und mit ihrem Drang, gegenüber den Zeitschriftenherausgebern einen guten Eindruck zu machen.

- Die Beziehungen zwischen Herausgebern und Referees scheinen eher von *Nicht-Kommunikation und gleichgültigem Desinteresse* an den referierten Arbeiten geprägt zu sein: lt. Glogoff 1988 hat die Hälfte der Referees (n=121 von 31 wissenschaftlichen Zeitschriften aus dem Bibliothekswesen) keinerlei formale Bewertungskriterien, sind mehrheitlich nicht informiert, ob ihre Kommentare von den Herausgebern an die Autoren weitergeleitet werden, ungefähr drei Viertel sind nicht informiert über das endgültige „Schicksal" der von ihnen referierten Manuskripte.

- Auch die Qualität veröffentlichter Artikel, also jener Artikel, welche die Hürden der Begutachtungsfahren überwunden haben, wird in einer Reihe von inhaltsanalytischen Studien auf der Basis größerer Fallzahlen ei-

ner kritischen Prüfung unterzogen. Meinefeld 1985 moniert bei einer Untersuchung von Artikeln in angesehenen deutschsprachigen soziologischen Journalen, eine kritische intersubjektive Überprüfung der Journalaufsätze sei nicht möglich. *Relevante Angaben fehlen demnach in fast allen überprüften Beiträgen,* die Gutachter dürften also darauf keinerlei Wert gelegt haben. Die häufig vorgebrachte Entschuldigung der „Platznot" sei bloß Ausrede, da für die relevanten Angaben (z. B. Verweigerungsraten) kleingedruckt nur wenig Platz benötigt werde.

4.3 Befragungen Begutachteter

Erhebungen zu den Einstellungen betroffener, aktiver WissenschaftlerInnen ergeben *primär negative Bewertungen* der Gutachterverfahren, vor allem hinsichtlich ihrer Rationalität, Effizienz, Objektivität und Innovationsfreudigkeit. Die negativen Urteile beziehen sich sowohl auf Herausgeberentscheidungen als auch auf das Gutachterwesen insgesamt:

- Bradley 1981 befragte 672 UniversitätsprofessorInnen über ihre Erfahrungen mit Gutachtern und Gutachten. Sie beklagten *Konformitätsdruck* bezogen auf Idiosynkrasien (d. h. auf höchst eigenwillige bis eigenbrötlerische Wünsche) der Gutachter, beanstandeten unzutreffende Kritik, mangelnde fachliche Kompetenzen der Gutachter, die Konzentration der Referees auf Trivialitäten, eine inferiore Behandlung der eigenen Person durch die Gutachter, achtlose, unaufmerksame Lektüre der Manuskripte durch die „Peers". Die Gutachten schienen eher verfasst worden zu sein, um den Zeitschriftenherausgeber zu beeindrucken, als um sachliche Kritik zu üben.

- Laufend werden in der Literatur Gutachter *als Gegner theoretischer wie methodischer Innovation* charakterisiert. So zieht Armstrong (1997) in einer umfassenden Überblicksarbeit (seine Literaturbasis: 68 Aufsätze, publiziert seit 1975) folgende Bilanz der empirischen Evidenz zum Journal-Peer-Review: „Current procedures .. seem to discourage scientific advancement, especially important innovations, because findings that conflict with current beliefs are often judged to have defects." Ähnlich konstatiert Yamazaki 1992, 1995 (Japan, Pharma-Forschung) die Gefahr des Konservativismus.

Die Kritiker der (meist einseitigen) Anonymität des Referee-Systems befürchten, dass diese Mechanismen unter dem schützenden Mantel (oft einseitiger) Anonymität greifen: die Gutachter wüssten zwar um die Identität der Einreichenden, jedoch nicht umgekehrt. Letzteren werde vielfach sogar der Text der Gutachten vorenthalten, sie könnten daher nicht einmal Fehler und Irrtümer der GutachterInnen richtigstellen.

Letztlich können die Klagen der Betroffenen auf eine Kurzformel gebracht werden: *Das* (dem Gutachter aufgrund seiner eigenen Kompetenzen, seiner eigenen wissenschaftlichen Entwicklungsgeschichte) *Bekannte wird gefördert, das weniger Bekannte behindert*. Wie vielfach in Literatur und informeller Kommunikation beklagt wird, hätten bekannte, eingefahrene Themen, Modelle und Theorien bei der Begutachtung gute Chancen, da diese den Gutachtern in der Regel geläufig oder zumindest bekannt seien. Bei innovativen bzw. dem jeweiligen Mainstream weniger ausdrücklich verpflichteten Ansätzen sei hingegen die Gefahr äußerst groß, dass eingereichte Anträge oder Manuskripte von den Gutachtern - aufgrund eigener Unkenntnis - abgewertet und abgelehnt würden. Als Standard-Killer-Argument diene in diesem Konnex gewöhnlich die stereotype Formulierung, dass der Bezug zur aktuellen Diskussion, zum derzeitigen Stand der Forschung (der immer mit dem Kenntnisstand des Gutachters gleichgesetzt werde) nicht gegeben und sohin auf Basis des verwendeten Ansatzes kein wesentlicher Erkenntnisfortschritt zu erwarten sei. Auch die mehrfach empirisch belegte schlechtere Behandlung, d. h. die *höheren Abweisungsraten* von Projekten bzw. Manuskripten, welche *interdisziplinäre* Konzepte und Zugänge verfolgen, lässt sich hier einreihen.[6]

Peer-Review-Hardliner könnten kontern: negative Bewertungen kämen von enttäuschten, gescheiterten WissenschaftlerInnen. Doch bundesdeutsche (Mittelbau-)Angehörige führen gerade ihre bisherigen Veröffentlichungserfolge eher auf „Glück und soziale Beziehungen" als auf Fähigkeit und Anstrengung zurück (Borkenau/Kammer 1988). Ähnlich berichteten die von Harriet Zuckermann 1977 untersuchten Nobelpreisträger von ungerechtfertigter Überhäufung mit Belohnungen und Bevorteilungen.

4.4 Übereinstimmungen von Gutachterurteilen untereinander / mit Zitationshäufigkeiten

Vielfach wird auch die *mangelnde Übereinstimmung* der Gutachter untereinander kritisiert bzw. nachgewiesen. So fokussiert z. B. Cicchetti 1997 die geringe Reliabilität der Peer-Review-Prozesse: Oft müssten die Herausgeber Entscheidungen auf der Basis konfligierender Gutachterempfehlungen fällen. Die wenigen offensiven Verteidiger des etablierten Review-Systems deklarieren hingegen Korrelationen von 0,2 bis 0,3 (vgl. Lindsey 1991) als hoch oder interpretieren die mangelnde Übereinstimmung zwischen den Referee-Urteilen als Beweis für die paradigmatische Vielfalt des Gutachterpools der jeweiligen Zeitschrift. Sie implizieren allerdings damit, ob sie das wollen oder nicht, das Fehlen paradigmenübergreifender wissenschaftlicher Qualitätskriterien.

Verteidiger des Status Quo vergleichen auch Bewertungen von Referees mit gleichlaufenden *Zitationshäufigkeiten*. Doch diese, höchst umstritten, sind für genau denselben Bias anfällig: Bereits etablierte Theorien, Methoden, WissenschaftlerInnen, Institutionen, werden durch „Impact"-Werte belohnt (z. B. durch Aufnahme in Datenbanken), ihre Resonanz mithin weiter verstärkt, innovative, noch statusniedrige Außenseiter hingegen bestraft (vgl. Fröhlich 1999a).

5 Der Widerstand gegen die eigene Objektivierung

Doch die Gutachter dienen, trotz aller negativer Befunde zur ihren Performanzen, z. T. zu unrecht als Prügelknaben. Viel zu wenig im Blicklicht und in der Kritik: Die Arkanpraxis, die Willkür der Herausgeber, die mitunter ihr Journal als diktatorischen Einmannbetrieb führen. Kaum untersucht: Wie wird man Herausgeber, wie verlaufen die Entscheidungsprozesse unter den Herausgebern? In paradigmenschwachen Fächern entscheidet bereits die Vergabe an Gutachter bekannter Schulrichtung über das Schicksal eines Manuskripts. Auch zu den Begutachtungsverfahren und Entscheidungsprozessen bei den Einreichungen von Projektvorschlägen bzw. Förderungsansuchen bei EU, Ministerien, Akademien und privaten Stiftungen finden sich zwar in Literatur und Internet-Listen zahlreiche (und z. T. recht negative) anekdotische Informationen, aber kaum solide kritische empirische Untersuchungen.

Herausgebern, Forschungsförderern, Gutachtern ist negativ anzurechnen, dass sie ganz offensichtlich die zahllosen einschlägigen empirischen Befunde der Wissenschaftsforschung (mit leider großteils negativen Resultaten) nicht kennen, jedenfalls ignorieren - Objektivierer leisten bekanntlich oft größten Widerstand gegen die *eigene* Objektivierung. Es erfolgt auch keine Information, geschweige denn eine Schulung der Gutachter, es wird keine systematische Förderung von Gutachterkompetenz betrieben - anscheinend wird diese als eine Art Natur-

gabe angesehen: in einer engen Spezialdisziplin emporgehantelt, sollen Wissenschaftler auf einmal alles und jedes objektiv bewerten können.

Abschließend sei noch einmal betont, dass sich in den einzelnen wissenschaftlichen Disziplinen *höchst unterschiedliche „Betriebssitten"* finden, und auch innerhalb der Disziplinen unterscheiden sich Journale, Verlage, Forschungsförderungsorganisationen etc. in ihren Usancen beträchtlich. Nur zwei Beispiele: Es ist in einigen Disziplinen (z. B. der Nationalökonomie) eher üblich, dass AutorInnen Einsicht in die Gutachtenstexte nehmen können, in anderen wiederum nicht. Eine Anonymisierung der AutorInnen wiederum, vielfach als Reform vorgeschlagen, ist in weiten Teilen der Physik bzw. von „Big Science" kaum möglich, da hier in einzelnen Teilgebieten oft weltweit ganz wenige Forschergruppen miteinander konkurrieren bzw. kooperieren, d. h. bereits die verwendeten Großapparaturen bzw. Verfahren die Forschergruppen hinreichend identifizieren. Diese Einschränkungen sollten bei den folgenden Reformvorschlägen mitbedacht werden.

6 Reformvorschläge

Die Reformvorschläge in Literatur und Internet-Listen streben *mehr Transparenz* an: Sie pochen auf Einsichtnahme in die vollständigen Gutachtentexte durch die Begutachteten, fordern verbindliche, offengelegte Kriterienlisten für Gutachter, seltener die regelmäßige Begutachtung der Referees selbst.

Radikalere KritikerInnen verlangen statt der bisherigen Arkanpraxis ein *offenes System*, inkl. der Lagerung von Manuskripten, Gutachten, Stellungnahmen und Kritiken auf öffentlich zugänglichen WWW-Servern.[7] Auf jeden Fall solle die Anonymität der Gutachter aufgehoben werden, sie beeinträchtige offene Diskussion und Kritik, aus Angst vor der Rache der Kritisierten im Schutze der Anonymität.

Andere hingegen schlagen die allgemeine Durchsetzung der *Doppelblindbegutachtung* vor (d. h. auch die Identität der Begutachteten wird während des Verfahrens geheimgehalten; vgl. z. B. Peters/Ceci 1982 und Kommentare im selben Heft). Im Vergleich zu Einfachblindverfahren schneiden Doppelblindverfahren i. d. R. besser ab, auch die Bevorteilung von WissenschaftlerInnen aus statushöheren („near-top-ranked") Institutionen bzw. die Benachteiligung von WissenschaftlerInnen aus nicht-akademischen Institutionen verringert sich (vgl. z. B. Blank 1967).

Zumindest ein *Gedankenexperiment* wert wäre auch ein wissenschaftliches Veröffentlichungssystem *ohne Peer Review,* d.h. ohne offizieller Zensur, am Beginn des Publikationsprozesses:

- Gerade das Vertrauen in die „qualitätsgeprüften" Publikationen führt zum hemmungslosen Zitieren und Übernehmen. Es heizt so die Publikationsflut durch rasch und flüchtig erstellte Veröffentlichungen an. Zur Zeit dominiert offensichtlich diese entlastende, weil komplexitätsreduzierende (Luhmann) Strategie des Vertrauens. Wir wissen aus etlichen Studien, dass AutorInnen die zitierte Literatur vielfach nie im Original eingesehen haben - sie vertrauen darauf, dass sie ohnehin von einigen kompetenten „gate keepers" kontrolliert wurde.

- Ohne Peer-Review-Sicherheit wäre sohin die Eigenverantwortlichkeit der Autoren hinsichtlich der kritischen Überprüfung der zu zitierenden Arbeiten wieder stärker gefordert. Aufgrund des größeren Aufwands bei Kontrolle und Bewertung der Zitate ohne Peer-Review-Etikett könnte sich die völlig 'überhitzte' Publikationsmaschinerie wieder etwas 'abkühlen'. Dass sich ohne Peer Review das Journalwesen oder der wissenschaftliche Sektor des Internet (etwa: die Internet-Aktivitäten unter der Patronanz wissenschaftlicher Gesellschaften) zu einem Tummelplatz von Dilettanten entwickeln würden, wäre nicht zu erwarten. Konsequent digitale wissenschaftlicher Kommunikation würde über Suchmaschinen bzw. Datenbank-Links Auffindbarkeit im Volltext und kritische Kommentierung durch LeserInnen erleichtern. Dann wäre die Gefahr öffentlicher Blamage wesentlich höher als bisher - zur Zeit bleiben allfällige Peinlichkeiten hinter den Kulissen der Institutionen und Journale verborgen.

Allerdings würde sich damit das vorgeschaltete und wichtigere Problem der Mittelverteilung bei der Forschungsförderung nicht lösen. Funktionstüchtigkeit und Prestige der physikalischen Preprint-Server dürften auch damit zusammenhängen, dass in Big-Science-Disziplinen bereits die Zugehörigkeit zu einer Projektgruppe, der Zugang zu aufwendigen Apparaturen bzw. die Finanzierung der Forschung eine hohe Selektionshürde darstellt. In den textdominierten Wissenschaften reichen noch heute vielfach als Forschungsmittel im Prinzip Papier und Bleistift (neben dem Zugang zu einer wissenschaftlichen Bibliothek).

Wenn am Peer Review festgehalten werden soll, dann wären Reformen dringend empfehlenswert:

- In der BRD werden Zivilprozesse den Richtern vom Computer per Zufall zugewiesen, um eine Verzerrung durch willkürliche Zuordnung zu verhindern. Dies sollte auch auf das Gutachterwesen übertragen werden. Die Zuteilung der Gutachter sollte *per Zufall* erfolgen, d. h. der gesamten Gutachterpool sollte systematisch ausgeschöpft werden - unter Ausschaltung von Willkür und der Machtkonzentration bei einigen wenigen „old boys". Bei paradigmenschwachen Fächern könnte dies nach Paradigmen geschichtet erfolgen, mit je einem paradigmeneigenen und einem fremdparadigmatischen Gutachter Minimum.

- Die Beschleunigung der Kommunikation zwischen AutorInnen, Redaktionen, Referees etc. per Internet sollte zur Selbstverständlichkeit werden.

- Am besten wäre eine Dreifachblindbegutachtung (nicht nur den Gutachtern, auch den Herausgebern bzw. Forschungsförderern sollte die Identität der Autoren bzw. Antragsteller vorenthalten werden).

- Den Begutachteten wäre der Inhalt sämtlicher Gutachten und Stellungnahmen in voller Länge mitzuteilen.

- Ein Berufungsrecht der Einreichenden sollte ebenfalls zum Standard werden. Vor allem muss ihnen die Möglichkeit eingeräumt werden, eindeutige Fehler und Missverständnisse der Gutachter zu korrigieren.

- Nach erfolgreich (oder auch negativ, zumindest bei Zustimmung der Einreichenden) absolviertem Prozess sollte alle Äußerungen, Dokumente und Prozesse *offengelegt* werden, auch die Kritik der Gutachter und Herausgeber, damit die gesamte scientific community von der Kritik profitieren kann. Der potentiell unendliche digitale Raum macht dies möglich.

- Bisher ist die Refereetätigkeit i. d. R. ehrenamtlich; dies ist sicherlich eine Ursache für die des öfteren beklagte Schwierigkeit, Gutachter zu finden, und für den empirisch belegten geringen (Zeit-)Aufwand bei der Erstellung der Gutachten. Die Tätigkeit des Peer Reviewing sollte daher angemessen honoriert werden.

- Die Herausgeber- und Gutachter-*kompetenzen* sollten systematisch gefördert werden. Dies sollte allerspätestens in den Graduiertenkollegs bzw. Doktorratsstudiengängen beginnen. Die kritischen Befunde der Wissenschaftsforschung sollten zur Kenntnis gebracht und zum Anlass von Reformen genommen werden. Generell sollte

bereits in der Lehre, aber auch durch das wissenschaftliche Belohnungssystem Kritikfähigkeit und -praxis systematisch gefördert werden. Dabei wären auch die wissenschaftstheoretische und forschungsethische Sensibilität und Reflexionsfähigkeit zu steigern.

- Die bisherige Praxis der Editoren und Referees wirkt *recht altertümlich-handwerklich* - z.T. mit fatalen Folgen. Jede einfache Datenbankrecherche hätte z. B. bereits anhand der einige Minuten erfordernden Lektüre der Abstracts ergeben, dass Zwillingsforscher Burt in drei Publikationen mit angeblich unterschiedlichen Stichproben für ein- und zweieiige Zwillinge stets jeweils genau dieselben Korrelationskoeffizienten eingesetzt hatte - auf drei Kommastellen genau (vgl. Fröhlich 2001 und dort zitierte Literatur). Die Identität der Visualisierungen in verschiedenen Artikeln und auf angeblich unterschiedliche Untersuchungen bezogen wurde erst im Laufe jahrelanger Nachforschungen zu neueren Betrugsaffären entdeckt. Diese hätten bei Volltext-Konsultierung bisheriger Artikel im Rahmen des Refereeing rasch und relativ aufwandslos entdeckt werden können. Auch Plagiatsüberprüfungsprogramme oder die simplen „related documents"-Funktionen (weisen Dokumente mit identer oder ähnlicher Zitationsstruktur nach) könnten schon lange eingesetzt werden. Die „Modernisierung" der altmodischen Überprüfungspraktiken der Herausgeber- bzw. Referees sollte daher ein vordringliches Reformanliegen sein, die Kompetenzen von Herausgebern, Redaktionen, Referees im Umgang mit Datenbanken bzw. Methoden der Informationswissenschaft müssten dringend entwickelt werden. Zumindest müssten entsprechend qualifizierte InformationswissenschaftlerInnen und WissenschaftsforscherInnen beigezogen werden.

- Die rechtlichen Voraussetzungen für eine Offenlegung der Akten von Stiftungen und sonstigen Forschungsförderern für empirische Forschungen sollten verbessert bzw. erst geschaffen werden.

Ich schlage zudem vor, dem *Zufall* noch mehr Raum zu geben:

- Finanzämter fordern bei der Bearbeitung der Einkommensteuererklärungen nicht von vornherein alle Belege an. Die Steuerpflichtigen müssen aber damit rechnen, dass ihr Fall aufgerollt wird und sie sämtliche Unterlagen vorlegen müssen. Es würde analog dazu die Kontrollmöglichkeiten von Schlamperei und Betrug verbessern, wenn Projekt- bzw. Manuskripteinreicher einkalkulieren müssten, mit einem Risiko von 5-10% zufällig ausgewählt zu werden, ihre Rohdaten und Labor-Tagebücher vorzuweisen (bisher sind diese lt. Nachfolgestudien meist „verlegt, verbrannt, verloren" gegangen, vgl. z. B. Wolins 1962). Sorgfalt und Qualität der Gutachten könnten gesteigert werden, wenn die Referees wüssten, dass mit einer Wahrscheinlichkeit von 5-10% ein weiterer Referee ihre Stellungnahmen begutachten würde, und dies auch negative Konsequenzen haben könnte. *Systematische stichprobenartige Kontrollen* würde vor allem jenen Verteidigern des bisherigen Systems den Wind aus den Segeln nehmen, die Reformen mit dem Hinweis auf den unermesslich steigenden Aufwand an Zeit und Kosten ablehnen.

- Aus der Klage, das Peer-Review-System sei ein „Lotteriespiel", könnte man eine Tugend machen: Warum nicht einen kleineren Teil der Forschungsförderungen tatsächlich in Form einer *Innovations-Lotterie* nach Losentscheid ausschütten?

Zu teuer? Die AutorInnen bzw. die Institutionen, die ihre Evaluationsprobleme auf den Publikationssektor abwälzen (dessen Kritik- und Kommunikationsfunktionen darunter leiden), sollen die Kosten hierfür übernehmen. Bei Umstellung auf digitale wissenschaftliche Kommunikation würden sich die sonstigen Kosten ohnehin drastisch senken. Da die Publikation weniger den LeserInnen als den Autoren und ihren Institutionen dient (die interessierten WissenschaftlerInnen besorgen sich die tatsächlich relevanten Informationen ohnehin direkt, in Form informeller Tauschhändel), sollten die eigentlichen Nutznießer auch dafür zahlen. Nicht zuletzt belohnt das derzeitige System ja die hartnäckig (bei anderen Journalen) kostenlos immer wieder einreichenden SchreiberInnen: Je „schlechter" ein Aufsatz, umso mehr Leser (in Form von Herausgebern und Referees) findet er also.

Geld für „Evaluation" wird ohnehin ausgegeben. Private Marktforschungsfirmen reißen z. B. ministeriell beauftragt Hörsaaltüren auf und schätzen mit einem Blick die vorhandene StudentInnenmasse ab (die nicht selten aufgrund von Raumverlegungen zu einer ganz anderen Lehrveranstaltung als offiziell verzeichneten gehört) oder befragen angebliche „Opinion Leader", welchen Wissenschaften sie universitäre Existenzrechte zubilligen. Die emsige Betriebsamkeit der sogenannten evaluativen Szientometrie (v. a. der einzelne Themen, Disziplinen und Journaltypen systematisch begünstigende 2-Jahres-Impact Faktor des ISI) führt zu Artefakten und kurzatmig-kurzsichtiger Salamipublikationstaktik (vgl. Fröhlich 1999a).

Im Vergleich zu solchen Vorgangsweisen ist Kritik, d.h. die möglichst kompetente *qualitative Bewertung geistiger Leistungen,* notwendig und sinnvoll. Eigentlich sollte sie alltägliche Praxis *aller* WissenschaftlerInnen[8] darstellen und nicht nur an einige wenige delegiert werden. Bei allem Evaluations- und Reformeifer sollten wir nicht den in der bisherigen Wissenschaftsgeschichte so oft nachgewiesenen *Widerstand* wissenschaftlicher Establishments *gegen Innovationen* vergessen. Ein technokratisch durchgestyltes, lükkenloses Bewertungs- und damit Hierarchiesystem dürfte fatale Folgen haben: Nicht nur in der Landwirtschaft haben sich Monokulturen als recht anfällig erwiesen. *Offenheit und Pluralismus* der wissenschaftlichen Felder sollten daher oberstes Gebot bleiben.

Literatur

Abramowitz, S. I. et al. (1975): Publish or politic: Referee bias in manuscript review. Journal of Applied Social Psychology 5 (3), 187-200

Armstrong, J. S. (1997): Peer Review for Journals: Evidence on Quality Control, Fairness, and Innovation. Science and Engineering Ethics 3 (1), 63-84

Baxt, W. G. et al. (1998): Who reviews the reviewers? Feasibility of using a fictitious manuscript to evaluate peer reviewer performance. Annals of emergency medicine 32 (3, 1), 310-317

Blank, R. M. (1967): The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from the American Economic Review. American Economic Review 8, 1041-1967

Borkenau, P./Kammer, D. (1988): Publizieren oder resignieren? Subjektive berufliche Perspektiven und Bewältigungsstrategien des Mittelbaus der Universität Bielefeld. Zeitschrift für Soziologie 17 (1), 72-79

Bradley, J. V. (1981): Pernicious publication practices. Bulletin of the Psychonomic Society 18 (1), 31-34

Cicchetti, D. V. (1997): Referees, Editors, and Publication Practices: Improving the Reliability - and Usefulness of the Peer Review System. Science and Engineering Ethics 3 (1), 51-62

Cole, S. et al. (1981): Chance and Consensus in Peer Review. Science 214, 881-886

Eldredge, J. D. (1997): Identifying peer-reviewed journals in clinical medicine. Bulletin of the Medical Library Association 85 (4); 418-422

Finetti, M./Himmelrath, A. (1999): Der Sündenfall. Betrug und Fälschung in der deutschen Wissenschaft. Stgt. etc.

Fröhlich, G. (1998): Optimale Informationsvorenthaltung als Strategem wissenschaftlicher Kommunikation, in: Harald H. Zimmermann / Volker

Schramm (Hg.), Knowledge Management und Kommunikationssysteme. Konstanz, 535-549. Online: <http://www.agmb.de/mbi/8/mb8.pdf>

Fröhlich, G. (1999a): Das Messen des leicht Messbaren. Output-Indikatoren, Impact-Maße: Artefakte der Szientometrie?, in: Becker, J./ Goehring, W. (Hg.): Kommunikation statt Markt. Gesellschaft für Mathematik und Datenverarbeitung, Sankt Augustin, GMD Report 61, online s.a. <http://info.uibk.ac.at/sci-org/voeb/om65.html#gf>

Fröhlich, G. (1999b): Von den Tempelwächtern der Wissenschaft. Der Standard - Album Wissenschaft (bm:wv; FWF) 23. 10. 1999, W1

Fröhlich, G. (2000): Wissenschaftsforschung: Die theoretisch-empirische Erforschung der „Medien und Institutionen des Wissens", in: AG Kulturwissenschaften Graz (Hg.): Kulturwissenschaften in Österreich. Graz, 28 - 31

Fröhlich, G. (2001): Betrug und Täuschung in den Sozial- und Kulturwissenschaften, in: Hug, T. (Hg.), Wie kommt die Wissenschaft zu ihrem Wissen? Hohengehren/Baltmannsweiler: Bd. 4/CD-ROM 2: Einführung in die Wissenschaftstheorie und Wissenschaftsforschung der Sozial- und Kulturwissenschaften, 261-273

Fröhlich, G. (2002a): verein.wissenschaft: Entstehung und Funktionen wissenschaftlicher Gesellschaften, in: Ulrike Kammerhofer-Aggermann (Hg.): Ehrenamt und Leidenschaft. Salzburg, 255-278

Fröhlich, G. (2002b): Peer Review: Contra. Forschung und Lehre 9 (2), 313

Fröhlich, G. (2002c): „Einschlagende Neuigkeiten. Der „Impact Factor" entscheidet über das Gewicht von Artikeln und ForscherInnen. Der Standard, Wissenschaft, 9.4. 2002, online: <http://DerStandard.at/?id=917598>

Glogoff, S. (1988): Reviewing the Gatekeepers: A Survey of Referees of Library Journals. American Society for Information Science Journal 39 (6), 400-407

Harnad, S. (1996) Implementing Peer Review on the Net: Scientific Quality Control in Scholarly Electronic Journals, in: Peek, R. / Newby, G. (Eds.) Scholarly Publishing: The Electronic Frontier. Cambridge / Mass., 103-118

Herrera, A. J. (1999): Language bias discredits the peer-review system. Nature 397 (6719), 467

Jasanoff, S. (1985): Peer Review In The Regulatory Process. Science, Technology and Human Values 10, 20-32

Lindsey, D. (1991): Precision in the Manuscript Review Process: Hargens and Herting Revisited. Scientometrics 22 (2), 313-325

Lowen, R. S. (1977): Creating the Cold War University. The Transformation of Standford. Berkeley etc.

Meinefeld, W. (1985): Die Rezeption empirischer Forschungsergebnisse - eine Frage von Treu und Glaube? Resultate einer Analyse von Zeitschriftenartikeln. Zeitschrift für Soziologie 14 (4), 297-314

Munthe, C/Welin, S. (1996): The Morality of Scientific Openness. Science and Engineering Ethics 2 (4), 411-428

Nylenna, M. (1994): Multiple blinded reviews of the same two manuscripts: Effects of referee characteristics and publication language. JAMA - Journal of the American Medical Association 272 (2), 149-151

Peters, D. P./Ceci, S. J. (1982): Peer review practices of psychological journals: The fate of published articles, submitted again. Behavioral and Brain Science 5, 187-195

Savage, J. D. (1999): Funding science in America: Congress, universities, and the politics of the academic pork barrel. Cambridge etc.

Yamazaki, S. (1992): Production and transfer of information: the increasing volume of information and the referee system. Pharmaceutical Library Bulletin 37 (3), 187-90

Yamazaki, S. (1995): Refereeing system of 29 life science journals preferred by Japanese scientists. Scientometrics 33 (1), 123-9

Peters, J. (1996): The Hundred Years War started today: an exploration of electronic peer review. Management Decision 34 (1), 54-59

Wenneras, C. /Wold, A. (1997): Nepotism and sexism in peer-review. Nature 387. 341-343

Wolins, L. (1962): Responsibility for Raw Data. American Psychologist 17, 657-58

Zuckermann, H. (1977): Scientific Elite. Nobel Laureates in the United States. N. Y./London

****

* Einige Teile dieses Artikels wurden bereits gekürzt veröffentlicht in Fröhlich 1999b, 2002b. Dank für kritisches Gegenlesen an Simone Griesmayr, für wertvolle Hinweise an Michael Strähle.

[1] Die Wissenschaftsforschung ist die (Selbst-)Anwendung wissenschaftlicher Methoden auf WissenschaftlerInnen bzw. wissenschaftliche Institutionen und ihre Praxis (vgl. Fröhlich 2000). Jene Sektoren der Informationswissenschaft, die sich mit wissenschaftlicher Kommunikation und Evaluation wissenschaftlicher Leistungen befassen, sind der Wissenschaftsforschung zuzurechnen.

[2] Bei (Einfach-)Blindverfahren (am häufigsten praktiziert) kennen die Begutachteten nicht die Identität der Gutachter. Bei Doppelblindverfahren wird auch den Gutachtern die Identität der AutorInnen bzw. AntragstellerInnen vorenthalten, zumindest wird dies versucht. Bei Dreifachblindverfahren wird zusätzlich den Herausgebern die Identität der AutorInnen vorenthalten. Dieses Verfahren wird leider selten praktiziert, z. B. von der Zeitschrift für Soziologie, Bielefeld. Dort sollen auch die Manuskripte von Starautoren abgelehnt worden sein.

[3] Das marktbeherrschende Institute for Scientific Information (ISI) definiert seinen Impact Faktor folgendermaßen: Es zählt alle Zitate, die ein Zeitschriftenjahrgang in den ersten zwei Jahren nach Erscheinen in den in den ISI-Datenbanken enthaltenen Journalen (einem Bruchteil des weltweiten Outputs an wissenschaftlichen Publikationen) erbringt. Diese Gesamtsumme wird dividiert durch die Zahl der „citable" (zitierfähigen) Artikel - und nur durch diese. Bevorteilt werden so allgemeine Journale wie „Science" mit vielen „non-citable" Editorials, Briefen, journalistischen Beiträgen, die z. T. noch dazu von bekanntesten AutorInnen stammen. Denn diese Beiträge werden von den ISI-Indexierern als nicht zitierfähig eingestuft und bringen dennoch viele Zitate für die Gesamtsumme ein, erhöhen also den ISI-Impact Faktor, vgl. Fröhlich 2002c.

[4] Ein Beispiel: Stanford-Leitungsfunktionär Terman akzeptierte bereitwillig die Bedingungen für die „funds that had been provided by William Robertson Coe, a generous benefactor ... Coe, a right-wing businessman, had placed restrictions on his patronage; his money was to be used to meet „the threat of Communism, Socialism, collectivism, Totalitarianism and other ideologies opposed to the preservation of our System of Free Enterprise." (Lowen 1977, 216, fn67) Der bisherige Politikwissenschaftsprofessor wurde daher entlassen, ein strammer Antikommunist eingestellt.

[5] Vgl. dazu die Fallstudie zu Stanford von R. S. Lowen 1977. Eine andere US-Universität „to fire anyone who doesn't attract research grants that provide between 50 and 100 per cent of his or her salary." (Lowen 1977, 237)

[6] Eine positive Ausnahme dürfte hier die Betonung von Interdisziplinarität bei EU-Projekten darstellen.

[7] Zur Diskussion von „electronic peer review conferences" vgl. Peters 1996, zu Praxiserfahrungen „interaktiver Publikation" im Internet vgl. Harnad 1996.

[8] Je nach Problemstellung sollten natürlich auch sonstige Betroffene und Kompetente im Sinne partizipativer Gestaltung einbezogen werden.

Univ.-Prof. Dr. Gerhard Fröhlich
Institut für Philosophie und Wissenschaftstheorie, Universität Linz
A-4040 Linz, Freistädterstraße 315
E-Mail: gerhard.froehlich@jku.at

Zuerst erschienen in: ODOK'01. Drehscheibe E-Mitteleuropa. Information: Produzenten, Vermittler, Nutzer. Die gemeinsame Zukunft. Tagungsberichte vom 9. Österreichischen Online-Informationstreffen und 10. Österreichischen Dokumentartag. Hrsg. V. Eveline Pipp. Wien: Phoibos, 2002 (Biblos-Schriften 173), 129-146.

Mit freundlicher Genehmigung des Verfassers und der Herausgeberin.