

Anreize im Wissenschaftssystem

von

Margit Osterloh und Bruno S. Frey

Universität Zürich*

und

CREMA – Center for Research in Economics, Management and the Arts, Switzerland

(diese Version: 12. September 2008)

Zusammenfassung:

Das deutsche Wissenschaftssystem ist gekennzeichnet durch die Einführung an Outputs orientierter Anreizsysteme. Wissenschaftliche Arbeit ist besonders unsicher und deshalb auf ein funktionierendes Peer-Review-System angewiesen. Qualitatives Peer Review und quantitative bibliometrischen Verfahren haben viele Mängel. Die steigende Rolle von Drittmitteln als Leistungsindikator und Forschungsanreiz macht WissenschaftlerInnen noch stärker vom Peer Review System abhängig. Monetäre Anreizsysteme verstärken die Mängel noch. Wir stellen drei Verbesserungsvorschläge vor: „Pay for performance“ im Wissenschaftssystem ist zurückzunehmen; die Bedeutung des „Peer Review Systems“ ist einzudämmen und sein Schwerpunkt in die Eingangsphase wissenschaftlicher Karrieren zu verlegen; und die Autonomie für wissenschaftliche Arbeit ist zu verstärken.

(97 Worte)

Keywords: Peer Review, Pay for Performance, Autonomie, Forschung, Wissenschaft

JEL Classification: A1, D01, I28, J45, M5.

* Prof. Dr. Dr.h.c. Margit Osterloh, Institut für Organisation und Unternehmenstheorien, Universitätsstrasse 84, CH-8006 Zürich. E-mail: m.osterloh@iou.uzh.ch.
Prof. Dr. Dr.h.c.mult. Bruno S. Frey, Institut für Empirische Wirtschaftsforschung, Winterthurerstr. 30, CH-8006 Zürich. E-mail: bsfrey@iew.uzh.ch.

1. Einleitung

Im deutschen Wissenschaftssystem finden derzeit zahlreiche Veränderungen statt, z.B.

- die Zunahme und Allgegenwart von Evaluationen, Rankings und Ratings,
- der Wechsel von einer Steuerung über Mittelzuwendungen zu einer an Zielvorgaben orientierten Outputsteuerung,
- die relative Zunahme drittmittelgeförderter Forschung,
- die Einführung eines „pay-for-performance“-Systems durch die variable W-Besoldung,
- die verstärkte Vertikalisierung des Hochschulsystems durch die Exzellenzinitiative.

Insgesamt sind die Reformen gekennzeichnet durch die Einführung an Outputs orientierter Anreizsysteme. Universitäten sollen unternehmerischer werden (Bok 2003; Maasen & Weingart 2008). Will man die Erfolgchancen dieser Reformen einschätzen, so sollte man die Unterschiede zwischen der Wissensproduktion in gewinnorientierten Unternehmen und im durch öffentliche Gelder geförderten Wissenschaftssystem bedenken.

Mit diesen Fragen hat sich seit geraumer Zeit die Wissenschaftsökonomik („economics of science“) befasst (z.B. Nelson 1959, 2004, 2006; Dasgupta 1994; Stephan 1996), die Einsichten der Wissenschaftssoziologie, insbesondere von Merton (1973) und Polanyi (1962) aufgreift. Deren Ergebnisse werden im *zweiten* Abschnitt zusammenfassend dargestellt. Es wird gezeigt, dass sich aus dieser Sicht in der Wissenschaft ein besonderes Erfolgskriterium etabliert hat, nämlich die Feststellung der Ersterfindung durch die Peers (Prioritätsregel). Diese Besonderheit bewirkt aber zugleich, dass wissenschaftliche Arbeit durch besonders hohe Unsicherheit gekennzeichnet ist. Daraus resultiert, wie im *dritten* Abschnitt dargelegt wird, dass auch das Belohnungs- und Anreizsystem in der Forschung besondere Ausprägungen haben muss. Es ist auf ein funktionierendes Peer-Review-System angewiesen. Monetäre Anreize spielen eine vergleichsweise geringe Rolle. Inwieweit das Peer-Review-System diese zentrale Rolle tatsächlich erfüllt, wird allerdings in der Literatur zur Wissenschaftsökonomik nicht untersucht. Deshalb erfolgt im *vierten* Abschnitt eine Bestandsaufnahme des Peer Review Systems. Nach einer Darlegung der Probleme von qualitativen Peer Review Verfahren und von quantitativen bibliometrischen Verfahren wird gefragt, ob die Kombination von quantitativen und qualitativen Kriterien („Informed Peer Review“) zu einem Ausgleich der Fehlerquellen führt. Argumentiert wird, dass bei sorglosem Einsatz die Fehlerwahrscheinlichkeit sogar überproportional wachsen kann. Dies ist eine negative Bilanz für die Funktionsfähigkeit eines Belohnungs- und Anreizsystems, das mit der Qualität seines Peer Review Systems steht und fällt. Im *fünften*

Abschnitt gehen wir auf die steigende Rolle von Drittmitteln als Leistungsindikator und Forschungsanreiz ein. Diese Entwicklung verstärkt die Abhängigkeit der WissenschaftlerInnen vom Peer Review System. Die Probleme bestehen nicht nur in möglicherweise falschen Gutachterentscheidungen, sondern auch darin, dass zu hohe Drittmittelanteile zu „research empires“ und abnehmenden Grenzerträgen führen können. Im *sechsten* Abschnitt legen wir dar, dass die geschilderten Probleme durch die Einführung monetärer Anreizsysteme wie die W-Besoldung verstärkt werden. Im *siebten* Abschnitt wird diskutiert, welche Konsequenzen daraus für die Ausgestaltung des Anreiz- und Belohnungssystems für die Wissenschaft gezogen werden sollten. Wir stellen drei Vorschläge zur Diskussion: Erstens die Zurücknahme von „pay for performance“ im Wissenschaftssystem, zweitens die Eindämmung der Bedeutung des „Peer Review Systems“ und die Verlagerung seines Schwerpunktes in die Eingangsphase wissenschaftlicher Karrieren und drittens die Gewährung von Autonomie als wichtigstem Anreiz für wissenschaftliche Arbeit.

2. Besonderheiten der Wissensproduktion in der Forschung

Aus ökonomischer Sicht unterscheidet sich (Grundlagen-) Forschung¹ idealtypisch in vierfacher Hinsicht von der Wissensproduktion in gewinnorientierten Unternehmen.

2.1. Grundlagenwissen ist ein öffentliches Gut, dessen Wert von der Öffentlichkeit nur schwer beurteilt werden kann

Grundlagenwissenschaft stellt zu einem hohen Anteil ein öffentliches Gut dar (Nelson 1959, 2006), gekennzeichnet durch zwei Merkmale:

- Nicht-Rivalität im Konsum. Die Nutzung von Wissen durch eine Person beeinträchtigt nicht deren Nutzung durch andere Personen. Es ist im Gegenteil von einem Wissens-Spillover auszugehen (Nelson 1959; Rosenberg 1974; Stern 2004), welcher die Quelle ökonomischen Wachstums darstellt (Romer 1994; Stephan 1996)
- Nicht-Ausschliessbarkeit vom Konsum auch für diejenige, die zu seiner Produktion nichts beigetragen haben. Wenn Wissen veröffentlicht wird, ist es nicht mehr

¹ Die Grenzen von Grundlagen- und anwendungsorientierter Forschung sind fließend. Beide Arten von Wissenschaft schliessen einander nicht aus, vgl. Stokes (1997). Die folgenden Ausführungen gelten in erster Linie für die Grundlagenwissenschaft, obwohl z.B. Nelson (2004) dafür plädiert, dass der Grundsatz der „scientific commons“ auch für die Entwicklung von Technologien bzw. von angewandter Wissenschaft gelten sollte.

appropriierbar. Die Offenlegung von Wissen stellt zugleich die Voraussetzung für die kumulative Weiterentwicklung von Wissen dar.²

Unternehmen haben nur in Ausnahmefällen ein Interesse daran, öffentliche Güter zu produzieren. Das gilt auch für die Produktion und Veröffentlichung von Forschung³. Aus diesem Grund muss der grösste Teil der Erzeugung von Grundlagenwissen vom Staat finanziert werden.⁴ Grundlagenwissenschaft produziert deshalb zu einem hohen Anteil öffentlich verfügbares Wissenskapital („open science“).⁵ Gewinnorientierte Unternehmen sorgen üblicherweise durch Geheimhaltung oder durch gewerbliche Schutzrechte dafür, dass sich die Investition in die Wissensproduktion auszahlt („private knowledge“). Dadurch wird jedoch die kumulative Weiterentwicklung von Wissen behindert, weil das private Wissen der Öffentlichkeit nicht unentgeltlich zur Verfügung steht.⁶

Anders als bei manch anderen öffentlichen Gütern, wie z.B. bei einem Leuchtturm oder der Verkehrsinfrastruktur, kann der Wert von Grundlagenwissen nicht von der Öffentlichkeit beurteilt werden. Ob dieses Wissen zu verkäuflichen oder anderweitig sinnvollen Produkten führt, ist häufig ungewiss und, wenn überhaupt, nur langfristig feststellbar (z.B. Nelson 2004). Deshalb versagt der Markt in der Regel bei der Feststellung der Qualität von Grundlagenforschung. Darüber hinaus ist die Informationsasymmetrie zwischen WissenschaftlerInnen und Laien, aber auch zwischen den spezialisierten Wissenschaftsdisziplinen so gross, dass eine Überwachung der angewandten Verfahren in der Wissenschaft durch Externe nicht möglich ist (Dasgupta &

² Darauf beruht u.a. der Erfolg der Opensource Software, vgl. Osterloh & Rota (2007).

³ Wenn Unternehmen Grundlagenforschung betreiben, tun sie das aus drei Gründen: Erstens, weil sie sonst mangels „absorptiver Kapazität“ nicht in der Lage sind, von Wissen-Spillovern zu profitieren (Cohen & Levinthal 1990). Zweitens, weil sie sonst nicht genug qualifizierte WissenschaftlerInnen rekrutieren können, welche oft Publikationsmöglichkeiten zur Bedingung ihrer Anstellung machen. Darüberhinaus reagieren Forschende oft auf Publikationsverbote mit einem Absinken der Arbeitsmoral. Drittens, weil Unternehmen mit Publikationen Reputation erwerben, die auf dem Kapitalmarkt positive Signale setzt und beim Einwerben von Regierungsaufträgen und Forschungsgeldern hilft. Immer aber entsteht hierbei ein Konflikt zwischen den proprietären Interessen des Unternehmens und dem Interesse an Veröffentlichung der Forschungsergebnisse, vgl. Stephan (1996).

⁴ Patente, welche die Appropriierbarkeit der gewerblichen Nutzung von Wissen ermöglichen, setzen in Deutschland die gewerbliche Anwendbarkeit voraus, die bei Grundlagenforschung meist nicht gegeben ist. Zur kritischen Diskussion der Absenkung der Schwellen für die Patentierbarkeit in den USA siehe Nelson (2004, 2006).

⁵ In jüngerer Vergangenheit gibt es allerdings eine Tendenz zur Privatisierung ehemals öffentlicher Forschung. Bahnbrechend hierfür war in den USA der Bayh-Dole Act, welcher Universitäten ermuntert, öffentlich finanzierte Forschungsergebnisse zu patentieren. Diese Entwicklung wird allerdings von vielen Innovationsforschern kritisch kommentiert, vgl. z.B. Nelson (2004, 2006), Dosi, Marengo & Pasquali (2006), Osterloh & Lüthi (im Druck).

⁶ Urheberrechte gewähren allerdings oft keinen ausreichenden Schutz vor Imitation. Darüberhinaus sind in vielen Fällen die Kosten von Patenten höher als der Nutzen, sodass Unternehmen und Individuen häufig darauf verzichten. vgl. z.B. Dosi, Marengo & Pasquali (2006), Osterloh & Lüthi (im Druck).

David 1994: 505)⁷. Deshalb kann wissenschaftliche Qualität in der Regel nur innerhalb des Wissenschaftssystems von unmittelbaren Peers beurteilt werden. Dies bildet die Grundlage für den Anspruch des Wissenschaftssystems auf weitgehende Autonomie der „Republic of Science“, welche auf die Selbstkoordination unabhängiger WissenschaftlerInnen setzt (Polanyi 1962).

2.2. Das Erfolgskriterium in der Wissenschaft ist an die Prioritätsregel geknüpft.

Wegen der Ungewissheit über den Erfolg am Markt und wegen der Informationsasymmetrie zwischen Insidern und Outsidern hat sich in der Wissenschaft – wie von Robert Merton (1973) überzeugend dargestellt – mit der sog. Prioritätsregel ein besonderes Erfolgskriterium etabliert. Die Prioritätsregel schreibt den Erfolg nur derjenigen Person oder Institution zu, welche (a) eine Entdeckung als erste gemacht hat und (b) dies von der Scientific Community bestätigt erhält. Dieses Erfolgskriterium führt zu einem symbolischen Eigentumsrecht via wissenschaftlicher Reputation (Stephan 1996). Es ersetzt reale Eigentumsrechte, die bei öffentlichen Gütern nicht möglich sind. Das hat zwei Implikationen:

- Der oder die Zweite im Konkurrenzkampf erleidet einen hohen Verlust, auch wenn der zeitliche Abstand nur gering war. Es gilt „the winner takes all“ (Dasgupta & David 1994; Stephan 1996; Sorensen & Fleming 2004). Kleine Unterschiede in der Leistungsfähigkeit, aber auch in Bezug auf glückliche Umstände, werden dadurch in riesige Unterschiede in der Reputation übersetzt.
- Die Entdeckung muss den Peers so schnell wie möglich kommuniziert werden. Daraus resultiert, dass man die Entdeckung veröffentlichen und damit zum öffentlichen Gut machen muss. Die Allgemeinheit profitiert davon in dreifacher Weise. Zum *einen* können die Peers die Erfindung evaluieren, d.h. Risiken oder Irrtümer offen legen. Zum *zweiten* können sie die Entdeckung nutzen, um ihre eigene Forschung weiterzutreiben, Redundanzen zu vermeiden und ihrerseits Priorität für ihre Forschung anzustreben. Zum *dritten* kann die Umsetzung des neuen Wissens in verwertbare Technologien verbilligt und beschleunigt werden (Dasgupta & David 1994: 500).

Im Gegensatz zur Wissenschaft gibt es auf Märkten ein eindeutiges Erfolgskriterium, den Gewinn, auch wenn sich dieser mitunter erst längerfristig feststellen lässt.

⁷ D.h. es ist von Outsidern weder eine Output- noch eine Prozesskontrolle möglich, sondern nur eine sog. „Clan-Kontrolle“ durch die Peers, vgl. Ouchi (1977).

2.3. Wissenschaftlicher Erfolg ist mit hohen persönlichen Ungewissheiten verbunden

Forschung ist mit grossen Ungewissheiten für die ForscherInnen verbunden.

- Die jeweils Zweiten in der Entdeckungs-Konkurrenz gehen aufgrund des „winner takes all“ Phänomens leer aus. Dieses hohe Risiko wird noch verstärkt durch den Matthäus-Effekt, d.h. „wer hat, dem wird gegeben“ (Merton 1968). Es sind dies kumulative Vorteile der jeweiligen Ersten bei Zitierungen und bei der Einwerbung von Forschungsmitteln.
- Auch ist der Forschungsprozess mit grossen Unsicherheiten behaftet. Die Forscher müssen zwischen verschiedenen Forschungsansätzen auswählen. Ein neuer, unorthodoxer Ansatz verspricht grosse Erfolge gemäss der Prioritätsregel,⁸ kann aber dazu führen, dass man am Ende mit leeren Händen dasteht. Verstärkt wird diese Unsicherheit dadurch, dass Diversität von Forschungsansätzen für die Wissenschaft zentral ist, um Irrtümer auszugleichen (Merton 1973:364 ff.; Dasgupta & David :506; Nelson 2004).
- Forschung beantwortet oft Fragen, die gar nicht gestellt wurden (Polanyi 1962), oder führt zu Entdeckungen, die gar nicht angestrebt wurden, den sog. Serendipitätseffekten (Stephan 1996).⁹ Selbstgesetzte Ziele oder die Einhaltung von Zielvereinbarungen werden deshalb nicht selten verfehlt, ohne dass dies immer ein Nachteil wäre.

Der Erfolg von MitarbeiterInnen in Unternehmen ist im Gegensatz zur Wissenschaft in der Regel wesentlich stärker durch eigene Anstrengung und den eigenen Einsatz beeinflussbar, weil die Prioritätsregel weitaus weniger Gültigkeit hat.¹⁰

2.4. Forschung findet im Spannungsfeld zwischen dem Wettbewerb um Priorität und der Notwendigkeit zum Teilen von Wissen statt.

Trotz der gegenseitigen Konkurrenz aufgrund der Prioritätsregel ist Wissenschaft auf Kooperation angewiesen. Wissenschaftlicher Fortschritt beruht häufig auf einem evolutionären Prozess. Ausgehend von früheren Erfolgen und Misserfolgen findet ein

⁸ Gemäss der Untersuchung von Gaston (1971) gaben 44 Prozent von erfolgreichen Physikern im Energiebereich an, dass sie sich Forschungsgebiete ausgesucht haben, in denen die Konkurrenz zwischen den Forschern gering war. Ein Albert Einstein zugeschriebenes Zitat besagt. „The secret to creativity is knowing how to hide your sources“.

⁹ Zu einer kritischen Diskussion von Serendipitätseffekten vgl. Godoe (2000).

¹⁰ Häufig scheitern die Technologieführer am Markt, d.h. Technologiefolger gewinnen. Erfolg auf dem Markt setzt ein Bündel von Management-Anstrengungen voraus, die nichts mit der Prioritätsregel zu tun haben.

kollektiver Lernprozess statt (Nelson 2004). Dieser wird durch den Zwang begünstigt, die Priorität durch rasche Veröffentlichung zu dokumentieren. Dieses Wissen ist durch Urheberrechte nur unzureichend geschützt und die Gefahr von Plagiaten ist allgegenwärtig. Als Ausweg zur Minimierung dieses Risikos gilt die Kooperation und der Gedankenaustausch in überschaubaren Forschungs-Teams. Innerhalb von Teams kann man darüber hinaus noch das eigene Forschungs-Portfolio erweitern und auf diese Weise das Risiko vermindern, einen erfolglosen Forschungspfad eingeschlagen zu haben (Dasgupta & David 1994; Stephan 1996). Forschungsteams funktionieren aber nur, wenn eine hohe wechselseitige Bereitschaft besteht, sich nicht als Trittbrettfahrer zu verhalten, sondern freiwillig zum teamspezifischen Wissen beizutragen. Die Voraussetzung hierfür ist ein hohes Vertrauen in die prosoziale, intrinsische Motivation der Teammitglieder (Osterloh & Frey 2000; Osterloh & Weibel 2006).

Auch in Unternehmen gibt es Spannungsfelder zwischen opportunistischer Wissenszurückhaltung und funktionierender Teamarbeit (Osterloh 2007), jedoch ist dieser Konflikt wegen der geringeren Bedeutung der Prioritätsregel weniger ausgeprägt.

3. Besonderheiten des Belohnungs- und Anreizsystems in der Forschung

Die besonderen Charakteristiken von Forschung bewirken, dass deren Belohnungs- und Anreizsysteme mit denen in Unternehmen nur eingeschränkt vergleichbar sind. Zwar ist beiden Bereichen gemeinsam, dass Individuen mit ihrer Arbeit Geld verdienen (extrinsische Motivation) sowie eine inhaltlich befriedigende Aufgabe erfüllen wollen (intrinsische Motivation). Aber aus den bisherigen Ausführungen folgen vier besondere Merkmale für die Funktionalität eines wissenschaftlichen Belohnungs- und Anreizsystems.

Zum *ersten* hat ein funktionierendes Peer-Review-System innerhalb der „Republic of Science“ eine überragende Bedeutung für die Entwicklung und Beurteilung der Qualität von Wissenschaft und damit für das Belohnungs- und Anreizsystem (z.B. Nelson 2004).

Zum *zweiten* muss in der Wissenschaft der fixe Anteil des Einkommens hoch sein. Der Grund liegt im hohen Risiko, nicht der oder die Erste zu sein und durch das „winner takes all“ System zusätzlich bestraft zu werden. Im Wissenschaftsbetrieb gibt es aber für die „Pechvögel“ immer die Möglichkeit, fehlende Erfolge in der Forschung durch Lehre und Forschungsadministration wettzumachen. Dieses System sichert zugleich die Effizienz einer solchen Entlohnungsform (Dasgupta & David: 499).

Zum *dritten* sollten hohe Lohndifferenzen in Forschungsteams zur Aufrechterhaltung der Kooperationsbereitschaft vermieden werden. Empirische Befunde

zeigen, dass hohe Lohndifferenzen die Team-Produktivität vermindern (Siegel & Hambrick 2005; Torgler, Schmidt & Frey 2006), auch in der Forschung (Pfeffer & Langhton 1993).

Zum *vierten* gibt es in der Wissenschaft ein besonderes Präferenzsystem, welches Merton (1973) „taste for science“ nennt. In diesem spielt die nichtmonetäre Belohnung eine grosse Rolle. Insbesondere zwei nichtmonetäre Anreize sind bedeutsam. Es ist dies zum einen die Anerkennung durch die wissenschaftliche Gemeinschaft in Form von Preisen, Ehrendoktoraten oder Mitgliedschaften in prestigereichen Akademien (Stephan 1996; Frey & Neckermann 2008).¹¹ Als höchste Anerkennung gilt ein Eponym, d.h. – wie etwa beim Haleyschen Komet - die Zuschreibung des Namens des Entdeckers zur Entdeckung. Für die grosse Mehrzahl von WissenschaftlerInnen sind allerdings Publikationen und Zitationen die relevante Form der Anerkennung. Zum anderen ist die gewährte Autonomie zentral. Stern (2004) zeigt, dass frisch promovierte Biologen, die in der kommerziellen Forschung arbeiten, eine beträchtliche Lohneinbusse in Kauf nehmen, wenn sie dafür mehr Autonomie in ihrer Arbeit erhalten. Ein weiterer Indikator hierfür ist, dass das Einkommen von Akademikern in Deutschland an Hochschulen, welche traditionell ein grosses Ausmass an Autonomie gewähren, etwa 20 Prozent geringer ist als ausserhalb der Hochschule (Jansen, Schomburg & Teichler 2007: 106). Da Autonomie zugleich die wichtigste Voraussetzung für kreative Arbeit (Amabile 1998) und hohen Forschungsoutput darstellt (Mudambi u.a. 2007), ist ein niedrigeres Einkommen an der Hochschule in einem gewissen Ausmass durchaus funktional. Es bewirkt einen Selektionseffekt. Es zieht diejenigen an, die im Vergleich zu monetären Entgelten eine höhere Präferenz für Autonomie und für Anerkennung durch die Peers haben – beides Voraussetzungen für eine Erfolg versprechende wissenschaftliche Tätigkeit.

4. Das Peer Review System auf dem Prüfstand

Das Peer-Review-System hat – wie geschildert – eine überragende Bedeutung als Basis für das Belohnungs- und Anreizsystem in der Forschung. Die Funktionsfähigkeit dieses Systems hängt damit entscheidend von der Fähigkeit der Peers ab, diejenigen zu identifizieren, welche beanspruchen dürfen, die Ersten zu sein. Diese Fähigkeit wird in der Literatur zur Wissenschaftsökonomik eigenartigerweise kaum hinterfragt (vgl. z.B. Dasgupta & David 1994; Nelson 2004, 2006), ebenso wenig wie die Fähigkeit des Peer-

¹¹ Zuckerman (1992) schätzt, dass es anfangs der 90er Jahre allein in Nordamerika über 3000 wissenschaftliche Preise gab.

Review-Systems, Rückkopplung und Hilfestellung auf dem Weg zu wissenschaftlicher Exzellenz zu geben. Deshalb soll im Folgenden eine Bestandaufnahme der Funktionsfähigkeit des (Informed) Peer-Review-Systems vorgenommen werden, und es sollen einige der Reformen im deutschen Wissenschaftssystem daraufhin geprüft werden, inwieweit sie die Funktionsfähigkeit dieses Systems unterstützen oder schwächen.

4.1.Probleme des qualitativen Peer Review Systems

Das traditionelle Begutachtungssystem steht schon seit geraumer Zeit aufgrund umfangreicher empirischer Evidenz auf dem Prüfstand (Starbuck 2006). Die Kritik lässt sich folgendermassen zusammenfassen:

- *Geringe Reliabilität der Gutachter-Urteile.*

Die Übereinstimmung zwischen Gutachterurteilen ist gering (z.B. Miner & MacDonald 1981; Peters & Ceci 1982; Cole 1992; Weller 2001; Bornmann & Daniel 2003).¹² Sie ist bei ablehnenden Gutachten höher als bei zustimmenden (Cichetti 1991). Daraus folgert Lindsey (1991), dass ohne Berücksichtigung der Übereinstimmung bei den ablehnenden Gutachten die Reliabilität bei der Auswahl der Besten minimal ist.

- *Geringe prognostische Qualität von Gutachten .*

Gottfredson (1978) ermittelt , dass die Ratings von GutachterInnen über die Qualität eines Manuskriptes nur mit 0,24 mit den späteren Zitationen korrelieren.¹³ Starbuck (2006: 83.) schätzt anhand verschiedener Indikatoren die Korrelation von Reviewer-Urteilen mit der tatsächlichen Qualität eines Manuskripts auf 0,25 bis 0.3. Er folgert daraus, dass sich mehr als 50 Prozent nicht unter den besten der eingereichten Artikel befinden. Es ist vielfach dokumentiert, dass in hochrangigen Journals Artikel zurückgewiesen wurden, die später hohe Preise gewonnen haben, einschliesslich des Nobel-Preises (Gans & Shepherd 1994; Campanario 1996). Dies sind keineswegs Einzelfälle (Weingart 2005; Starbuck 2006). Horrobin (1996) behauptet sogar, dass Laien bei der Beurteilung medizinischer Forschung nicht weniger geeignet wären als die Fachkollegen.

- *Opportunistisches Verhalten der Gutachter und Editoren*

GutachterInnen beurteilen Artikel besser, die ihre eigenen Arbeiten zustimmend zitieren, und weisen Artikel eher zurück, die ihre Arbeit in Frage stellen (Lawrence 2003: 260). In einer experimentellen Studie fand Mahoney (1977) heraus, dass Gutachter zu 71 Prozent

¹² Allerdings gibt es auch Gegenstimmen, z.B. Sonnert (1995). Zur kritischen Diskussion vgl. z.B. Hirschauer (2004), Hornbostel (2008). Die Autoren stellen fest, dass die niedrige Gutachter-Reliabilität hinzunehmen sei, weil Übereinstimmung nicht das Ziel der Wissenschaft wäre. Offen bleibt, welche Auswirkung dies auf die Ablehnung oder Akzeptanz von Manuskripten haben soll.

¹³ Vgl. aber die Kritik an der Auswertung von Zitierungen in Abschnitt 4.2.

methodische Mängel in einen Manuskript finden, wenn es der herrschenden theoretischen Ansicht widerspricht, hingegen nur 25 Prozent, wenn es im Einklang damit ist. Editoren wenden längere Zeit für den Begutachtungsprozess auf, wenn das Manuskript in ihr Spezialgebiet fällt (Ellison 2002), möglicherweise weil sie in diesen Fällen kleinlicher sind (Starbuck 2006: 81).

- *Opportunistisches Verhalten der Autoren*

Autoren fühlen sich von Gutachtern gedrängt, ihre Manuskripte zu ändern, um die Akzeptanz zu erhöhen, auch wenn dies ihrer Überzeugung widerspricht. Bedeian (2003, 2004) fand heraus, dass dies bei nicht weniger als 25 Prozent von Autoren der Fall ist, die ihr Manuskript wiedereinreichen. Frey (2003) nennt dieses Verhalten „Akademische Prostitution“.

Zusammengenommen ist damit festzustellen: Die Annahme eines Manuskriptes gleicht einem Zufallprozess. Das ist der Grund, warum sogar der Editor eines angesehenen Journals den Ratschlag gibt „Just Try, Try Again“ (Durso 1997). Das System der qualitativen Peer Reviews beruht damit – trotz seiner hohen Bedeutung für die Funktionsfähigkeit des Wissenschaftssystems – auf einer erstaunlich fragwürdigen wissenschaftlichen Grundlage (z.B. Atkinson 2001).

4.2. Probleme bibliometrischer Verfahren

Als Reaktion auf die Kritik an qualitativen Peer Reviews haben sich zunehmend quantitative, bibliometrische Verfahren etabliert, die auf Zitationsanalysen und Impact-Faktoren beruhen.¹⁴ Von quantitativen Methoden erhofft man sich *zum einen*, dass sie objektiver als qualitative Methoden seien (z.B. Weingart 2005), weil

- sie auf weitaus mehr Einschätzungen beruhen als die drei bis vier üblichen Peer Reviews,
- sie eine Kontrolle von Begünstigungen durch „old boys networks“ darstellen,
- sie nicht-reaktiv seien, d.h. keine Rückwirkungen auf den Review-Prozess hätten.

Zum *anderen* wird als Vorteil quantitativer Messgrößen angesehen, dass sie auch von Personen ausserhalb der „Republic of Science“ benutzt werden können, insbesondere von fachfremden Kollegen, der Presse, der Wissenschaftsadministration und der Politik. Auf diese Weise wird der interne Evaluationsprozess, der für die Selbstorganisation der

¹⁴ Die Britische Regierung will nach eigenem Bekunden in der nächsten Runde des „Research Assessment Exercise (RAE)“ das System der Peer Reviews gänzlich durch bibliometrische Daten ersetzen, vgl. Evidence Report (2007).

Wissenschaft typisch war, externalisiert.¹⁵ Outsider des Wissenschaftssystems erhalten – so hofft man – einen transparenten und eindeutigen Einblick in eine ansonsten undurchschaubare Wissenschaft und damit die Möglichkeit zur Kontrolle. Das erklärt die Beliebtheit von bibliometrischen Verfahren als Grundlage von Rankings und Ratings. Es erklärt auch den Druck der Politik, Rankings bzw. Ratings als Ausgleich für die den Hochschulen neuerdings gewährte grössere Autonomie zur Verfügung zu stellen. Es sind diese angeführten Vorteile, welche vor allem die derzeitigen Reformbestrebungen im deutschen Wissenschaftssystem legitimieren sollen.

Allerdings sind die genannten Vorteile quantitativer, bibliometrischer Verfahren durchaus problematisch. Es existiert hierüber jedoch nur wenig konzeptuelle und empirische Forschung. Eine vorläufige Bestandsaufnahme der Literatur lässt sich zu drei Gruppen von Problemen bibliometrischer Verfahren zusammenfassen, die kumulativ wirken.

Zum *ersten* bilden dieselben Peer Reviews, deren Probleme überwunden werden sollen, die Basis aller bibliometrischen Verfahren, denn gezählt werden zumeist nur Veröffentlichungen in begutachteten Journals. Wie gezeigt wurde, sind diese Probleme beträchtlich. Bibliometrische Verfahren und die darauf basierenden Evaluationen können nur dann an Exaktheit gewinnen, wenn durch die Aggregation mittels Zitaten und Impact Faktoren ein Fehlerausgleich entsteht (Weingart 2005). Dieser Ausgleich ist aber wegen der methodischen Probleme und der Reaktivität der Messverfahren fraglich. Dies ist Gegenstand der folgenden beiden Punkte.

Zum *zweiten* gibt es beträchtliche methodische Probleme (vgl. z.B. Adler, Ewing & Taylor 2008), z.B.

- Selektionsprobleme: Werden Bücher *und* Zeitschriften ausgewertet und welche Zeitschriften werden ausgewählt? Bibliometrische Daten repräsentieren immer nur einen Teil des wissenschaftlichen Kommunikationsprozesses (Reedijk 1998). Die Dominanz von wenigen Indizes, wie dem des ISI (Institute of Scientific Information) und des SSCI (Social Science Citation Index) hat dazu geführt, dass die Selektionskriterien, die Datenqualität und die Methoden der Datenaufbereitung dieser Indizes kaum mehr kritisch in die Einschätzung der Ergebnisse einbezogen werden. Sie werden „off the shelf“ benutzt, obwohl kleine Veränderungen der Kriterien und der Methoden grosse Wirkungen in Bezug auf die Ergebnisse haben können (Ursprung & Zimmer 2006).

¹⁵ Informationsökonomisch gesprochen wird damit versucht, ein Vertrauensgut in ein Suchgut überzuführen, vgl. Darby & Karni (1973).

- Abgrenzungsprobleme: Insbesondere interdisziplinäre Forschung leidet darunter, dass sie oft fehlerhaft ausgewertet wird, weil die Zuordnung zu Forschungsfeldern unvollständig oder falsch ist. Dies ist besonders problematisch, weil Grenzgängertum häufig eine Voraussetzung für Kreativität darstellt (Amabile u.a. 1994; Dogan 1999).
- Mangelndes Wissen über Zitier-Gewohnheiten: Diese sind nicht nur in einzelnen Disziplinen, sondern auch in Sub-Disziplinen sehr verschieden, sodass ein Vergleich von bibliometrischen Daten schon innerhalb einer Disziplin problematisch sein kann (Ursprung & Zimmer 2006). Darüber hinaus hängen sie von der Erreichbarkeit der Zitationsquellen ab.¹⁶
- Mangelndes Wissen über die Art der Zitate: Diese können eine zustimmende oder ablehnende Bedeutung haben. Mitunter folgen sie einfach einem Herdenverhalten oder spiegeln akademische Popularität von Star-Papieren wider (Faria 2005). Viele Zitate werden nur deshalb eingefügt, um die Wahl bestimmter Gutachter zu begünstigen und um diese dann freundlich zu stimmen. Simkin und Rowchowdhury (2005) zeigen erstens, dass die Wahrscheinlichkeit, zitiert zu werden, eine Funktion früherer Zitate ist. Dies entspricht dem von Merton (1968) beschriebenen „Matthäus-Effekt“.¹⁷ Die Autoren zeigen zweitens aufgrund einer sorgfältigen Analyse von beim Zitieren übertragenen Fehler, dass etwa 70 – 90 Prozent aller zitierten Papiere vermutlich gar nicht gelesen wurden.
- Irreführung durch den Impact-Faktor: Die relative Bedeutung einer wissenschaftlichen Zeitschrift, gemessen an der Häufigkeit, mit der diese Zeitschrift anderen Orten zitiert wird, besagt nicht, dass die dort veröffentlichten Artikel diesem Durchschnitt entsprechen¹⁸. Adler, Ewing & Taylor (2008:14) stellen denn auch in einem Gutachten für die „International Mathematical Union“ fest, dass die Verwendung von Impact-Faktoren zu grossen Fehlerwahrscheinlichkeiten führen kann und „atemberaubend naiv“ sei.

Es erstaunt nicht, dass angesichts dieser methodischen Probleme die Übereinstimmung der verschiedenen Rankings gering ist (vgl. z.B. Frey & Rost 2008; Küpper & Ott 2002; Maasen & Weingart 2008). Für die Ökonomik haben Ursprung & Zimmer (2006) gezeigt,

¹⁶ Ein besonders frappierender Fall ist Reinhard Selten, der bislang einzige deutsche Nobelpreisträger für Ökonomik. Der bahnbrechende Aufsatz, für den er den Nobelpreis erhielt, ist im Social Science Citation Index nicht enthalten, weil er in einem nicht berücksichtigten Journal veröffentlicht wurde.

¹⁷ Dies gilt nicht für den sog. Hirsch-Index (Hirsch 2005), vgl. Ursprung & Zimmer (2006). Dieser Index hat aber den schwerwiegenden Nachteil, dass er die Publikationsleistung von Autoren, die wenig publizieren, aber viel zitiert werden, nicht angemessen wieder gibt, vgl. Adler, Ewing & Taylor (2008).

¹⁸ Vgl Oswald (2007), der dies für die angesehensten ökonomischen Zeitschriften belegt hat.

dass im einflussreichen Ranking des „Handelsblattes“ jede dritte der einbezogenen Personen an der Spitze stehen könnte, wenn man die Methode ändert.

Die *dritte* Problemgruppe ist zugleich diejenige, die am wenigsten diskutiert wird.¹⁹ Sie resultiert daraus, dass auch bibliometrische Verfahren reaktiv sind, d.h. dass die Wissenschaftler und wissenschaftliche Institutionen in strategischer Weise ihr Verhalten danach ausrichten. Im Ergebnis werden diejenigen belohnt, die am besten verstehen, „to game the system“. Wichtig ist, dass die daraus entstehenden Verzerrungen des Anreiz- und Belohnungssystems auch für den hypothetischen Fall auftreten, dass die bibliometrischen Verfahren ohne Probleme durchgeführt worden wären.

Solche Verhaltensweisen wurden schon vom Sozialpsychologen Campbell (1976) beschrieben; sie sind als „Campbell`s Law“ in die Literatur eingegangen. „Je stärker ein einzelner quantitativer sozialer Faktor dazu benutzt wird, soziale Entscheidungen zu begründen, desto stärker ist er verzerrenden Einflüssen ausgesetzt und desto mehr führt er selbst dazu, die sozialen Prozesse zu verzerren und zu verfälschen, die eigentlich untersucht und verbessert werden sollen“ (Campbell 1976:49)²⁰. Beispiele aus dem Bildungsbereich sind zahlreich:

- In den USA sind regelmässige quantitativ ausgewertete Tests im Schulbereich üblich, an die auch Sanktionen geknüpft sind. Empirische Befunde zeigen, dass die Ergebnisse künstlich nach oben gedrückt werden, indem etwa schlechte Schüler ermutigt werden, am Testtag zuhause zu bleiben oder gar zum Schulabbruch bewegt werden. Beobachtet wurde auch ein oberflächliches „teaching to the test“ auf Kosten des gründlichen Lernens. Im Ergebnis verschwinden die positiven Ergebnisse der Evaluationen, wenn man für diese Faktoren kontrolliert. Darüberhinaus sinkt die Anzahl der Kinder mit einem Schulabschluss (Lind im Druck).
- In Italien werden finanzielle Ressourcen an die Hochschulen in Abhängigkeit von den bestandenen Examen verteilt. Als Folge schneiden die Absolventen derjenigen Universitäten auf dem Arbeitsmarkt am schlechtesten ab, welche die höchsten Absolventenzahlen aufweisen, weil diese ihr Anforderungsniveau gesenkt haben (Bagues, Labini & Zinoveya 2008).
- Hochschulen in Grossbritannien und in den USA engagieren Star-Forscher, insbesondere kurz vor Evaluationen, um in den Rankings und Ratings gut

¹⁹ Vgl. z.B. das Schwerpunkt-Heft der Zeitschrift „Perspektiven der Wirtschaftspolitik“ 2008, Heft 3, in dem dieses Problem nicht einmal erwähnt wird.

²⁰ Übersetzung übernommen aus Lind (in Druck).

abzuschneiden. Diese Stars werden hoch bezahlt, obwohl sie nur eine begrenzte Zeit an der entsprechenden Hochschule verbringen. Die Mittel werden eingespart, indem weniger Nachwuchskräfte eingestellt und gefördert werden und die Anzahl der Postdocs pro Vollprofessur wächst (Stephan 2008).

- Herausgeber von Zeitschriften veranlassen ihre Autoren, Publikationen ihrer Zeitschrift zu zitieren, um ihren Impact-Faktor zu verbessern (Garfield 1997; Smith 1997; Atkinson 2001).

Diese Befunde zeigen, dass quantitative Evaluationskriterien immer die Gefahr in sich bergen, dass professionelle ethische Normen erodiert werden (Kieser 1998). Etwas weniger manipulativ, im Ergebnis aber ebenfalls hochgradig verzerrend ist ein Sachverhalt, der in der Ökonomik unter dem Begriff „multiple tasking effekt“ behandelt wird (Holmstrom & Milgrom 1991): Komplexe Aufgaben sind durch eine grosse Anzahl von verschiedenen Kriterien gekennzeichnet, die unterschiedlich gut quantifizierbar sind. Die Evaluierten orientieren sich in erster Linie an den leicht messbaren Kriterien und lassen die schwer messbaren ausser Acht, obwohl diese häufig die wichtigeren sind. Im Hochschulbereich bedeutet das, dass vielfach die Qualität und Innovativität auf dem Altar der Quantität von Publikationen geopfert oder dass Forschungsmittel verschwendet werden. Folgende Beispiele zeigen dies:

- Forscher wenden die „Salamitaktik“ an, d.h. neue Ideen oder interessante Datensätze werden so dünn wie Salamischeiben aufgeschnitten, um die Anzahl der Publikationen zu maximieren (Weingart 2005). Diese Reaktion wird verstärkt, wenn monetäre Belohnungen an die Anzahl der Publikationen geknüpft werden. Ein solches Verhalten wurde von Butler (2003) für Australien nachgewiesen. Möglicherweise besteht ein kausaler Zusammenhang zur Tatsache, dass Australien bezüglich fast aller OECD Standards zurückfiel.
- Die Neigung wird verstärkt, konventionelle oder modische Ansätze zu präsentieren, die wenig Widerspruch hervorrufen und leichter veröffentlicht werden können (Horrobin 1996; Pritchard & Wilmott 1997; Muller-Carmen & Salzgeber 2005). Dies führt zu einer Homogenisierung der Forschung, wie dies für Business Schools (Gioia & Corley 2002) und die Volkswirtschaftslehre (Holcombe 2004) nachgewiesen wurde.
- Die Orientierung an Impact-Faktoren bewirkt, dass kleinere, spezialisierte und nicht-englischsprachige Zeitschriften an Attraktivität verlieren. Werden die Veröffentlichungen in Journals mit hohem Impact-Faktor als Kriterium für die

Beurteilung von Fachbereichen verwendet, so werden im Ergebnis weniger WissenschaftlerInnen eingestellt, welche Randbereiche vertreten und in diesen Zeitschriften publizieren. Dies wurde für Grossbritannien am Beispiel der Ökonomik gezeigt. Der Anteil heterodoxer, d.h. nicht streng neoklassisch orientierter ÖkonomInnen (z.B. in den Fächern Geschichte oder der Methoden der Ökonomik) ist seit Einführung des „Research Assessment Exercise“ drastisch zurückgegangen. Die Mehrheit der Studierenden hat keine Gelegenheit mehr, alternative Lehrmeinungen zu hören (Lee 2007).

- Es besteht ein Anreiz, nur noch erfolgreiche Tests zu publizieren und die negativen Ergebnisse zu verschweigen oder sogar zu beseitigen, weil sich negative Ergebnisse schlechter publizieren lassen. Dies widerspricht dem Ideal einer Wissenschaft, welche die Falsifikation von Hypothesen als ihre Kernaufgabe ansieht. Lernen aus den Fehlern der Forschungs-Gemeinschaft ist nicht mehr möglich.
- Wird Forschungsoutput anhand der Höhe eingeworbener Drittmittel gemessen, wie dies heute zunehmend der Fall ist (vgl. Abschnitt 5), so entsteht die Tendenz, mit Forschungsmitteln verschwenderisch umzugehen (Holcombe 2004).

Eine weitere, indirekte Folge der Orientierung an Quantität anstelle schwerer messbarer Qualität ist die zunehmende Last der Begutachtung und die damit einhergehende Wahrscheinlichkeit, dass die Qualität der Gutachten sinkt (Neidhardt 2006). Gutachten werden immer häufiger weniger qualifizierten Forschern weitergereicht, welche nicht in der Lage sind, die Innovativität einer Arbeit wirklich einzuschätzen.²¹ Darüber hinaus fehlt die Zeit, die Qualität der Daten und die statistischen Methoden zu überprüfen, wie Beispiele aus der Ökonomik zeigen (Hamermesh 2007).

Insgesamt ist festzustellen, dass es Einzelbefunde zu Problemen quantitativer Methoden der Messung von Forschungsleistungen gibt, dass aber über deren systematische Auswirkungen nur wenige Erkenntnisse existieren. Gleichwohl haben mittlerweile diese Methoden die Steuerungsmechanismen in der Wissenschaft umfassend verändert, ohne dass dies wissenschaftlich reflektiert wurde.

²¹ Weller (2004) fand allerdings eine negative Beziehung zwischen argumentativem Aufwand eines Gutachtens und dem formalen Qualifikationsniveau des Gutachtenden. Ob dadurch die Einschätzung der Innovativität eines Beitrags steigt, ist eher fraglich.

4.3. Probleme des „Informed Peer Review“

Vielen ForscherInnen und Forschungsinstitutionen gelten „informed Peer Reviews“, d.h. die Kombination aus quantitativen und qualitativen Evaluationsverfahren, als Königsweg (vgl. z.B. Deutscher Wissenschaftsrat 2008; Weingart 2005). Bibliometrische Verfahren sollen qualitative Peer Reviews korrigieren und ergänzen. Jedoch ist es keineswegs selbstverständlich, dass eine Kombination von quantitativen und qualitativen Kriterien zum Ausgleich der Fehlerquellen führt. Ebenso gut ist möglich, dass diese Kombination bei fahrlässigem Gebrauch der Kriterien kumulativ oder sogar multiplikativ wirkt. Das ist zum Beispiel dann der Fall, wenn Peers bei der Beurteilung von Forschungsanträgen, bei Berufungsentscheidungen oder bei der Evaluation von Abteilungen quantitative Indikatoren benutzen, deren methodische und technische Grundlage sie nicht kennen. Die Indikatoren haben dann eine beachtliche nicht kontrollierte Hebel-Wirkung. Wie gezeigt, haben schon geringe Änderungen der Kriterien und Methoden grosse Auswirkungen auf die Position in der Beurteilung. Aber selbst die Berücksichtigung dieser Faktoren gibt nicht annähernd ein objektives Bild, weil die geschilderten reaktiven Auswirkungen nur schwer einschätzbar sind. Es ist anhand von quantitativen Indikatoren kaum zu beurteilen, inwieweit gute Forschungsleistungen oder gutes „window dressing“ vorliegen. Wenn – wie in manchen Berufungsverfahren – Peers nicht offenlegen, dass ihr Urteil vorwiegend auf Publikations- und Zitationslisten beruht, oder wenn die Vorauswahl der Kandidierenden ausschliesslich auf bibliometrischen Daten beruht, ist eine Kumulation der Fehler quantitativer und qualitativer Beurteilungsverfahren nicht zu vermeiden.

5. Die Vergabe von Forschungsmitteln auf dem Prüfstand

In den letzten Jahren hat sich die Bedeutung von Drittmitteln als Finanzierungsquelle stark erhöht. Geplant ist, immer mehr Basis-Finanzierung durch Drittmittel zu ersetzen (Hornbostel 2001; Teichler 2005).²² Deren Höhe wird als ein Indikator für Forschungsleistungen angesehen. Begründet wird dies damit, dass die Vergabe von Drittmitteln auf Peer Reviews beruht (z.B. Deutscher Wissenschaftsrat 2005). Knappe Mittel sollen auf diese Weise effizienter eingesetzt und leistungssteigernde Anreize für die ForscherInnen gesetzt werden. Zusätzlich gehen im W-Besoldungssystem die eingeworbenen Drittmittel als Leistungskriterium ein. Die Bedeutung der Peer Reviews

²² Der Anteil drittmittelfinanzierter Forschung an deutschen Universitäten ist von 1980 bis 2000 von 22 Prozent auf 28 Prozent gestiegen, vgl. Jansen u.a. (2007). Geplant ist, Drittmitteln bei der leistungsorientierten Budgetierung verstärkte Relevanz zuzumessen vgl. Jaeger (2005).

nimmt auf diese Weise doppelt zu: bei der Ressourcenausstattung und bei der Besoldung. Die Begutachtung von Forschungsanträgen und ihre Auswirkungen sind jedoch selten untersucht worden (Bornmann & Daniel 2003; Jansen u.a. 2007).

5.1. Probleme bei der Begutachtung von Forschungsanträgen

Die wenigen Studien zur Begutachtung von Forschungsanträgen zeigen gemischte Ergebnisse (Neidhardt 1988; Wessely 1998; Weller 2001; Bornmann & Daniel 2003). Einige Autoren kommen zum Schluss, dass das Schicksal eines Antrages zur Hälfte von seiner Qualität und zur Hälfte vom Glück abhängt (Cole 1992). Andere finden eine befriedigende Korrelation zwischen der Gutachter-Reliabilität und der Akzeptanz von Anträgen (z.B. Neidhardt 1988; Bazeley 1998) und eine mittlere Korrelation zwischen erfolgreichen Forschungsanträgen und akademischem Erfolg (z.B. Daniel 1993). Aber die Kausalität ist unklar (Bornmann & Daniel 2003). Einige Autoren stellen Patronage fest (Over 1996; Wenneras & Wold 1999). So weist Neidhardt (1988) für das Gutachter-Verfahren bei der Deutschen Forschungsgemeinschaft (DFG) nach, dass ihre Gutachter besser abschneiden, wenn sie selber Anträge stellen. Andere Autoren widersprechen dieser Kritik mit dem Argument, dass Vetternwirtschaft ein Artefakt sein könnte „from the better applicants coming from the better institutions which supply the better reviewers“ (Wessely 1998:302).²³ Der resultierende kumulative Vorteil wäre jedoch nur dann fair zu nennen, wenn der anfängliche Zugang zu den Ressourcen ausschliesslich von der Leistung abhängen würde. Genau das stellen Viner, Powell & Green (2004) anhand britischer Daten in Frage.

5.2. Probleme bei den Auswirkungen von falschen Gutachterentscheidungen

Gillies (2005a, 2005b) diskutiert unter Bezugnahme auf die Theorie der statistischen Tests zwei Typen des Irrtums. Ein *Irrtum vom Typ I* entsteht, wenn ein Test zur Zurückweisung einer richtigen Hypothese führt. In diesem Fall würde ein Forschungsantrag abgelehnt, der wichtige Ergebnisse hätte bringen können. Ein bemerkenswertes Beispiel hierfür ist der Fall von Ignaz Semmelweis, dem Entdecker des Erregers des Kindbettfiebers im Jahr 1848 (Gillies 2005b). Die grosse Mehrheit der medizinischen Profession wies diese Entdeckung und die daraus folgenden praktischen Empfehlungen zurück. Semmelweis musste seine Wirkungsstätte am Wiener Hospital

²³ So begründet auch die Deutsche Forschungsgemeinschaft (DFG)(2003: 89) die hohe Korrelation zwischen DFG-Bewilligungen und der Zahl der DFG-Gutachter an einer Universität. Zur Kritik vgl. Münch (2006).

verlassen und tausende von Patientinnen starben. Erst um 1880 wurde seine Entdeckung anerkannt. Ein *Irrtum vom Typ II* entsteht, wenn ein Test zur Bestätigung einer falschen Hypothese führt. Dieser Irrtum ist in Bezug auf Forschungsanträge weniger schwerwiegend als der Irrtum vom Typ I. Er führt zur Verschwendung von Ressourcen, die in Wahrheit zu keinem brauchbaren Ergebnis führen.

Die meisten Begutachtungsverfahren konzentrieren sich im Sinne einer effizienten Forschungsförderung auf die Vermeidung des Fehlers vom Typ II. Die viel schwerwiegenderen Fehler vom Typ I können angesichts der Serendipitätseffekte und anderer Ungewissheiten in der Forschung kaum ausgeschaltet werden. Um die negativen Konsequenzen zu mildern, plädiert Gillies (2005a) deshalb dafür, dass *jedem* Forschungsansatz, der definierte Eingangshürden genommen hat, zumindest ein Minimum an Förderung zugesprochen werden sollte.

5.3. Probleme bei der Verteilung von Forschungsmitteln bei richtigen Gutachterentscheidungen

Aber auch wenn Gutachter tatsächlich in der Lage wären, die richtigen Entscheidungen zu treffen, entstehen weitere Probleme. Es bleibt völlig offen, was daraus zu schliessen ist. Sollten den für schlecht befundenen Projekten und Institutionen die Mittel gekürzt werden? Oder sollten ihnen nicht *zusätzliche* Mittel bewilligt werden, damit sie ihre Qualität erhöhen können? Sollte umgekehrt den für gut befundenen Projekten und Institutionen die Mittel gekürzt werden, weil sie ja ohnehin erfolgreich sind? Führen bei ihnen zusätzliche Mittel möglicherweise zu einem abnehmenden Grenzertrag? Eine sinnvolle Evaluation müsste den *marginalen Effekt* einer Änderung der Mittel erfassen (Frey 2008). Jansen u.a. (2007) zeigen in einem empirischen Vergleich von Drittmittelanteilen in der Astrophysik, der Nanotechnologie und der Mikroökonomik, dass es einen abnehmenden Grenzertrag und sogar einen abnehmenden absoluten Ertrag steigender Drittmittelanteile gibt,²⁴ gemessen an der Anzahl der Publikationen.²⁵ Die AutorInnen führen dies darauf zurück, dass hohe Drittmittelanteile mit höherem Akquisitionsaufwand sowie mit grösseren Forschungsteams einhergehen. Dies verursacht relativ höhere Koordinationskosten, welche den positiven Effekt zusätzlicher WissenschaftlerInnen überwiegt. Das Ergebnis steht in Übereinstimmung mit Befunden aus der Teamforschung, dass bei hochkomplexen Aufgaben effiziente Teams

²⁴ Der kritische Schwellenwert variiert allerdings disziplinspezifisch stark.

²⁵ Die AutorInnen betonen, dass dieses Leistungskriterium durch weitere Indikatoren zu ergänzen sind.

vergleichsweise klein sind (Steward 2006). Jansen u.a. (2007) ziehen daraus den Schluss, dass ein steigender Drittmittelanteil zu unerwünschten, weil ineffizienten Konzentrationstendenzen führen kann. Dies stimmt mit den Argumenten von Horrobin (1982, 1996) überein, dass „research empires“ nicht nur weniger effizient arbeiten als kleine, freiwillig vernetzte Einheiten. Vielmehr wird durch die damit verbundene Machtkonzentration auch die notwendige Vielfalt an Forschungsansätzen reduziert. Hartmann (2006) und Münch (2006) äussern ähnliche Bedenken in Bezug auf die Exzellenzinitiative. Als Konsequenz hieraus wäre eine sorgfältigere Abwägung der Nachteile von „research empires“ mit möglichen Grössenvorteilen notwendig.

6. Wie wirkt die Verstärkung monetärer Anreize?

Die Ausführungen haben gezeigt, dass die Funktionsfähigkeit des Peer Review Systems zur Implementation der Prioritätsregel von jeher eingeschränkt war. Die flächendeckende Einführung quantitativer Evaluationsverfahren hat dieses Problem nicht beseitigt, sondern eher vergrössert. Gleichwohl steigt die Bedeutung von Peer Reviews zusätzlich aufgrund der steigenden Abhängigkeit von Drittmitteln. Die Probleme verstärken sich, wenn die Begutachtungsergebnisse – wie in der variablen W-Besoldung vorgesehen – mit monetären Konsequenzen verknüpft werden.

Die neue W-Besoldung ist am 1.1.2005 obligatorisch für alle neu berufenen ProfessorInnen eingeführt worden. Sie enthält niedrigere Grundgehälter als die C-Besoldung. Dafür wurden zum ersten Mal so genannte leistungsorientierte, variable Elemente eingeführt, die über die bisher üblichen Funktionsbezüge und Bezüge aus Anlass von Berufungs- und Bleibeverhandlungen hinausgehen. Sie werden für besondere Leistungen in Forschung und Nachwuchsförderung sowie für die Drittmittelinwerbung gewährt. Die W-Besoldung bedeutet damit die Einführung eines „pay for performance“-Systems, das dem deutschen Hochschulsystem bisher fremd war.

Aus den geschilderten Besonderheiten der Wissensproduktion in der Forschung ergeben sich jedoch vier schwerwiegende Einwände.

- Der risikobehaftete Anteil des Einkommens wächst

Diese von der Politik durchaus beabsichtigte Wirkung entspricht nicht der dargelegten Besonderheit des Wissenschaftssystems. Unterstellt wird eine naive Annahme des Prinzipal-Agenten-Ansatzes (Jensen & Murphy 1990), dass nämlich monetäre Anreize immer zu einer Erhöhung der Anstrengung in die erwünschte Richtung führen. Diese Annahme hat sich schon im Bereich der Management-Entlohnung als äusserst fragwürdig

herausgestellt (Frey & Osterloh 2005; Rost & Osterloh 2008) und wird von einflussreichen Personalökonominnen zunehmend kritisiert (z.B. Lazear & Shaw 2007).

- Der Anreiz zu strategischem Verhalten wächst

Wird das monetäre Interesse von Forschern gestärkt, dann sollte damit gerechnet werden, dass rationale Forscher ihr Einkommen mit den Mitteln aufzubessern suchen, die am wenigsten Aufwand erfordern. Angesichts der hohen Risiken aufgrund der Prioritätsregel, der Serendipitätseffekte und der Zufallsprozesse im Peer-Review-Verfahren ist dies gerade *nicht* die Investition in innovative Forschung. Kommen zu den hohen Risiken für den Erfolg in der Forschung auch noch höhere Einkommensrisiken hinzu, so wird das Bemühen zunehmen, „to game the system“, nicht aber das Bemühen, in die Qualität der Forschung zu investieren. Forschende werden sich der weitaus sichereren Bemühung zuwenden, das System der Erfolgzuschreibung zu durchschauen und für sich zu instrumentalisieren. Dadurch entsteht ein Lock-in-Effekt, der es auch KritikerInnen dieses Systems immer schwerer macht, sich ihm zu entziehen. Es ist auch zu erwarten, dass das geschilderte opportunistische Verhalten von Reviewern, Editoren und Autoren zunimmt.

- Die Teamproduktion wird gestört

Ungleichheit des Einkommens bewirkt, dass die für wissenschaftliche Arbeit zunehmend wichtiger werdende Teamproduktion gefährdet wird. Die Literatur zu Leistungsturnieren zeigt, dass mit zunehmenden Leistungsprämien die Neigung zu Sabotageakten zunimmt (z.B. Lazear & Shaw 2007).

- Intrinsische Motivation wird verdrängt

Wie erwähnt, setzt kreative Arbeit intrinsische Motivation voraus (Stephan 1996; Amabile 1998; Osterloh 2007). Monetäre Anreize können, wie die Sozialpsychologie (z.B. Deci 1980) und die Psychologische Ökonomik (Frey 1997) zeigt, unter bestimmten Bedingungen diese Motivation verdrängen. Dies steht im Gegensatz zur Standard-Ökonomik, welche den theoretischen Hintergrund von „pay for performance“ Systemen darstellt. Sie nimmt an, dass variable Gehaltsbestandteile immer die Leistung erhöhen, d.h. einen Preiseffekt erzeugen. Sie berücksichtigt nicht, dass gleichzeitig zum Preiseffekt ein Verdrängungseffekt stattfindet, der zu einer Verringerung der Motivation in die erwünschte Richtung führt, sofern (a) die Tätigkeit ursprünglich intrinsisch motivierend war, (b) die Entlohnung als kontrollierend empfunden wird und (c) die durch die Entlohnung verstärkte extrinsische Motivation den Verlust an intrinsischer Motivation nicht ausgleicht (vgl. Osterloh & Weibel 2008). Der Verdrängungseffekt beruht in erster Linie auf der Verringerung der empfundenen Autonomie (Gagne & Deci 2005). Daraus folgt, dass

intrinsische Motivation in erster Linie durch Erhöhung der Autonomie gesteigert werden kann. Gleichwohl ist auch für die intrinsische Motivation die Höhe der Entlohnung nicht nebensächlich. Eine als fair empfundene fixe Entlohnung wird als Signal für Wohlwollen und Wertschätzung wahrgenommen und mit höherer freiwilliger Kooperation beantwortet als eine variable Entlohnung (Irlenbusch & Sliwka 2003; Kuvaas 2006).

Die W-Besoldung setzt in erster Linie auf den kontrollierenden Effekt. Es ist deshalb zu erwarten, dass die für die Forschung und die Kooperation in Teams zentrale intrinsische Motivation geschwächt wird und die Orientierung an monetären Entgelten zunimmt. Zudem wird durch diese Besoldungsform den WissenschaftlerInnen das Vertrauen entzogen, eigenverantwortlich eine hohe Leistung zu erbringen. Diese Misstrauenskundgebung verringert die Loyalität zur beschäftigenden Institution.²⁶ In der Folge ist mit einer Zunahme der Neigung zu rechnen, die Schwächen des Peer-Review-Systems für sich zu instrumentalisieren.

7. Ein radikaler Vorschlag

Angesichts der schwerwiegenden Mängel des Peer Review Systems stellen wir einen radikalen Vorschlag zur Diskussion²⁷. Er besteht aus drei Komponenten, die im Folgenden diskutiert werden. Es ist dies erstens die Zurücknahme von „pay for performance“ im Wissenschaftssystem. Zweitens schlagen wir vor, die Bedeutung des „Peer Review Systems“ zu beschränken und seinen Schwerpunkt in die Eingangsphase einer wissenschaftlichen Karriere zu verlagern. Drittens plädieren wir dafür, anstelle bloss monetärer Anreize auf die Gewährung von Autonomie als Anreiz zu setzen. Auf diese Weise wird dem „taste for science“ (Merton 1973) Rechnung getragen, welcher Voraussetzung für kreative wissenschaftliche Arbeit ist.

²⁶ In der Folge sinkt die Bereitschaft sich jenseits der Eigeninteressen für diese Institution zu engagieren, vgl. Falk & Kosfeld (2006); Osterloh & Weibel (2006). Dieser Schluss lässt sich aufgrund von Laborexperimenten ziehen, welche zeigen, dass Individuen unter bestimmten Bedingungen bei einer fixen Entlohnung eine höhere Leistungsbereitschaft und ein höheres Interesse am Wohlergehen des Arbeitgebers zeigen als bei variablen Löhnen, vgl. Irlenbusch & Sliwka (2003).

²⁷ Die Kritiker des Peer Review Systems haben bisher wenig Ideen für Reformen vorgelegt. So schlagen Ietto-Gillies (2008) und Hornbostel (2008) vor, das open-access-System zu verstärken. Das Problem der Verteilung von Ressourcen ist damit aber nicht gelöst. Atkinson (2001) fordert eine Überwachung des Review-Prozesses. Aber wer überwacht die Überwacher? Horrobin (1996) sieht die Lösung darin, alle staatlichen zur Verfügung stehenden Ressourcen zu gleichen Teilen auf die Forscher zu verteilen. Nach zehn Jahren soll die Fortsetzung der Förderung von Publikationen und Peer Reviews abhängig gemacht werden. Dies bewirkt lediglich ein Hinausschieben der Probleme, die Horrobin (1996) so hart kritisiert.

7.1. Rücknahme von „pay for performance“ im Wissenschaftssystem

Ein hoher variabler Anteil des Einkommens widerspricht den Besonderheiten der Wissensproduktion in der Forschung. Er erhöht das ohnehin hohe Risiko des Misserfolgs, verringert die Bereitschaft zur Wissensweitergabe im Team, erhöht den Anreiz, die Schwächen des Peer Review Systems strategisch auszunutzen, verdrängt die intrinsische Motivation und unterminiert insgesamt den „taste for science“. Hingegen sind die positiven Anreizwirkungen der W-Besoldung äusserst ungewiss. Es ist fraglich, ob WissenschaftlerInnen wie AkkordarbeiterInnen reagieren.²⁸ Da bekannt ist, dass Einkommen für Akademiker an den Hochschulen geringer sind als in Unternehmen, ist von vorneherein ein – durchaus funktionaler – Selbstselektionseffekt anzunehmen: An hohen Einkommen Interessierte wählen ohnehin keine Wissenschaftskarriere.

7.2. Eindämmung der Abhängigkeit vom Peer Review System

Im derzeitigen System sind Forscherkarrieren total und lebenslang von einem fehlbaren Review-System abhängig. Diese Abhängigkeit wird durch den „Evaluations-Hype“ (Weingart 2005:130) und durch die steigenden Drittmittelanteile an den Forschungsressourcen noch gesteigert. Um diese Abhängigkeit zu verringern, schlagen wir zwei Massnahmen vor:

Zum ersten lässt sich der Einfluss der Peer Reviews eindämmen, wenn die Forschenden und Lehrenden sorgfältig sozialisiert und ausgelesen werden und ihnen anschliessend eine hohe Autonomie gewährt wird.²⁹ Bei dieser zukunftsorientierten Strategie müssen durchaus qualitative und quantitative Peer Review Verfahren zum Einsatz kommen. Dies sichert bei aller Fehlbarkeit des Peer Review Systems, dass Standards der Wissenschaftlichkeit erfüllt werden, und gibt Hinweise auf das Potential der Kandidaten. Nach der Ernennung als ordentliche ProfessorIn nach strengen Kriterien muss darauf vertraut werden, dass die berufene Person die erwarteten Leistungen auch ohne ständige Kontrolle erbringt. Dadurch wird die für Wissenschaft unerlässliche Autonomie zumindest ab der Berufung auf eine ordentliche Professur gewährleistet. Deshalb sind Berufungsverfahren das mit Abstand wichtigste Geschäft einer wissenschaftlichen Institution. Dabei ist durchaus mit einer gewissen Varianz zu rechnen. Manche unter den

²⁸ Ein viel zitiertes Beispiel ist das US-amerikanische Unternehmen Safelite Glass, vgl. Lazear (2000): Nach der Umstellung von fixen Stundenlöhnen auf Stücklöhne stieg die Produktivität um 36% (Anzeizeffekt: 20% , Selektionseffekt: 16%), während die Lohnkosten nur um 9% stiegen. Mittlerweile warnen Lazear & Shaw (2007) vor der Generalisierung dieses Beispiels.

²⁹ Dies entspricht den von Ouchi (1980) sowie der in der umfangreichen kontrolltheoretischen Literatur empfohlenen Massnahmen, wenn sowohl Output- als auch Verfahrenskontrollen nicht ausreichen, vgl. z. B. Simons (1995); Schreyögg & Steinmann (1987).

Ausgewählten werden in ihrer Leistung nachlassen, andere hingegen werden durch den gewährten Freiraum beflügelt und zu Spitzenleistungen motiviert. In der Wissenschaft sollten letztere zählen. Unwillige und Versager müssen als notwendiges Übel betrachtet werden, damit das Wissenschaftssystem als Ganzes Höchstleistungen erbringen kann.

Das Prinzip der strengen Auslese und der anschließenden Gewährung von Autonomie ist an Universitäten wie der Harvard Universität lebendig. Es wurde von James Bryan Conant, dem Präsidenten der Harvard Universität folgendermassen zum Ausdruck gebracht (Renn 2002):

„There is only one proved method of assisting the advancement of pure science – that is picking men of genius, backing them heavily, and leaving them to direct themselves.“ (Letter to the New York Times, 13. August 1945).

Es ist Teil der „Principles Governing Research at Harvard“:³⁰

„The primary means for controlling the quality of scholarly activities of this Faculty is through the rigorous academic standards applied in selection of its members.“

ForscherInnen sind in Deutschland im Durchschnitt 41 Jahre alt, bis sie eine ordentliche Professur erlangen (Janson, Schomburg & Teichler 2007)³¹, sie durchlaufen also im Vergleich zu anderen Berufen eine ungewöhnlich lange Selektions- und Sozialisationsphase, in der ihre Autonomie eingeschränkt ist. Nach dieser Phase sind sie selbstverständlich nicht unabhängig, sondern müssen sich bei jeder beabsichtigten Publikation und bei jedem Forschungsantrag erneut dem Urteil der Peers unterwerfen. Aber erstens wird die fast vollständige Abhängigkeit vom Peer Review System stark reduziert. Zweitens wird der Druck zur Produktion quantitativ messbaren Outputs und damit die Belastung der Gutachter gemildert, was sorgfältigere Gutachten erwarten lässt. Drittens wird der beherrschende Einfluss von Indizes wie demjenigen des ISI und des SSCI abgebaut und damit die Gefahr der Homogenisierung der Forschung durch das Regime der Impact-Faktoren.

Zweitens schlagen wir vor, die Finanzierung der Forschung wieder zu einem höheren Anteil durch Grundausstattungen und zu einem geringeren Teil durch Einwerbung von Drittmitteln zu gewährleisten. Auch diese Massnahme verringert den Einfluss der Peer Reviews. Sie verhindert darüber hinaus das Entstehen von ineffizienten und die Vielfalt

³⁰ See <http://www.fas.harvard.edu/research/greybook/principles.html>.

³¹ In USA ist das Alter beim Erreichen einer Tenure-Professur mit 40 bis 41 Jahren etwa gleich hoch, vgl. Janson, Schomburg & Teichler (2007): 82.

reduzierenden „research empires“. Zusätzlich wäre eine grössere Dezentralisierung und Vielfalt von Fördereinrichtungen jenseits der dominierenden Deutschen Forschungsgemeinschaft wünschenswert (vgl. ebenso Neidhardt 2006).

7.3. Autonomie als Anreiz

Autonomie ist eine wichtige Voraussetzung für Kreativität (Amabile 1996, Gagne & Deci 2005). Sie ist aber zugleich Teil des Belohnungssystems in der Wissenschaft, welches dem „taste for science“ entspricht. ForscherInnen nehmen Einkommenseinbussen in Kauf, wenn sie dafür mehr Autonomie erhalten. Die Attraktivität von Universitäten sinkt, wenn zusätzlich zum niedrigeren Einkommen auch noch die Autonomie in der Forschung reduziert wird. Daraus resultieren zwei Vorschläge, die über die Abschaffung von „pay for performance“ und die verringerte Abhängigkeit vom Peer Review System hinausgehen.

Zum ersten ist das Ausmass an regelmässigen Evaluationen von Individuen zu reduzieren. Sie schränken die Autonomie ein, insbesondere wenn sie zu häufig und zu eng durchgeführt werden. Der Aufwand ist gross und noch grösser sind die verborgenen Kosten, welche durch reaktives Verhalten der Evaluierten entstehen (Frey & Osterloh 2007).

Zum zweiten sind die Evaluationen von Institutionen, welche zur Verteilung von Ressourcen unvermeidlich sind, in erster Linie prozess- und nicht outputorientiert durchzuführen. Angesichts der geschilderten Probleme des Peer-Review-Prozesses schlagen wir vor, dass sich die Evaluation von Institutionen auf folgende Fragen konzentriert: Ist ein sorgfältiger Prozess der Sozialisation und Selektion der WissenschaftlerInnen gesichert? Ist ein hoher Grad an Autonomie im Forschungsprozess gewährleistet?³²

Ein nach diesen Gesichtspunkten gestaltetes Anreizsystem hat der deutschsprachigen Wissenschaft zu Beginn des 20. Jahrhunderts Weltgeltung verschafft und die Attraktivität von „akademischen Leuchttürmen“ wie die Harvard University befördert. Es wird derzeit massiv durch ein mit dem Charakter der Forschung unvereinbares Anreizsystem verändert. Dieses kündigt nicht nur das institutionalisierte Vertrauen in die Selbststeuerung der „Republic of Science“ auf, sondern unterminiert auch

³² Ein prominentes Beispiel für ein nach diesen Prinzipien geführtes Forschungsinstitut war das sehr erfolgreiche Basler Institute of Immunology, welches nach einem rigorosem Auswahlprozess den Forschern einen grossen Spielraum gewährte, vgl. Melchers (1990), Neumann (2003).

den für kreative Forschung nötigen individuellen „taste for science“. Unser Vorschlag kann zwar nicht gänzlich das verloren gegangene Vertrauen in die Selbststeuerungsfähigkeit der Wissenschaft wiederherstellen, aber er kann gute Voraussetzungen dafür schaffen, dass die Attraktivität des deutschen Wissenschaftssystems für kreative ForscherInnen wieder zunimmt.

Literatur

Adler, R., Ewing, J., & Taylor, P., 2008. Citation Statistics, A Report from the Joint Committee on Quantitative Assessment of Research (IMU, ICIAM, IMS).

Amabile, T., 1996. Creativity in context: Update to the social psychology of creativity. Westview Press, Boulder.

Amabile, T., 1998. How to kill creativity. Harvard Business Review 76, 76-87.

Amabile, T.M., Conti, R., Coon, H., Lazenby, J., & Herron, M. 1994. The Context of Innovation. The Role of Work Environment, Waltham, Mass.: Brandeis University

Atkinson, M., 2001. "Peer Review" Culture. Science and Engineering Ethics 7, 193-204.

Bagues, Manuel, Mauro S. Labini, & Natalia Zinovyeva, 2008. Differential Standards and University Funding: Evidence from Italy. CESifo Economic Studies 54: 149-176.

Bazeley, P., 1998. Peer Review and Panel Decisions in the Assessment of Australian Research Council Project Grant Applications: What Counts is a Highly Competitive Context. Higher Education 35, 435-452.

Bedeian, A. G., 2003. The manuscript review process: The proper roles of authors, referees and editors. Journal of Management Inquiry 12, 331-338.

Bedeian, A. G., 2004. Peer review and the social construction of knowledge in the management discipline. Academy of Management Learning and Education 3, 198-216.

Bok, D., 2003. Universities in the Marketplace. The Commercialization of Higher Education. Princeton und Oxford: Princeton University Press.

Bornmann, L., & Daniel, H.-P., 2003. Begutachtung durch Fachkollegen in der Wissenschaft. Stand der Forschung zur Reliabilität, Fairness und Validität des Peer-Review-Verfahrens, in: Schwarz, S., Teichler, U. (Eds.), Universität auf dem Prüfstand. Konzepte und Befunde der Hochschulforschung. Campus, Frankfurt (Main), pp. 207-225.

Butler, L., 2003. Explaining Australia's increased share of ISI publications - the effects of a funding formula based on publication counts. Research Policy 32, 143-155.

- Campanario, J. M., 1996. Using Citation Classics to study the incidence of serendipity in scientific discovery. *Scientometrics* 37, 3-24.
- Campbell, D. T., 1976. Assessing the impact of planned social change. Occasional papers # 8. Social research and public policies: The Dartmouth/OECD Conference, Hanover, NH: Dartmouth College, The Public Affairs Center.
- Cicchetti, D. V., 1991. The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavioral and Brain Sciences* 14, 119-135, (Discussion 135-186).
- Cohen, W. M., & Levinthal, D. A., 1990. Absorptive capacity: a new perspective on learning and innovation, *Administrative Science Quarterly* 35, 128-152.
- Cole, S., 1992. *Making Science. Between Nature and Society*, Harvard University Press, Cambridge.
- Daniel, H.-D., 1993. *Guardians of science: Fairness and reliability of peer reviews*, Wiley, Weinheim.
- Darby, M.L., & Karni, E., 1973. Free competition and the optimal amount of fraud. *Journal of Law and Economics*, 16:67-88.
- Dasgupta, P., & David, P., A., 1994. Towards a new economics of science. *Research Policy* 23, 487-521.
- Deci, E.L., 1980. *The Psychology of Self-Determination*, Lexington, Mass.
- Deutsche Forschungsgemeinschaft (DFG), 2003. Förder-Ranking 2003. Institutionen - Regionen - Netzwerke. DFG-Bewilligungen und weitere Basisdaten öffentlich geförderter Forschung, Bonn.
- Deutscher Wissenschaftsrat, 2005. *Stellungnahme zu Leistungsfähigkeit, Ressourcen und Grösse universitätsmedizinischer Einrichtungen*.
- Deutscher Wissenschaftsrat, 2008. *Empfehlungen zum Forschungsrating, Mai 2008 (Drs. 8485-08)*
- Dogan, M., 1999. Marginality. *Encyclopedia of Creativity* 2: 179-184.
- Dosi, G., Marengo, L., & Pasquali, C., 2006. How much should society fuel the greed of innovators? On the relations between appropriability, opportunities and rates of innovation. *Research Policy* 35:1110-1121.

Durso T.W., 1997. Editor's advice to rejected authors: just try, try again. *The Scientist* 11, 13.

Ellison, G., 2002. The Slowdown of the Economics Publishing Process. *Journal of Political Economy* 110, 947-93.

Evidence Report, 2007. The use of bibliometrics to measure research quality in the UK higher education system <http://bookshop.universitiesuk.ac.uk/downloads/bibliometrics.pdf>.

Falk, A., & Kosfeld, M., 2006. The hidden cost of control. *American Economic Review* 96, 1611-1630.

Faria, J. R., 2005. The game academics play: editors versus authors. *Bulletin of Economic Research* 57: 1-12.

Frey, B. S., & Neckermann, S. 2008 Academics Appreciate Awards. A New View on Incentives in Research. Arbeitspapier, Institut für Empirische Wirtschaftsforschung Universität Zürich

Frey, B. S., 1997. Not just for the money: An economic theory of personal motivation, Edward Elgar, Cheltenham.

Frey, B. S., 2003. Publishing as prostitution? - Choosing between one's own ideas and academic success. *Public Choice* 116, 205-223.

Frey, B. S., 2008. Evaluitis - eine neue Krankheit, in: Matthies, H., Simon, D. (Eds.), *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen*. Leviathan Sonderheft 24, VS Verlag für Sozialwissenschaften, Wiesbaden, pp. 125-140.

Frey, B.S., & Osterloh, M 2005. Yes, Managers should be paid like Bureaucrats. *Journal of Management Inquiry*, 2005, Vol. 14, Nr. 1, S. 96-111.

Frey, B. S., & Osterloh, M., 2007. Evaluations: Hidden Costs, Questionable Benefits, and Superior Alternatives. Working Paper University of Zurich, SSRN download: <http://ssrn.com/abstract=928354>.

Frey, B.S., & Rost, K., 2008. Do rankings reflect research quality?, Working Paper University of Zurich.

Gagne M., & Deci, E.L., 2005 Self-determination theory and work motivation. *Journal of Organizational Behavior* 26:331-362.

Gans, J. S., & Shepherd, G. B., 1994. How are the mighty fallen: Rejected classic articles by leading economists. *Journal of Economic Perspectives* 8, 165-179.

Garfield, E., 1997. Editors are justified in asking authors to cite equivalent references from same journal. *British Medical Journal* 314, 1765.

Gaston J. 1971. Secretiveness and Competition for Priority of Discovery in Physics, *Minerva* 9/1.

Gillies, D., 2005a. Lessons from history and philosophy of science regarding the research assessment exercise. Paper read at the Royal Institute of Philosophy on 18 November 2005, download: www.ucl.ac.uk/sts/gillies.

Gillies, D., 2005b. Hempelian and Kuhnian approaches in the philosophy of medicine: The Semmelweis case. *Studies in history and philosophy of biological and biomedical sciences* 36, 159-181.

Gioia, D. A., & Corley, K. G., 2002. Being good versus looking good: Business school rankings and the circean transformation from substance to image. *Academy of Management Learning and Education* 1, 107-120.

Godoe, H., 2000. Innovation regimes, R & D and radical innovations in telecommunications. *Research Policy* 29, 1033-1046.

Gottfredson, S. D., 1978. Evaluating Psychological Research Reports: Dimensions, Reliability, and Correlates of Quality Judgements. *American Psychologist* 33, 920-934.

Hamermesh, D., 2007. Replication in Economics. IZA Discussion Paper No. 2760.

Hartmann, M., 2006. Die Exzellenzinitiative - ein Paradigmenwechsel in der deutschen Hochschulpolitik. *Leviathan* 34, 447-465.

Hirsch, J.E., 2005. An Index to Quantify an individual's scientific research output, <http://arXiv.org/abs/physics/0508025>

Hirschauer, S., 2004. Peer Review Verfahren auf dem Prüfstand. Zum Soziologiedefizit der Wissenschaftsevaluation. *Zeitschrift für Soziologie* 33/1: 62-83.

Holcombe, R. G., 2004. The national research council ranking of research universities: Its impact on research in economics. *Econ Journal Watch* 1, 498-514.

Holmstrom, B. P., & Milgrom, P. 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization*, 7(2): 24-52.

Hornbostel, S. 2001. Third party funding of German universities. An indicator of research activity? *Scientometrics* 50(3): 523-537.

Hornbostel, S., 2008. Neue Evaluationsregime? Von der Inquisition zur Evaluation, in: Matthies, H., Simon, D. (Eds), *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen*. Leviathan Sonderheft 24, VS Verlag für Sozialwissenschaften, Wiesbaden, pp. 59-82.

- Horrobin, D. F., 1982. In praise of non-experts. *New Scientist* 94, 1942-1944.
- Horrobin, D. F., 1996. Peer review of grant applications: a harbinger for mediocrity in clinical research? *Lancet* 348, 1293-1295.
- Ietto-Gillies, G., 2008. A XXI-century alternative to XX-century peer review. *Real World Economics Review* No. 45, 10-22.
- Irlenbusch, B., & Sliwka, D., 2003. Steigern variable Löhne die Leistung? *Arbeitspapier Institut zur Zukunft der Arbeit IZA, Bonn*.
- Jansen, D., Wald, A., Frenke, K., Schmoch, U., & Schubert, T., 2007. Drittmittel als Performanzindikator der wissenschaftlichen Forschung. Zum Einfluss von Rahmenbedingungen auf Forschungsleistung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 59(1):125-149.
- Janson, K., Schomburg, H., & Teichler, U., 2007. Wege zur Professur. Qualifizierung und Beschäftigung an Hochschulen in Deutschland und den USA. Münster u.a.: Waxmann.
- Jensen, C. M., & Murhpy, J. K., 1990. Performance Pay and Top-Management Incentives, *The Journal of Political Economy* 98(2), 225-264.
- Kieser, A., 1998. Going Dutch - Was lehren niederländische Erfahrungen mit der Evaluation universitärer Forschung? *Die Betriebswirtschaft* 58, 208-224.
- Küpper, H.-U., & Ott, R., 2002. Objektivierung von Hochschulrankings - Analyse ihrer Mess- und Bewertungsprobleme auf der Grundlage eines Vergleichs deutscher und US-amerikanischer Ranglisten. *Betriebswirtschaftliche Forschung und Praxis* 6: 614-630.
- Kuvaas, B., 2006. Work Performance, Affective Commitment, and Work Motivation: The Roles of Pay Administration and Pay Level. *Journal of Organizational Behavior*, 27, S. 365-385.
- Lawrence, P.A., 2003. The politics of publication - authors, reviewers, and editors must act to protect the quality of research. *Nature* 422, 259-261.
- Lazear, E.P., 2000. Performance Pay and Productivity. *American Economic Review*, 90, 1346-1361.
- Lazear, E. P. & Shaw, K., 2007. Personal Economics: The Economist's View of Human Resources. *Journal of Economic Perspectives*, 21(4),91-114.
- Lee, F.S., 2007. The Research Assessment Exercise, the state and the dominance of mainstream economics in British Universities. *Cambridge Journal of Economics* 31: 309-325.

Lind, G. (im Druck) Amerika als Vorbild? In: Bohl, T. & Kiper, H. (Hrsg.). Lernen aus Evaluationsergebnissen – Verbesserungen planen und implementieren. Bad Heilbrunn: Klinkhardt.

Lindsey, D., 1991. Precision in the manuscript review process: Hargens and Herting revisited. *Scientometrics* 22: 313-325.

Maasen, S., & Weingart, P., 2008. Unternehmerische Universität und neue Wissenschaftskultur. in: Matthies, H., Simon, D. (Eds), *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen*. Leviathan Sonderheft 24, VS Verlag für Sozialwissenschaften, Wiesbaden: 141-160.

Mahoney, M.J., 1977. Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System. *Cognitive Therapy Research* 1: 161-175.

Melchers, Fritz, 1990. The Basel Institute for Immunology. In: Arje Scheinin (Hrsg.) *The Foundation of New Technology, Conference on High Technology and Society*, Turku, Finnland: 35-41.

Merton, R.K., 1968. The Matthew Effect in Science. *Science*, Jan 5 1968, 159/3810: 56-63.

Merton, R.K., 1973. *The sociology of science: theoretical and empirical investigation*. University of Chicago Press, Chicago, IL.

Miner, L., & McDonald, S., 1981. Reliability of peer review. *Journal of the Society of Research Administrators* 12, 21-25.

Mudambi, R., Mudambi, S. & Navarra, P., 2007. Global innovation in MNCs: The effects of subsidiary self-determination and teamwork. *Journal of Product Innovation Management*, 24(5), 442-455.

Muller-Camen, M., & Salzgeber, S., 2005. Changes in academic work and the chair regime: The case of German business administration academics. *Organization Studies* 26, 271-290.

Münch, R., 2006. *Wissenschaft im Schatten von Kartell, Monopol und Oligarchie. Die latenten Effekte der Exzellenzinitiative*. Leviathan 34, 466-486.

Neidhardt, F., 1988. *Selbststeuerung in der Forschungsförderung: Das Gutachterwesen in der DFG*, Westdeutscher Verlag, Opladen.

Neidhardt, F., 2006. Fehlerquellen und Fehlerkontrollen in den Begutachtungssystemen der Wissenschaft. In: Hornbostel, Stefan & Simon, Dagmar (Hrsg.) *Wieviel (In-)Transparenz ist notwendig? Peer Review Revisited*. IFQ-working paper No. 1/ Dezember 2006: 7-13.

Nelson, R. R., 1959. The simple economics of basic scientific research. *Journal of Political Economy* 67, 297-306.

Nelson, R. R., 2004. The market economy, and the scientific commons. *Research Policy* 33, 455-471.

Nelson, R. R., 2006. Reflections on "The Simple Economics of Basic Scientific Research": looking back and looking forward. *Industrial and Corporate Change*, 15/6:903-917).

Neumann, Ralf ,2003. Gespräch mit Fritz Melchers: Im Chaos ist alles möglich. *Laborjournal* 10/2003: 28-32.

Osterloh, M., 2007. Human resources management and knowledge creation, in: Nonaka, I., Kazuo, I. (Eds.), *Handbook of Knowledge Creation*. Oxford University Press, Oxford, pp. 158-175.

Osterloh, M. & Frey, B. S., 2000. Motivation, knowledge transfer, and organizational forms. *Organization Science* 11, 538-550.

Osterloh, M. & Lüthi, R. (im Druck) Commons without Tragedy. Das Beispiel Open Source Software, in; Hoffmann-Riem, Wolfgang und Eifert, Martin (Hrsg.): *Geistiges Eigentum und Innovation*.

Osterloh, M. & Rota, S., 2007. Open source software production - Just another case of collective invention? *Research Policy* 36, 157-171.

Osterloh, M. & Weibel, A., 2006. *Investition Vertrauen. Prozesse der Vertrauensentwicklung in Organisationen*, Gabler, Wiesbaden.

Osterloh, M., & Weibel, A., 2008. Managing Motivation – Verdrängung und Verstärkung der intrinsischen Motivation aus Sicht der psychologischen Ökonomik. *Wirtschaftswissenschaftliches Studium* 8, 406-411.

Oswald, A. J., 2007 An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers. *Economica* 74: 21-31.

Ouchi, W. G., 1980. Markets, bureaucracies and clans. *Administrative Science Quarterly* 25, 129-141.

Over, R., 1996. Perceptions of the Australian Research Council Large Grants Scheme: differences between Successful and Unsuccessful Applicants. *Australian Educational Researcher* 23, 17-36.

Peters, D., & Stephen, C., 1982 Peer Review Practices of Psychological Journals: The Fate of Published Articles, submitted Again. *The Behavioral and Brain Sciences* 5: 187-195.

- Pfeffer, J., & Langhton, N., 1993. The Effect of Wage Dispersion on Satisfaction, Productivity, and Working Collaboratively: Evidence from College and University Faculty. *Administrative Science Quarterly* 38, 382-407.
- Polanyi, M. 1962. The Republic of Science: Ist Political and Economic Theory, *Minerva* 1/1: 54-73.
- Prichard, C., & Willmott, H., 1997. Just how managed is the McUniversity? *Organization Studies* 18, 287-316.
- Reedijk, J., 1998. Sense and nonsense of science citation analyses: comments on the monopoly position of ISI and citation inaccuracies. Risks of possible misuse and biased citation and impact data. *New Journal of Chemistry* 22(8): 767-770.
- Renn, J., 2002. Challenges from the past. Innovative structures for science and the contribution of the history of science. Paper presented at the Max Planck Forum 5, Innovative Structures in Basic Decision Research, Ringberg Symposium, München.
- Romer, P. M. 1994. The Origins of Endogeneous Growth, *Journal of Economic Perspectives*, 8/1: 3-22.
- Rosenberg, N. 1974. Science, Innovation, and economic growth. *Economic Journal* 84/333: 90-108.
- Rost, K., & Osterloh, M., 2008 Unsichtbare Hand des Marktes oder unsichtbares Handschütteln? Wachstum der Managerlöhne in der Schweiz. *Die Unternehmung*, 3/, S. 274- 304.
- Schreyögg, G., & Steinmann, H., 1987. Strategic Control. A New perspective. *Academy of Management Review* 12, 91-103.
- Siegel, P. A., & Hambrick, D. C., 2005. Pay Disparities within Top Management Groups: Evidence of Harmful Effects of Performance of High-Technology Firms. *Organization Science* 16, 259-274.
- Simkin, M. V., & Roychowdhury, V. P., 2005. Stochastic modeling of citation slips. *Scientometrics* 62(3): 367-384.
- Simons, R., 1995. Control in an age of empowerment. *Harvard Business Review* 73, 80-88.
- Smith, R., 1997. Journal accused of manipulating impact factor. *British Medical Journal* 314, 463.
- Sonnert, G., 1995. What makes a good Scientist? Determinants of peer evaluation among biologists. *Social Studies of Science* 25: 35-55.

- Sorenson, O., & Fleming, F., 2004. Science and the diffusion of knowledge. *Research Policy* 33, 1615-1633.
- Starbuck, W. H., 2006. *The Production of Knowledge. The Challenge of Social Science Research*. Oxford University Press, Oxford.
- Stephan, Paula E., 1996 *The Economics of Science*. *Journal of Economic Literature* 34, September: 1199-1235
- Stephan, P. E., 2008. Science and the University: Challenges for Future Research. *CESifo Economic Studies*, 54: 313-324.
- Stern, S., 2004. Do scientists pay to be scientists? *Management Science* 50, 835-853.
- Steward, G. L. 2006. A Meta-Analytic Review of Relationships between Team Design Features and Team Performance. *Journal of Management* 32:29-54.
- Stokes, D. E., 1997. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington D.C.: Brookings Institution Press.
- Teichler, U., 2005. *Hochschulstrukturen im Umbruch*, Campus, Frankfurt/New York.
- Torgler, B. S., Schmidt, S. L., & Frey, B. S., 2006. Relative Income Position and Performance: An Empirical Panel Analysis. Working Paper Series No. 282, Institute for Empirical Research in Economics, University of Zürich.
- Ursprung, H. W., & Zimmer, M., 2006. Who is the „Platz-Hirsch“ of the German Economics Profession? A Citation Analysis. *Jahrbücher für Nationalökonomie und Statistik* 227/2:187-202.
- Viner, N., Powell, P., & Green, R., 2004. Institutional biases in the award of research grants: a preliminary analysis revisiting the principle of accumulative advantage. *Research Policy* 33, 443-454.
- Weingart, P., 2005. Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics* 62, 117-131.
- Weller, A. C., 2001. *Editorial Peer Review. Its Strength and Weaknesses*, Information Today Inc., New Jersey.
- Weller, Ch., 2004. Beobachtungen wissenschaftlicher Selbstkontrolle. Qualität, Schwächen und die Zukunft des Peer-Review-Verfahrens. *Zeitschrift für Internationale Beziehungen* 11/2: 365-394.
- Wenneras, C., Wold, A., 1999. Bias in peer review of research proposals in Peer Reviews in Health Sciences, in: Godlee, F., Jefferson, T. (Eds.), *BMJ Books*, London, pp. 79-89.

Wessely, S., 1998. Peer review of grant applications: what do we know? *The Lancet* 352, 301-305.

Zuckerman, H.A., 1992. The Proliferation of Prizes: Nobel Complements and Nobel surrogates in the Reward System of Science. *Theoretical Medicine* 13: 217-231.