

Unternehmen Wissenschaft?

Alfred Kieser

Zusammenfassung: Der Aufsatz befasst sich mit den Auswirkungen der Ökonomisierung der Universität auf die Forschung, vor allem mit den Auswirkungen einer Quantifizierung von Forschungsleistungen und der Vorgabe von Anreizen zur ihrer Steigerung. Es wird gezeigt, dass Impact Factors, die zur Gewichtung von Zeitschriftenaufsätzen herangezogen werden, die wissenschaftliche Bedeutung von Zeitschriften nicht valide erfassen. Vor allem bewirken sie eine Änderung des Verhaltens der Wissenschaftler. Intrinsische Motivation wird zunehmend durch extrinsische verdrängt; Wissenschaftler sind nicht mehr bestrebt, neue wichtige Erkenntnisse zu gewinnen und die Entwicklung der Wissenschaft voranzutreiben, sondern sammeln Punkte, indem sie Bewährtes variieren. Zulagen für Forschungsleistungen verstärken die negativen Effekte einer bloßen Messung des Forschungsoutputs und darauf aufbauender Rankings. Der Aufsatz schließt mit einigen radikalen Vorschlägen zur Abwendung der negativen Auswirkungen der Ökonomisierung der Forschung.

Schlüsselwörter: Impact Factors · Leistungszulagen in der Forschung · Motivation von Wissenschaftlern · Rankings · Peer Reviews

Science as business?

Abstract: This article discusses the effects of managerializing the university, in particular of a quantification of research performance and the provision of incentives for increasing it. It is shown that impact factors do not provide a valid measure of journals' scientific impact.

They influence scientists' behaviour with regard to research. Intrinsic motivation is increasingly replaced by extrinsic motivation. Scientists no longer strive to gain important insights and to advance science, they rather collect points by varying well-tried approaches.

Incentives for research performance amplify the negative effects of a quantification of research output and of the rankings that follow from it. The article concludes with some radical proposals on how to reduce the negative effects of managerializing research

Keywords: Impact factors · Incentives for research performance · Motivation of scientists · Rankings · Peer reviews

© VS Verlag für Sozialwissenschaften 2010

Prof. (em.) Dr. Dr. h.c. A. Kieser (✉)
Lehrstuhl für Allg. BWL und Organisation,
Universität Mannheim, 68131 Mannheim, Deutschland
E-Mail: kieser@bwl.uni-mannheim.de

1 Einführung

In seiner Antrittsvorlesung *Was heißt und zu welchem Ende studiert man Universalgeschichte?* 1789 an der Universität Jena unterschied Schiller (1789, S. 106) zwei Typen von Gelehrten: den „Brodgelehrten“ und den „philosophischen Kopf“.

Dem Brodgelahrten ist es „einzig und allein darum zu thun ..., die Bedingungen zu erfüllen, unter denen er zu einem Amte fähig und der Vortheile desselben theilhaftig werden kann.“ Er setzt „nur darum die Kräfte seines Geistes in Bewegung, um dadurch seinen sinnlichen Zustand zu verbessern und eine kleinliche Ruhmsucht zu befriedigen“. „[J]e weniger seine Kenntnisse durch sich selbst ihn belohnen“, fährt Schiller fort, „desto größere Vergeltung heischt er von außen ...“. Alle Bestrebungen des „philosophischen Kopfes“ sind dagegen „auf Vollendung seines Wissens gerichtet; seine edle Ungeduld kann nicht ruhen, bis alle seine Begriffe zu einem harmonischen Ganzen sich geordnet haben, bis er im Mittelpunkt seiner Kunst, seiner Wissenschaft steht, und von hier aus ihr Gebiet mit befriedigtem Blick überschaut.“

Die erste These meines Vortrags lautet, dass die Ökonomisierung der Universität aus potentiellen „philosophischen Köpfen“, also aus intrinsisch motivierten Wissenschaftlern, extrinsisch motivierte „Brodgelehrte“, oder besser: „Brodhochschullehrer“, macht. Meine zweite These ist, dass die Ökonomisierung der Universität zu einer Hierarchisierung der Universitäten und zu einer Absenkung des Ausbildungsniveaus auf nationaler Ebene führt.

Die Ökonomisierung der Universität hat viele Aspekte (Müller-Böling 2000). Effizienz, Exzellenz, Herstellung von Wettbewerb zur Selektion akademischer Leistungen und Schaffung von Anreizen zur Verbesserung von Forschung und Lehre sind zentrale Schlagworte dieses Ansatzes. Es kann keine Frage sein, dass Kostenrechnung, strategische Planung oder Spendeneinsammeln („Fund Raising“), um nur einige Funktionen der ökonomisierten Universität zu nennen, sinnvolle Unterfängen darstellen können. Im Folgenden möchte ich mich jedoch auf einen Ausschnitt der Debatte zur Ökonomisierung der Universität konzentrieren: auf die Frage, ob es der Wissenschaft nutzt, wenn man versucht, die Leistungen von Wissenschaftlern exakt zu messen, mit Marktwerten zu versehen und für ihre Steigerung Anreize in Aussicht zu stellen.

Man muss an dieser Stelle vielleicht darauf hinweisen, dass die deutsche Universität im Laufe der Zeit immer wieder Strukturbrüchen unterworfen war. Schiller hielt seine Antrittsvorlesung an einer Universität des Landesherrn, die aus der Universität der Kirche hervorgegangen war. In der Aufklärung bildete sich die Universität des Beamten- und Bürgertums heraus, aus der dann wiederum die Ordinarienuniversität entstand, auf welche, nach 1968, die Gruppenuniversität folgte. Die Reformen der 90er Jahre mündeten schließlich in die ökonomisierte Hochschule.

Um die Wirkung der Ökonomisierung auf die Wissenschaft zu analysieren, gehe ich auf die folgenden Fragen ein:

- Was motiviert Wissenschaftler? Sichert hohe Motivation Erfolg?
- Wie wird Leistung in der Wissenschaft gemessen?
- Kommen wirklich die besten Aufsätze in die Top-Zeitschriften?

- Vermessen Rankings das Wissenschaftssystem? Oder schaffen sie ein neues Wissenschaftssystem, indem sie es zu messen vorgeben?
- Verbessern Leistungszulagen in der Wissenschaft die Leistung?
- Führen Rankings zu einer stärkeren Hierarchisierung von Fakultäten und Universitäten?
- Was ist zu tun? Welche Konsequenzen legt meine Analyse nahe?

2 Was motiviert Wissenschaftler? Sichert hohe Motivation Erfolg?

In seinem Vortrag „Wissenschaft als Beruf“ aus dem Jahre 1917 führt Weber (1994, S. 6) aus: „Ohne diesen seltsamen, von jedem Draußenstehenden belächelten Rausch, diese Leidenschaft, dieses: ‚Jahrtausende mußten vergehen, ehe du ins Leben tratest, und andere Jahrtausende warten schweigend‘ – darauf, ob dir diese Konjektur gelingt, hat einer den Beruf zur Wissenschaft *nicht* und tue etwas anderes. Denn nichts ist für den Menschen etwas wert, was er nicht mit Leidenschaft tun kann.“ Der „philosophische Kopf“ Schillers und Webers Wissenschaftler aus Leidenschaft sind Seelenverwandte.

Richard Feynman, amerikanischer Physik-Nobelpreisträger (1965), bringt die Motivation des Wissenschaftlers wesentlich prosaischer auf den Punkt: „Science is like sex. Sure, it may give some practical results, but that’s not why we do it.“ (Feynman et al. 1985, S. 283)

Es ist vor allem die Tätigkeit und erst in zweiter Linie der Erfolg, die den Forscher motivieren. Denn Erfolg in der Forschung ist höchst unsicher (s. dazu Osterloh u. Frey 2008): Die meisten Wissenschaftler machen nie sensationelle Entdeckungen, veröffentlichen nie Aufsätze in absoluten Spitzenzeitschriften, schreiben nie Bücher, die Furore machen, und werden nicht als Keynote Speaker zu internationalen Konferenzen geladen. Wenige heimsen großen Ruhm ein. Die Wissenschaft ist ein „Winner-Take-All-Spiel“ wie beispielsweise der Profisport oder die Popmusik (vgl. Frank u. Cook 1995): Es gibt Superstars und Kärner. Es herrscht der Matthäus-Effekt (Merton 1985): „Denn wer da hat, dem wird gegeben werden, dass er Fülle habe. Wer aber nicht hat, dem wird auch das genommen, was er hat.“ Dennoch macht den meisten Wissenschaftlern das Forschen Spaß.

Auf das hohe Risiko des Forschens ging auch Weber in seiner berühmten Rede ein: „Nun ist es aber Tatsache: daß mit noch so viel von solcher Leidenschaft, so echt und tief sie sein mag, das Resultat sich noch lange nicht erzwingen läßt.“ (Weber 1994, S. 6) Was er im Anschluss an diese Feststellung ausführt, verdient heute besondere Aufmerksamkeit: „Es ist ja wohl heute in den Kreisen der Jugend die Vorstellung sehr verbreitet, die Wissenschaft sei ein Rechenexempel geworden, das in Laboratorien oder statistischen Kartotheiken mit dem kühlen Verstand allein und nicht mit der ganzen ‚Seele‘ fabriziert werde, so wie ‚in einer Fabrik‘. ... Was bei einem solchen Vorgehen ‚schließlich herauskommt, ist oft blutwenig“ (S. 6).

Wie ich später ausführe, ist es genau das, was die Ökonomisierung der Wissenschaft dem Forscher nahe legt: Bewährtes „mit dem kühlen Verstand“ variieren, um möglichst viele Veröffentlichungen in hochrangigen Zeitschriften zu schinden. Allerdings, und da behält Weber immer noch Recht, was bei diesem Vorgehen „schließlich herauskommt“, ist in aller Regel „blutwenig“.

3 Wie wird Leistung in der Wissenschaft gemessen?

Welche wissenschaftliche Leistung ist höher einzustufen, die Entdeckung des Penicillins durch Alexander Fleming oder die These Max Webers vom Protestantismus als Beförderer des Kapitalismus? Diese Frage ist nicht zu beantworten, weil sie schlicht unsinnig ist. Ebenso unsinnig ist die Frage, ob ein Aufsatz im *Journal of Marketing* höher zu gewichten sei als ein Aufsatz im *Journal of Finance*. Selbst innerhalb eines Fachs erweisen sich Forschungsleistungen häufig als unvergleichbar und damit als nicht in eine Rangordnung zu bringen. Ein Forscher löst ein bestimmtes Problem auf der Basis der Theorie X und mit der Methode Y; ein anderer bearbeitet ein ganz anderes Problem auf der Basis der Theorie U und mit der Methode V. Die Unmöglichkeit, Forschungsergebnisse in eine Rangfolge zu bringen, ist häufig die Ursache mühseligen Ringens in Berufungskommissionen: Die eine Gruppe sieht ein höheres Potential in der Forschung der Bewerberin X, die andere ist überzeugt, dass dem Bewerber Y bald der Durchbruch gelingen wird. Subjektive Einschätzungen sind unvermeidlich und in Diskursen meist nicht auflösbar.

Um aus solchen Konflikten herauszukommen, greift man gerne auf quantitative Kriterien zurück, indem man etwa darauf hinweist, dass der eine Kandidat eine umfangreichere Veröffentlichungsliste vorzuweisen hat als der andere. Häufig wird auch argumentiert, dass, wenn die Aufsätze der einen Kandidatin mit dem wissenschaftlichen Rang der jeweiligen Zeitschriften gewichtet werden, sie klar höher einzustufen ist als die der anderen Bewerber. Solange der wissenschaftliche Rang einer Zeitschrift einfach behauptet wird, haftet einer solchen Argumentation jedoch immer eine subjektive Tönung an. Der Eindruck der subjektiven Wertung verschwindet erst aus der Diskussion, wenn alle Vertreter eines Fachs um eine Einschätzung der Bedeutung der wissenschaftlichen Zeitschriften gebeten werden und diese Einschätzungen dann gemittelt werden (das ist der Ansatz des Verbands der Hochschullehrer für Betriebswirtschaft, s Hennig-Thurau et al. 2004). Allerdings scheint dann immer noch die gemittelte Subjektivität der vielen Einschätzungen durch. Das Kunststück einer weitgehenden „Objektivierung“ hat Garfield (2006) mit seinem Impact Factor geschafft. Dieser Indikator zur Gewichtung des wissenschaftlichen Einflusses von Zeitschriften wird heute von Thomson Science, einer Division der Agentur Reuters, für eine große Zahl ausgewählter Zeitschriften verschiedener Disziplinen ermittelt und in einer Datenbank gegen Entgelt zur Verfügung gestellt. Wissenschaftsbewertung ist Big Business geworden. Der Impact Factor beruht auf der Annahme, dass Wissenschaftler Werke anderer Wissenschaftler vor allem deshalb zitieren, weil sie auf deren Ergebnissen aufbauen, und dass somit eine Zeitschrift, deren Aufsätze häufiger zitiert werden, einen höher zu bewertenden Beitrag zur Wissenschaft leistet, und sie deshalb eine höhere Qualität beanspruchen kann, als eine Zeitschrift, auf deren Aufsätze Wissenschaftler weniger häufig zugreifen bzw. die sie weniger häufig zitieren (Armin u. Mabe 2000; Ewing 2006). Der Impact Factor bietet auf dieser Basis eine Gewichtung der Bedeutung wissenschaftlicher Zeitschriften, die auf den ersten Blick nicht auf subjektiven Einschätzungen einzelner Zeitschriften beruht, sondern gewissermaßen aus vielen kleinen Würdigungen wissenschaftlicher Leistungen gewonnen wird, die in der täglichen Forschungsarbeit anfallen. Diese Würdigungen, so die Überlegung, entstammen nicht dem Kalkül, sondern der Redlichkeit des Forschers, denn es ziemt sich, Ergebnisse, auf denen man aufbaut, durch Zitierung anzuerkennen. Die Forschungsleistung

eines Wissenschaftlers kann dann als Summe seiner mit den jeweiligen Impact Faktoren gewichteten Zeitschriftenveröffentlichungen ermittelt werden. Auf der Grundlage dieses Wertes kann man dann auch eine Rangordnung von Wissenschaftlern erstellen, sei es ein Ranking der Bewerber auf eine Professur oder eines aller Professoren einer Disziplin, das in einem Presseorgan veröffentlicht wird. Man kann die Häufigkeiten der Zitierungen auch direkt zu einem Ranking benutzen (s. z. B. Podsakoff et al. 2008), und die Position einer Fakultät für ein Fakultäten-Ranking kann man aus den Ranglistenpositionen der in ihr tätigen Wissenschaftler aggregieren. So geht etwa das *Handelsblatt* bei der Ermittlung des Rankings deutschsprachiger BWL-Professoren vor. Es werden dort allerdings nicht nur die Impact Factors betriebswirtschaftlich relevanter Zeitschriften des Social Science Citation Index (SSCI) und des Science Citation Index (SCI) berücksichtigt, sondern auch noch das Ranking des Erasmus Research Institute of Management, Rotterdam, sowie das über eine Befragung aller Mitglieder ermittelte Ranking des Verbands der Hochschullehrer für Betriebswirtschaft, in dem deutschsprachige Zeitschriften stärker berücksichtigt sind (o. Verf. 2009).

Die ersten beiden Plätze des *Handelsblatt*-Rankings für BWL belegen zwei Professoren der Fakultät für Betriebswirtschaftslehre der Universität Mannheim. Das ist aus Sicht eines Mannheimer Universitätsangehörigen höchst erfreulich, aber gleichwohl unsinnig. Dass beide in ihren Fächern jeweils international hoch renommierte Wissenschaftler sind, kann einem jeder Insider versichern. Was die beiden jedoch forschen und was sie veröffentlichen, ist nicht vergleichbar und nicht in eine Rangfolge zu bringen. Der eine forscht zu Marketing, der andere zu Banken und Finanzierung. Die Feststellung, dass der eine Nummer eins und damit besser als die Nummer zwei ist, ist ungefähr so sinnvoll wie die, dass der Golfer Tiger Woods im Vergleich mit dem Tennisspieler Roger Federer der bessere Sportler ist. Rangordnungen von Wissenschaftlern verschiedener Disziplinen und selbst von Wissenschaftlern einer Disziplin aufzustellen, die unterschiedliche Fragestellungen mit unterschiedlichen Ansätzen und Methoden verfolgen, ist unsinnig. Wenn aber Rankings von Wissenschaftlern unsinnig sind, dann auch auf ihnen aufbauende Rankings von Fakultäten und Universitäten.

Trotz ihrer Absurdität sind Rankings von Wissenschaftlern und Wissenschaftsinstitutionen ungemein populär. Berufungskommissionen, Dekane und Universitätspräsidenten richten ihre Entscheidungen nach ihnen aus. Ihre Popularität gründet vor allem darauf, dass sie den Prozess der Bewertung abkürzen und Konflikte vermeiden helfen. Man muss sich zur Beurteilung von Bewerbern nicht mehr unbedingt der Mühe unterziehen, Veröffentlichungen der zu Beurteilenden zu lesen und untereinander zu vergleichen, was, wie wir gesehen haben, eine kaum lösbare Aufgabe darstellt. Man multipliziert einfach die Aufsätze der Bewerber mit den zugehörigen Impact Faktoren und addiert die so ermittelten Punkte oder schaut gleich in Rankings. Dazu muss man nicht einmal ein Angehöriger des betreffenden Fachs, ja nicht einmal Wissenschaftler sein. Rankings verheißten „die Geheimnisse der Welt der Forschung und der internen Verteilung von Reputation und Belohnungen [zu] lüften und es Außenseitern [zu] erlauben, einen direkten Blick auf die internationale Stellung oder die provinzielle Isolierung ihrer lokalen Wissenschaftler zu werfen und [geben] ihnen so die Macht ..., unberechtigte Ansprüche auf Ruhm zu entlarven“ (Weingart 2005, S. 105). In einer gewissen Weise signalisiert der Impact Factor einer Zeitschrift den Wert eines in ihr publizierten Aufsatzes wie z. B. der Preis eines Schrankes

einen Wert signalisiert (Hogler u. Gross 2009). Und so wie ein Schreiner gut beraten ist, solche Schränke herzustellen, die sich gut verkaufen und einen hohen Preis erzielen, so ist ein Wissenschaftler gut beraten, eine Art von Forschung zu betreiben, die sich zu Aufsätzen verarbeiten lässt, die mit großer Wahrscheinlichkeit von hoch rangierenden Zeitschriften zur Veröffentlichung angenommen werden. Und wie das Topmanagement eines diversifizierten Unternehmens weiß, welche Gewinnbeiträge die einzelnen Geschäftsbereiche bringen, wissen nun die Dekane und die Universitätsspitze, welche Fakultäten und Institute „gut am Markt ankommende“ Forschung generieren, und richten ihre Entscheidungen zur Mittelausstattung oder zu Stellenbesetzungen danach aus. Die Strategie der ökonomisierten Universität zielt darauf ab, in ein optimales „Portfolio der Forschungswertschöpfung“ zu investieren. Diese, aus der Sicht der Universitätsleitungen „rational“ begründbare Strategie determiniert in einem hohen Maße die Arbeit der Wissenschaftler. Von ihnen wird erwartet zu forschen, was ihnen, ihrer Fakultät und ihrer Universität Ranking-Punkte bringt, nicht aber, was sie selbst als die Wissenschaft vorantreibende Projekte ansehen.

Allerdings: Impact Faktoren – wie alle Ranking Systeme – weisen einige empfindliche Begrenztheiten auf (Adler u. Harzing 2009; Adler et al. 2008; Neuhaus et al. 2009; Svensson 2010):

- Es werden nur Veröffentlichungen in bestimmten Zeitschriften berücksichtigt. Dies sind fast nur englischsprachige, vorwiegend US-amerikanische Zeitschriften.
- Buchveröffentlichungen und Beiträge in Sammelbänden werden nicht berücksichtigt.
- Es werden nur Zitierungen berücksichtigt, die innerhalb von zwei Jahren nach der Publikation erfolgen. Viele Aufsätze werden aber über einen längeren Zeitraum zitiert.
- Zeitschriften zu Spezialgebieten, die in der Regel eine geringere Zirkulation aufweisen, werden benachteiligt.
- Vor allem aber ist es mehr als fraglich, ob Zitierungen vorwiegend Anerkennung für die Übernahme wichtiger Ergebnisse anderer Forscher in die Forschung der Zitierenden zum Ausdruck bringen. Ein Forscher zitiert einen anderen Forscher nicht nur, wenn er dessen Ergebnisse übernimmt, sondern auch, wenn er ihn kritisiert, sich seiner Schule zuordnet, sich von ihm abgrenzt, durch Belesenheit zu beeindrucken versucht ... Gerne zitiert er auch diejenigen, die generell häufig zitiert werden, denn das erhöht die Wahrscheinlichkeit, dass er selbst zitiert wird.

Erstaunlich ist nun, dass, wie wir gleich sehen werden, sich in Zeitschriften mit einem hohen Impact Factor Aufsätze finden, die nicht oder kaum zitiert werden. Dies legt die Frage nahe, ob das System der Begutachtung von Aufsätzen so funktioniert, dass tatsächlich die besten in die renommiertesten Zeitschriften kommen.

4 Erscheinen wirklich die besten Aufsätze in den Top-Zeitschriften?

Es ist nicht nur fraglich, ob Impact Faktoren die wissenschaftliche Qualität von Zeitschriften erkennen lassen, sondern auch, ob der Abdruck eines Aufsatzes in einer renom-

mierten Zeitschrift – wie auch immer das Renommee bestimmt wird – etwas über die Qualität dieses Aufsatzes aussagt. Schon die für renommierte Zeitschriften typischen Ablehnungsquoten von über 90 % der eingereichten Aufsätze lassen Zweifel aufkommen. Ein Großteil der eingehenden Manuskripte wird nämlich, um die Inanspruchnahme von Gutachtern in Grenzen zu halten, vom Haupt- oder einem Mitherausgeber nach einer ersten Durchsicht in einem sog. „desk-reject-Verfahren“ abgelehnt. Es ist von einer nicht geringen Irrtumswahrscheinlichkeit dieser ersten, notgedrungen nicht ganz so gründlichen Siebung auszugehen. Zu einer ausführlichen Beurteilung der Aufsätze, die diese Selektion überleben, wählt der Herausgeber in der Regel zwei, mitunter auch drei oder vier Gutachter aus. Diese empfehlen – u. U. nach mehreren Runden von Kommentierungen und Überarbeitungen – Ablehnung oder Annahme des eingereichten Manuskripts. Die Übereinstimmung zwischen Gutachterurteilen zu eingereichten Zeitschriftenaufsätzen ist in den Naturwissenschaften höher als in den Sozialwissenschaften und bei Ablehnungen höher als bei Annahmen (Fischer 2003), insgesamt aber recht gering (Fischer 2003; Miller 2006). Die geringe Übereinstimmung in den Sozialwissenschaften lässt manche Autoren zu dem Schluss gelangen, man könne die Entscheidung über Annahme oder Ablehnung eines Manuskripts auch dem Wurf eines Würfels anvertrauen (Bedeian 2004; Horrobin 2001; s. dazu auch die umfassende Auswertung von Ergebnissen bei Osterloh u. Frey 2008). Auch korrelieren die Urteile von Gutachtern über die Qualität eines Manuskriptes nur schwach mit den späteren Zitierungen der entsprechenden Aufsätze (Fischer 2003; Gottfredson 1978; Singh et al. 2007; Starbuck 2005). Folgerichtig stellte der US Supreme Court unlängst fest, dass von dem Umstand, dass Fachgutachten durchgeführt werden, nicht auf die Qualität der Inhalte wissenschaftlicher Zeitschriften geschlossen werden könne (Horrobin 2001).

Besonders aufschlussreich ist eine Studie von Peters und Ceci (1982). Diese wählten aus 12 angesehenen psychologischen Zeitschriften (mit hohen Ablehnungsquoten und Impact Faktoren) je einen Artikel der letzten 2–3 Jahre aus und änderten deren Autoren- und Institutionennamen vom Hochreputierten ins Namenlose. Auch die Titel, Zusammenfassungen und Einleitungsabschnitte veränderten sie. Dann reichten sie den ansonsten unveränderten Aufsatz bei der gleichen Zeitschrift ein, die ihn bereits publiziert hatte. Das erstaunliche Ergebnis: Nur 3 der 12 Manuskripte wurden als bereits publiziert erkannt, und 8 der 9 übrigen wurden nun *abgelehnt*.

Die geringe Übereinstimmung der Gutachter ist keineswegs erstaunlich. Wie wir bereits festgestellt haben, zeichnen sich Wissenschaftler durch Vorlieben für bestimmte Theorien und Methoden aus. Darüber hinaus wenden sie bei ihrer Beurteilung unterschiedliche Kriterien an und gewichten diese unterschiedlich. „[T]wo referees commissioned to review the same manuscript actually read different works“ (Bedeian 2004, S. 281). Und sie sind auch deswegen nicht unvoreingenommen, weil sie in aller Regel Konkurrenten der Autoren um die knappen Seiten in Zeitschriften und um Drittmittel sind. Insbesondere in Fächern mit hoher Binnenkonkurrenz wird im peer-review-Verfahren gewissermaßen „die Lobby der Böcke zu den Wahrern der Interessen der Gärtner“ (Fischer 2003, S. 61). In dem Experiment von Peters und Ceci hat sicher auch eine Rolle gespielt, dass die Autoren ins „Namenlose“ geändert wurden. Die Autoren der „echten“ Einreichungen waren ausnahmslos recht bekannt. Bei in der Community geschätzten Autoren geben sich die

Herausgeber Mühe, Gutachter zu finden, die die gleichen Ansätze schätzen wie die Autoren und sie entscheiden sich bei widersprüchlichen Gutachtern eher für eine Annahme.

Gutachter haben nicht nur Vorlieben für bestimmte Theorien und Methoden. Sie sind vor allem konservativ. Sie sind konservativ, weil sie Karriere machen wollen: vom Gutachter zum Mitherausgeber, zum Hauptherausgeber und zum Herausgeber noch bedeutenderer Zeitschriften, und deshalb meiden sie das Risiko, einen Aufsatz zur Annahme zu empfehlen, der hoch innovativ ist, nicht an Bewährtem anknüpft und bei dem damit zu rechnen ist, dass er in der Gemeinschaft der Wissenschaftler auf Ablehnung stößt. Von Herausgebern wird erwartet, dass sie wissenschaftliche Innovationen fördern. Tatsächlich fördern sie in aller Regel die herrschende Meinung, den *Mainstream*.

Die Impact Faktoren sozialwissenschaftlicher Zeitschriften spiegeln aufgrund ihres Ansatzes eine starke Voreingenommenheit zugunsten amerikanischer Forschungstraditionen wider. Europäische englischsprachige Zeitschriften enthalten einen hohen Anteil an von europäischen Forschungstraditionen geprägten Aufsätzen, die von amerikanischen Autoren wegen der Inkompatibilität mit der eigenen Forschung weniger häufig zitiert werden. Das hat jedoch einen niedrigeren Impact Factor dieser Zeitschriften zur Folge. Diese Voreingenommenheit zugunsten amerikanischer Forschungstraditionen ereilt auch europäische Autoren, die von europäischen Traditionen geprägte Aufsätze (etwa qualitative Analysen beinhaltende oder an bestimmten Wissenschaftlern anknüpfende Konzeptionen), bei amerikanischen Zeitschriften einreichen: Ihre Aufsätze werden von mit diesen Traditionen weniger vertrauten amerikanischen Gutachtern mit einer höheren Wahrscheinlichkeit abgelehnt. Insofern ist der den Impact Faktoren zugrunde liegende Wettbewerb „not a ‚rational game‘ of academic quality conducted through the score sheet of citations, journal rankings and so on. It is a political game in which the dice are loaded so heavily as to make the outcome pre-ordained.“ (Grey 2010, S. 689)

Man muss als Autor also nicht nur einen guten Aufsatz schreiben, man muss vor allem das Glück haben, dass der Herausgeber nicht nur einen, sondern zwei oder drei Gutachter auswählt, bei denen der eingereichte Aufsatz nicht Regungen der Missgunst, des Hochmuts oder des Dogmatismus weckt und die eine gewisse Sympathie für den gewählten theoretischen Ansatz haben. Fachgutachten sind „the least imperfect way of upholding the quality of scientific publications“ (Herausgeber der naturwissenschaftlichen Zeitschrift *Nature* in einem Editorial, zit. n. Biagioli 2002, S. 36, FN 6)

Nicht zuletzt wegen des äußerst schwachen Zusammenhangs zwischen Gutachterbeurteilungen und Zitierungen eines veröffentlichten Aufsatzes ist der Impact Factor ein schlechter Indikator der Qualität. Und so sagen auch Additionen der mit diesem Indikator gewichteten Aufsätze eines Autors nichts über dessen wissenschaftliche Qualifikation aus (Adler et al. 2008; Fischer 2003; Gottfredson 1978; Singh et al. 2007; Starbuck 2005).

Trotzdem erfreuen sich Impact Faktoren großer Beliebtheit bei der Bewertung individueller Bewerber durch Berufungskommissionen und durch externe Gutachter, die häufig explizit auf Impact Faktoren und Rankings verweisen. Und solche Beurteilungen einzelner Forscher werden in vielen Fakultätsräten und Senaten anerkennend zur Kenntnis genommen, selbst wenn in diesen Gremien Kompetenz für statistische Analysen vertreten ist. Zu groß sind die vermeintlichen Vorteile einer Entscheidungsunterstützung durch Impact Faktoren: Objektivität, Eindeutigkeit, Konfliktvermeidung.

5 Vermessen Rankings das Wissenschaftssystem oder schaffen sie ein neues, indem sie es zu messen vorgeben?

Das größte Problem von Rankingsystemen aber ist, dass sie das Verhalten derjenigen, die von ihm betroffen sind, beeinflussen, geradezu zu Manipulationen einladen. Bruno Frey (2004) stellt sinngemäß fest: Eine Verfassung gegen Gauner produziert Gauner. Ich würde nicht sagen, dass ein System zur Kontrolle des Forschungsoutputs, also eine Verfassung gegen faule Professoren, faule Professoren produziert, es produziert aber Professoren, die ihren Einsatz für Forschung und Lehre einem Kalkül unterwerfen.

Im bereits erwähnten populären *Handelsblatt*-Ranking für Betriebswirtschaftsprofessoren sind die 200 – nach dem Maßstab des *Handelsblatt* – erfolgreichsten BWL-Professoren aus deutschsprachigen Ländern gelistet. Unter diesen finden sich keine Professoren, die ihre Forschungsergebnisse vorwiegend in Monografien veröffentlichen, mögen diese auch noch so einflussreich sein. Es werden nämlich im auch im *Handelsblatt*-Ranking, wie in Rankings üblich, nur Zeitschriften-Veröffentlichungen berücksichtigt. Die Folge ist, dass Wissenschaftler weniger Monografien schreiben, was ich persönlich sehr bedauerlich finde, denn als Student habe ich vor allem aus Monografien gelernt. Und wenn man mich jetzt fragt, welche Veröffentlichungen ich für die wichtigsten halte, führe ich fast nur Monografien auf. Für deutschsprachige Zeitschriften schreiben Managementforscher, weil das keine oder nur wenige Punkte bringt, höchstens noch Zweitaufgüsse ihrer englischsprachigen Veröffentlichungen oder Zweitrangiges.

Rankings ermitteln Unterschiede zwischen solchen Wissenschaftlern und Wissenschaftsinstitutionen, deren Leistungen so dicht beieinander liegen, dass sie kaum zu unterscheiden sind. Aber die in Rankings ermittelten Unterschiede bestätigen sich selbst und führen zu Vergrößerungen der Abstände: Schlechter gerankte Wissenschaftler oder Institutionen verlieren Ansehen bei Studierenden, Politikern und Hochschuladministratoren. Sie erleiden Einbußen bei der Ressourcenausstattung, ihre Rankingposition verschlechtert sich weiter Rankings wirken wie self-fulfilling prophecies.

Das System verleitet zu Trickereien (Macdonald u. Kam 2007). Wissenschaftler zitieren beispielsweise ihre eigenen Schriften noch häufiger als bisher. Auch zitieren sie mit Vorliebe Aufsätze, die in Zeitschriften mit einem hohen Impact Factor erschienen sind. Das verleiht ihren eigenen Aufsätzen Bedeutung. *Zeitschriften mit einem hohen Impact Factor haben also nicht nur deswegen ein hohes Prestige, weil sie häufig zitiert werden, sie werden auch häufig zitiert, weil sie ein hohes Prestige aufweisen.* Das System der Impact Factors bestätigt sich selbst, es ist selbst-referenziell (Macdonald u. Kam 2007).

Mit dem Versprechen, dass sie den Großteil der Arbeit übernehmen, veranlassen weniger berühmte Wissenschaftler berühmtere Kollegen zu gemeinsamen Veröffentlichungen. Das erhöht nicht nur die Wahrscheinlichkeit der Annahme des resultierenden Manuskripts zur Veröffentlichung – für den Herausgeber sind die Autoren nicht anonym –, sondern auch, wenn der Aufsatz erschienen ist, die des Zitiert-Werdens. Autoren verlegen sich auch auf das Verfassen besonders „zitierrächtiger“ Aufsätze wie Metaanalysen oder Rezensionenartikel, was ihnen durch Aufsätze nahe gelegt wird (z. B. durch Judge et al. 2007).

Wenn man früher einen Bewerber auf eine Professur fragte, was seine Forschungsinteressen sind, erhielt man Antworten der folgenden Art: „Mich interessiert, wie Arbeits-

gruppen funktionieren“ oder „Ich erforsche, ob Anreize die Leistung erhöhen.“ Heutige Kandidaten wollen in A-Journals veröffentlichen. Und sie wissen auch, was sie dazu tun müssen: eine Thematik, eine Theorie und eine Methode wählen, die en vogue sind. Mit anderen Worten: Sie machen genau das, was sich in der jüngeren Vergangenheit bewährt hat, halt ein bisschen variiert. Am besten wiederholt man heute eine Untersuchung mit einer etwas leistungsfähigeren Methode, auch wenn das in einem Missverhältnis zur Datenqualität stehen sollte. Fortschritte in der Methodik sind leichter zu konstatieren als Fortschritte in den Forschungsergebnissen. Wenn nämlich Wissenschaftler Ergebnisse wirklich innovativer Forschung wiedergeben, riskieren sie, dass die Gutachter ihren Aufsatz nicht mit ihnen bekannten Forschungsergebnissen in Verbindung bringen können und ihn ablehnen. Etliche Aufsätze, die später berühmt und viel zitiert wurden, die gar Forschungsergebnisse zum Inhalt hatten, die ihren Verfassern später zu Nobelpreisen verhelfen, erlitten zunächst das Schicksal einer Ablehnung (Campanario 1996; Gans u. Shepherd 1994).

Wenn es der Autor nicht schon von sich aus gründlich genug macht, dann sorgen Gutachter und Herausgeber dafür, dass ein eingereicherter Aufsatz zu einem Mainstream-Aufsatz wird. Um die Gutachter bloß nicht zu vergraulen und eine negative Entscheidung zu provozieren, machen die Autoren alles, was die Gutachter vorschlagen, auch wenn sie den Eindruck haben, dass dadurch ihre Aufsätze eher schlechter als besser werden. Zu Recht spricht Bruno Frey (2003) daher von der „Veröffentlichung als Prostitution“.

Und Wissenschaftler versuchen, aus ihren Forschungsprojekten so viele Aufsätze wie möglich zu pressen. Das geht so ähnlich wie bei den Kombinationsbilderbüchern, bei denen man mit jeweils einigen Streifen verschiedener Hüte, Gesichter, Bäuche und Beine ganz viele unterschiedliche Figuren erzeugen kann. Mit dieser Methode bringt es etwa ein noch unter 35jähriger Betriebswirt auf 36 internationale Veröffentlichungen in drei Jahren (o. Verf. 2010b) (Abb. 1).

Herausgeber und Verlage tricksen ebenfalls: Sie drängen Autoren, möglichst viele Aufsätze zu zitieren, die in ihrer Zeitschrift erschienen sind. Hierzu ein Beispiel: „We have noticed that you cite *Leukemia* [die Zeitschrift, bei der der Autor einen Beitrag eingereicht hat] only once in 42 references. Consequently, we kindly ask you to add references of articles published in *Leukemia* to your present article.“ (Editor, *Leukemia*, in einer Mitteilung an einen Autoren, zit. in Smith 1997) Fakultäten und Universitäten beteiligen sich an den Trickereien. So drängen Dekane und Präsidenten Wissenschaftler ihrer Institutionen, vor allem in Zeitschriften zu veröffentlichen, die populären Rankings wie dem der *Financial Times* zugrunde liegen (Connelly u. Gallagher 2007). Sie stellen gerne Bewerber ein – auch als Gastwissenschaftler für begrenzte Zeitspannen –, die in high impact-Zeitschriften veröffentlichen und viele Zitierungen versprechen, um dadurch ihre Platzierungen in Rankings zu verbessern.

So schaffen Rankings Wirklichkeit. Sie stellen nicht Transparenz auf einem bereits existierenden „Markt“ her, vielmehr generieren sie einen Pseudomarkt (Voswinkel 2005). Sie verzerren den Wettbewerb innerhalb der Gemeinschaft der Wissenschaftler, wie er vor der Einführung der Rankings existierte. Dieser war diffus, aber keineswegs wirkungslos. Wissenschaftler und ihre Institutionen standen in einem Wettbewerb um Reputation (Luhmann 1998) oder um die Akkumulierung sozialen Kapitals (Bourdieu 1998). In diesem Wettbewerb fanden verschiedene Formen von Veröffentlichungen (Bücher, Zeit-

Abb. 1: Kombinationstechnik auf Basis von Bilderstreifen



schriftenaufsätze, Beiträge zu Sammelwerken), Würdigungen der Kollegen untereinander, Auszeichnungen und Karrieresignale (Rufe) ihren Niederschlag.

Selbst Akteure, die Rankings ablehnen, können sich ihrem Einfluss kaum entziehen. So meint die Dekanin einer US Law School: „I was one of those [faculty] ... who would say that we should disregard the rankings completely, and we should do whatever we need to do to make this a great place for our students and faculty. Now [as dean], I don't think I have any choice but to think in terms of the rankings.“ (Sauder u. Espeland 2009)

Rankings regen Wissenschaftsinstitutionen zur Profilbildung an, verhindern diese aber zugleich. Profilbildung nach Maßgabe der Rankings heißt nicht, ein anderes Profil auszubilden als konkurrierende Fakultäten, sondern das gleiche Profil mit größerer Perfektion

anzustreben. Maasen und Weingart (2007) meinen zu Recht: „Ein System, in dem sich alle Akteure an den gleichen Indikatoren orientieren, läuft Gefahr, seine Diversität und Innovativität zu verlieren.“ Akteure des Systems sind Herausgeber, Gutachter und Universitätsleitungen, die alle danach streben, dass sich Rankingpositionen, die sie betreffen, verbessern. Und weil alle diese Akteure karriererelevante Entscheidungen fällen, wollen die Wissenschaftler genau die Punkte erringen, die im System angerechnet werden. Sie gehen nicht mehr Forschungsfragen nach, die sie im Hinblick auf den Fortschritt der Wissenschaft als wichtig erachten, sie sammeln Punkte für Ranglisten. Sie begeben sich nicht mehr auf wissenschaftliche Entdeckungsreisen, sondern folgen den in Rankings ausgeflaggten Trampelpfaden. Die extrinsische Motivation verdrängt die intrinsische; „Philosophische Köpfe“ mutieren zu „Brodgelehrten“. Aufsätze, in denen immer wieder die gleichen Theorien und Methoden für die Bearbeitung leicht variiertes Forschungsfragen zum Einsatz kommen, verdrängen das wissenschaftliche Buch; Fakultäten verfolgen ähnliche Forschungsprogramme und bieten ähnliche Studiengänge (wofür auch die Akkreditierungsinstitute sorgen). Kreativität ist für die Erlangung wissenschaftlichen Erfolgs nicht mehr in einem höheren Maße erforderlich.

James March (2004) nährt die Hoffnung, dass irgendwann Stumpfsinn und Verzweiflung im Getriebe der ökonomisierten Wissenschaft dermaßen überhand nehmen, dass „scholars may come to embrace a dream of a research community that recognizes the exploratory advantages of fragmentation into enclaves of high but differentiated quality at the same time as it endorses the multidisciplinary, multinational and multilingual contacts that exploit the advantages in the service of a more inclusive vision of knowledge“.

6 Verbessern Zulagen die Leistungen von Wissenschaftlern?

Die 2002 eingeführte W-Besoldung sieht vor, dass Professoren zur Steigerung ihrer Motivation variable Leistungsbezüge erhalten können (Süß 2007). Eine der Initiatorinnen dieses Besoldungssystems, die damalige Bundeswissenschaftsministerin Edelgard Bulmahn, schätzte, dass zehn bis 15 % aller Professoren „zu müde“ seien (Horstkotte 2010). Der Präsident der Hochschulrektorenkonferenz, Klaus Landfried, sprach von „Faulpelzen“ unter den Kollegen (o. Verf. 1999). Da die neue Besoldung nicht zu Mehrausgaben gegenüber der alten Besoldung führen darf, musste man, um Volumen für diese Prämien zu schaffen, die Grundgehälter kräftig senken.

Die W3-Besoldung beruht also auf der Annahme, viele Professoren leisteten bei einer fixen Besoldung weniger, als sie könnten. Lockte man sie mit Prämien, so die zweite Annahme, würden sie ihre Leistungen steigern. Aber: Vielleicht sind viele Hochschullehrer an zusätzlichen finanziellen Leistungen gar nicht sonderlich interessiert, weil die Maximierung von Zulagen nicht mit ihrem Berufsverständnis in Einklang zu bringen ist? Oder sie sind vorwiegend extrinsisch motiviert, verdienen aber mit Nebentätigkeiten bereits soviel, dass sie Leistungszulagen nicht weiter motivieren können?

Problematisch dabei ist vor allem die Annahme, dass Leistungen von Wissenschaftlern so gemessen werden können, dass genau die gewünschten Leistungen prämiert und negative Effekte vermieden werden. Denn schon allein die Kontrolle des Forschungsoutputs verändert, wie wir gesehen haben, das Verhalten der Wissenschaftler. Eine finanzielle

| | |
|----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Stufe 5 | Herausragende, international beachtete und maßgebliche Beiträge in Forschung und/oder Lehre, Nachwuchsförderung u. Weiterbildung von herausragenden Wissenschaftlern von internationaler und fachüberschreitender Reputation |
| Stufe 4 | Beiträge in Forschung, Lehre, Nachwuchsförderung und/oder Weiterbildung von sehr hohen Standards und/oder herausragender Bedeutung für deren Pflege und Entwicklung an der Universität |
| Stufe 3 | Dauerhaft sehr gute Beiträge zur Entwicklung von Forschung, Lehre, Nachwuchsförderung und/oder Weiterbildung |
| Stufe 2 | Besondere Beiträge in Forschung, Lehre, Nachwuchsförderung und/oder Weiterbildung und ein bemerkenswertes Engagement für deren Pflege und Entwicklung an der Universität |
| Stufe 1 | Leistungen in Forschung, Lehre, Nachwuchsförderung und/oder Weiterbildung, die über üblicherweise zu erwartenden Leistungen von Professoren hinausgehen |

Abb. 2: Leistungszulagen-Stufen für Professoren der Universität Konstanz

Prämierung intensiviert die aufgezeigten Effekte, vor allem auch den der Produktion wenig innovativer Mainstream-Forschung.

Anreizsysteme lösen „irrationale“ Entscheidungen aus. Die Universität Konstanz hat z. B. verschiedene Zulagen-Stufen formuliert (Abb. 2). Um den Vergaberahmen nicht zu sprengen, ist eine prozentuale Verteilung über die Stufen verbindlich vorgegeben (Abb. 3). Nur zwei Prozent der Professoren dürfen absolute Spitzenforscher – „Leuchttürme der Forschung“, wie das in Schriften zur Hochschulpolitik oft so schön heißt – sein. Man stelle sich vor, eine Fakultät schreitet zu einer Berufung und stellt fest, dass sie schon alle Prozentvorgaben erfüllt oder übererfüllt hat. Nur in der Kategorie 0 – der Kategorie, der nur das Grundgehalt zusteht – sind noch Plätze frei. Jetzt muss diese Fakultät also einen Wissenschaftler suchen, dessen Leistungen unterdurchschnittlich sind und mit großer Wahrscheinlichkeit auf absehbare Zeit so bleiben.

Generell ist es so, dass die Professoren einer Fakultät sich einen größeren Anteil des Vergaberahmens sichern können, wenn sie Kollegen berufen, die mit großer Wahrchein-

Abb. 3: Prozentvorgaben für die Leistungszulagen-Stufen der Universität Konstanz

| | | |
|----------------------|----------|--------|
| Spitzenzuschuss | X.XXX,00 | € 2 % |
| Stufe 5 | 700,00 | € 5 % |
| Stufe 4 | 600,00 | € 8 % |
| Stufe 3 | 500,00 | € 15 % |
| Stufe 2 | 400,00 | € 30 % |
| Stufe 1 | 300,00 | € 30 % |
| Stufe 0: Grundgehalt | | € 10 % |

lichkeit nie umfangreichere Leistungszulagen beanspruchen können. Professoren können sich aber auch, um ihre Aussichten auf Zulagen zu verbessern, an eine Fakultät bewerben, deren bereits vorhandene Professoren sich nur gering Chancen auf Leistungsanreize ausrechnen können. Zum Glück ist es nach meiner Kenntnis bislang bei solchen Gedankenspielen geblieben.

Mit das größte Problem aber ist, dass Leistungszulagen die Verdrängung der intrinsischen Motivation, die bereits durch die Leistungskontrolle ausgelöst wird, verstärken (s. dazu Osterloh u. Frey 2008), den „Brodgelehrten“ zu weiterer Verbreitung verhelfen. Man kennt den Effekt der Verdrängung der intrinsischen Motivation durch extrinsische Belohnungen aus Experimenten mit Kindern. Lässt man Kinder, die das gerne machen, Bilder malen und belohnt diese nicht nur mit Lob, sondern auch mit vorher angekündigten Preisen, dann nimmt ihr Interesse zu malen ab, wenn keine Preise mehr winken (Lepper et al. 1973). Eine Verdrängung intrinsischer durch extrinsische Motivation findet jedoch auch in beruflichen Kontexten statt, in denen, wie im Wissenschaftssystem, die intrinsische Motivation eine große Rolle spielt (Frey u. Jegen 2001; Reeson u. Tisdell 2008). Leistungsorientierte Zulagen verdrängen die intrinsische Motivation vor allem durch die Reduzierung der wahrgenommenen Autonomie. Wissenschaftler können zum Beispiel nicht mehr so frei über Forschungsthemen, Art der Veröffentlichungen (Bücher oder Zeitschriftenartikel, generelle oder spezialisierte Zeitschriften) und die anzuwendenden Methoden (quantitativ vs. qualitativ) entscheiden.

Jedoch ist die Höhe der Entlohnung auch für intrinsisch Motivierte alles andere als nebensächlich. Eine als fair empfundene fixe Entlohnung wird von intrinsisch Motivierten als Signal für Wohlwollen und Wertschätzung wahrgenommen und im Prinzip mit höherer Leistung beantwortet als eine vergleichbare variable Entlohnung (Irlenbusch u. Sliwka 2003). Es ist nicht so, dass Professoren nicht gerne auch viel Geld verdienen. Es geht ihnen da so ähnlich wie Thomas Bernhard – einem zweifelsohne intrinsisch motivierten Schriftsteller. Er schrieb seinem Verleger Unseld: „die Arbeit ist meine einzige Freude, meine größtmögliche Unzucht“. Dennoch hatte er stets das Gefühl, für seine Werke nicht ausreichend honoriert zu werden. Er klagte seinem Verleger: „[I]ch brauche etwas zum Leben also, wenn ich nichts habe, muss ich, wie jeder andere Mensch auch, arbeiten gehen. Dagegen habe ich nichts. Im Gegenteil, Holzhacken oder ähnliches ist mir die längste Zeit lieber als schreiben, aber dann kann ich auch nicht daran denken, den Roman, an dem ich arbeite, weiter zu bringen und so fort. Wie stellen Sie sich vor, lebt ein Mensch mit einem Bauch? Man muss ihn füllen, ganz einfach.“ (Bernhard u. Unseld 2009, S. 76).

Nun mögen Sie fragen: Weshalb funktioniert die leistungsabhängige Vergütung in Unternehmen, aber nicht in der Universität? Sie funktioniert auch in Unternehmen bei komplexeren Tätigkeiten eher schlecht als recht (Bebchuk u. Fried 2004; Frey u. Osterloh 2005). Viele Manager befürworten leistungsabhängige Vergütung, weil sie ihnen ermöglicht, ihr Einkommen kontinuierlich zu steigern – in sehr vielen Fällen auch bei weniger guten Leistungen. Vor elf Jahren verdiente der Vorstandsvorsitzende eines deutschen Dax-Unternehmens 20 mal so viel wie der Durchschnitt seiner Mitarbeiter, 2008 war es das 49-fache (Palan u. Werres 2009). In Unternehmen sind erfolgsabhängige Gehälter im

Gegensatz zu deutschen Universitäten nicht gedeckelt. Zudem herrscht zwischen Vorständen und Aufsichtsrat in der Regel ein gutes Einvernehmen, so dass bei der Vereinbarung der Erfolgskriterien nicht unbedingt das Interesse der Anteilseigner dominiert.

In diesem Zusammenhang mag ein Hinweis auf die USA von Belang sein: An US-Forschungsuniversitäten ist eine leistungsabhängige Vergütung nicht verbreitet. Es gilt dort der Grundsatz, Wissenschaftler außerordentlich sorgfältig auszuwählen und ihnen ein marktdäquates Gehalt zu bieten. Wenn man einen hochkarätigen Wissenschaftler gewonnen hat, gewährt man ihm die Autonomie, seine Forschungsprojekte und die Art, wie er sie durchführt und veröffentlicht, selbst zu bestimmen. Eine leistungsabhängige Vergütung ist genauso wenig üblich wie eine formale regelmäßige Kontrolle auf der Basis vorgegebener Kriterien. Dass Spitzenforscher ein hohes Gehalt kassieren, ohne fortzuführen, ehrgeizige Projekte zu verfolgen, ist so gut wie noch nie vorgekommen. Diesen Grundsatz hat James Bryan Conant, ein Präsident der Harvard Universität im Jahre 1945 formuliert: „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).

Er ist immer noch Teil der „Principles Governing Research at Harvard“: „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“ (beide Zitate habe ich Osterloh u. Frey 2008 entnommen).

7 Weshalb akzeptieren Wissenschaftler Anreize, identifizieren sich sogar nach und nach mit ihnen?

Zunächst finden, wie bereits ausgeführt, Politiker und Universitätsadministratoren Rankings gut, weil diese ihnen die Illusion einer rationalen Steuerung der Wissenschaft vermitteln. Auch manche Kollegen, insbesondere Kollegen der Betriebswirtschaftslehre finden Leistungsanreize gut, weil sie Anhänger der Neuen Institutionenökonomie sind, die diese Leistungsanreize propagiert (Ferraro et al. 2005). Es ist aber nicht auszuschließen, dass auch Kollegen, die das System zunächst ablehnen, sich im Laufe der Zeit mit ihm identifizieren und Nachwuchswissenschaftler erst gar nicht auf den Gedanken kommen, eine ablehnende Haltung zu entwickeln. In dem Maße, in dem nämlich Wissenschaftler ihren Entscheidungen Rankings und Leistungsanreize zugrunde legen und womöglich von ihnen profitieren, werden sie mit einer gewissen Wahrscheinlichkeit diese nach und nach auch inhaltlich gutheißen, und sich die ihnen zugrunde liegende Logik zu eigen machen. Ein Bewerber, der in Berufungsverhandlungen Leistungszulagen akzeptiert – und er muss sie akzeptieren, weil er sonst ein niedrigeres Gehalt bezieht –, will diese dann auch erreichen. Er bekommt Schwierigkeiten mit seinem Selbstbild, sollte er versuchen, dieses, sein eigenes Streben, als nur durch äußeren Zwang zustande gekommen zu charakterisieren. So wird über die Zeit aus einer bürokratischen Kontrolle eine „konzertierte“, eine konsensuelle, eine vom Teamgeist getragene (Barker 1993).

8 Rankings führen zu einer stärkeren Hierarchisierung von Fakultäten und Universitäten und zu einer Absenkung des nationalen Ausbildungsniveaus

Es mehren sich Anzeichen, dass eine durch Ökonomisierung abgesicherte Bildungspolitik zur Ausbildung weniger Spitzenuniversitäten führt und die Ausbildung der Masse der Studierenden auf Hochschulen und Fachbereiche nachgeordneter Hierarchiestufen verteilt, welche kaum noch über nennenswerte Ressourcen zur Forschung verfügen (Hartmann 2010). Diese Einschätzung vertrat der scheidende DFG-Präsident Ernst-Ludwig Winnacker in seiner Abschiedsrede auf der Jahresversammlung der DFG am 31. Mai 2006. Seiner Meinung nach würden die zwischen den Universitäten bereits bestehenden Qualitätsunterschiede durch die Exzellenzinitiative – auch eine Art Ranking – zunehmen. Neben „reinen Forschungsuniversitäten, die sich auch in der Ausbildung an den Anforderungen moderner wissenschaftlicher Forschung ausrichten, wird es solche geben, die dies nur ansatzweise und in einzelnen Fächern versuchen, solche, die diesen Anspruch erst gar nicht anstreben, und solche, die ihre Stärke eher in der Praxisorientierung suchen“ (Winnacker 2006, S. IX ff.). Angesichts bestehender Finanzierungsengpässe dürften die auf den unteren Ebenen des Systems befindlichen Fachbereiche und Hochschulen Schwierigkeiten haben, sich emporzuarbeiten. Und damit schwindet auch zunehmend die Motivation ihrer Mitglieder, Verbesserungen in Angriff zu nehmen. Es ist anzunehmen, dass im Zuge der Exzellenzinitiative das Matthäus-Prinzip, demzufolge dem gegeben wird, der schon hat, die deutschen Universitäten so stark prägen wird wie niemals zuvor (Hartmann 2010). Es ist jedoch mehr als fraglich, ob ein stark hierarchisch gegliedertes Universitätssystem eine bessere Ausbildung gewährleistet als ein stärker egalitäres (s. auch Hartmann 2007). Erfahrungen mit höchst ungleich ausgestatteten Hochschulen in den USA geben Anlass zur Skepsis. An der Spitzengruppe der Forschungsuniversitäten entspricht die Qualität der Lehre, insbesondere in den undergraduate Kursen, meist nicht der Forschungsreputation (Kirp 2004; Latzer 2004; Newman et al. 2004). Und das Niveau der Lehre an Hochschulen der nachgeordneten Ebenen ist vergleichsweise schlecht. Wie Hartmann (2007, S. 90) aufzeigt, „studiert [zwar] ... ein fast doppelt so hoher Teil eines Altersjahrgangs wie in Deutschland ..., ca. 90% der Studierenden aus den sozial schwächeren Schichten besuchen jedoch Hochschulen, die bestenfalls das Niveau deutscher Berufsakademien aufweisen. Fast die Hälfte geht sogar nur auf Zwei-Jahres-Colleges, die in etwa den letzten beiden Klassen einer deutschen gymnasialen Oberstufe entsprechen. An jenen etwa 150 Hochschulen, die mit den hiesigen Universitäten vergleichbar sind, stammen dagegen gerade einmal neun Prozent der Studierenden aus der unteren Hälfte der Bevölkerung, während 74% aus dem oberen Viertel der Gesellschaft kommen.“

9 Was tun?

Aus den angestellten Überlegungen ziehe ich die folgenden Schlussfolgerungen:

1. Das System der leistungsabhängigen Zulagen für Ergebnisse in Forschung und Lehre sollte abgeschafft werden. Wissenschaftliche Reputation und staatliche Alimentierung sollten wieder stärker auseinandergehalten werden.

2. Fakultäten oder Vereinigungen wie der Verband der Hochschullehrer für Betriebswirtschaft sollten die Frage, ob sie sich an Rankings beteiligen, ihren Mitgliedern zur Entscheidung vorlegen. Hinzuweisen ist in diesem Zusammenhang auf die jüngste Weigerung des Historikerverbandes, sich an einer sogenannten „Informed Peer Review“ des Wissenschaftsrats, bei der methodisch besonders anspruchsvoll vorgegangen werden sollte, zu beteiligen. Die Ablehnung wurde begründet mit der „Unmöglichkeit, ein dynamisches Fach wie die Geschichtswissenschaft parametrisch gleichsam in einer Momentaufnahme abzubilden und wertend zu erfassen“ Was dabei herauskomme, seien „Informationen für politische Diskussions- und Entscheidungsprozesse, die gemessen an der Realität des Faches unterkomplex seien, der Politik aber das Gefühl des Informiertseins durch die Wissenschaft selbst vermittelten. Auf diese Weise bediene der Wissenschaftsrat letztlich die politische Illusion, Wissenschaft lasse sich parametrisch durch das Setzen bestimmter Anreize steuern ...“ (Plumpe 2010).
3. Entscheidungen, die Karrieren von Wissenschaftlern betreffen, dürfen sich nicht auf Rankings oder gewichtete Zeitschriftenpublikationen stützen. Indikatoren zur Gewichtung von Publikationen und Rankings widersprechen eklatant wissenschaftlichen Grundsätzen valider und reliabler Messung.
4. Gutachter zur Beurteilung wissenschaftlicher Werke oder Projekte sollten nicht nur nach Interessengebieten ausgewählt werden, sondern zunächst danach, ob sie das gleiche Paradigma vertreten wie der Autor. Wie gezeigt, haben Gutachter häufig Vorurteile gegen Paradigmen, die nicht die ihren sind.
5. Autoren sollten berechtigt sein, anonyme Gutachten abzulehnen, wenn sie aufzeigen können, dass diese von Vorurteilen beeinflusst sind.
6. Lokale innovative Ansätze sollten gefördert werden, auch wenn die Rezeption in der internationalen Scientific Community auf sich warten lässt (s. das obige Zitat von James March).

Ein an der Stanford University forschender und lehrender Kollege erläuterte mir einmal (zu einer Zeit, in der Impact Faktoren noch keine Rolle spielten), der Präsident seiner Universität könne sich in aller Regel darauf verlassen, dass sich die Qualität von Forschung und Lehre in den Schools und Departments auf höchstem Niveau fortentwickle. Dafür Sorge der Ehrgeiz der Forscher. Der Präsident würde aber auch Hinweise aus der Scientific Community aufgreifen, wenn es in dem einen oder anderen Fach Probleme gäbe. Dann würde er den Rat auswärtiger Experten einholen und gegebenenfalls Vorschläge für die Besetzung von Berufungskommissionen machen. So ähnlich stelle ich mir Qualitätspolitik an deutschen Universitäten vor: Man verlässt sich grundsätzlich auf den sich aus intrinsischer Motivation speisenden Wettbewerb zwischen den Institutionen. Dieser Wettbewerb ist nicht auf Impact Faktoren und andere Indikatoren angewiesen, die unweigerlich falsche Signale setzen und unerwünschtes Verhalten auslösen. Die Beteiligten sind sich einig, dass die Beurteilung von Forschungsqualität nur auf der Basis einer gründlichen inhaltlichen Auseinandersetzung mit Forschungsergebnissen erfolgen kann – einer Auseinandersetzung, die nur Wissenschaftler kompetent wahrnehmen können. Bei begründeten Hinweisen auf einen Forschungsertrag, der ein gesetztes Niveau unterschreitet, können übergeordnete Instanzen – der Präsident, das Wissenschaftsminis-

terium – Evaluationen initiieren. Solange gewährleistet ist, dass die Forschung an dem zu evaluierenden Institut auf anerkannten Paradigmen aufbaut bzw. neue Paradigmen von anerkannten Forschern erarbeitet werden, sind Forscher als Evaluatoren zu bestellen, die auf der Basis dieser Paradigmen selbst schon geforscht haben. Regelmäßige, gesetzlich vorgeschriebene Evaluationen der Forschung sind entbehrlich. Sie absorbieren viel Forschungskapazität bei Evaluierten und Evaluatoren und verleiten zur Konstruktion von Fassaden. Im Übrigen wird bei Akkreditierungen, die auf die Einhaltung eines Mindestniveaus der Lehre achten, auch ein Auge auf Forschung als Basis einer guten Lehre geworfen. Jedoch dürfen in Evaluationen Rankingdaten, die, wie gezeigt wurde, wissenschaftlich nicht zu rechtfertigen sind, keine Rolle spielen. Dass diese in den Köpfen der Mitglieder der Evaluationskomitees herumspuken, wird indes nicht zu vermeiden sein.

Um diese Abschiedsvorlesung mit einer persönlichen Bemerkung abzuschließen: Ich bin sehr froh, für akademische Leistungen nicht mit Leistungspunkten oder gar Boni entlohnt worden zu sein. Die Forschung hat mich meistens in einen Zustand angenehmer, mitunter auch unangenehmer Anspannung versetzt. Boni für Leistungen, die man einem Gremium vorrechnen muss, hätten mir die Freude an der Wissenschaft ganz bestimmt gründlich ausgetrieben.

Literatur

- Adler, Nancy J., und Anne Will Harzing. 2009. When knowledge wins: Transcending the sense and nonsense of academic rankings. *Academy of Management Learning & Education* 8 (1): 72–95.
- Adler, R., J. Ewing, und P. Taylor. 2008. Citation statistics. A report from the joint committee on quantitative assessment of research (IMU, ICIAM, IMS) <http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf>. 09. Juli 2010.
- Armin, M., und M. Mabe. 2000. Impact factors. Use and abuse. *Perspectives in Publishing* 1:1–6.
- Barker, James R. 1993. Tightening the iron cage: Concertive control in self-managing teams. *Administrative Science Quarterly* 38:408–437.
- Bebchuk, Lucien, und Jesse Fried. 2004. *Pay without performance: The unfulfilled promise of executive compensation*. Cambridge: Harvard University Press.
- Bedeian, Arthur G. 2004. Peer review and the social construction of knowledge in the management discipline. *Academy of Management Learning & Education* 3:198–216.
- Bernhard, Thomas, und Siegfried Unseld. 2009. *Der Briefwechsel*. Frankfurt a. M.: Suhrkamp.
- Biagioli, Mario. 2002. From book censorship to academic peer review. *Emergences: Journal for the Study of Media & Composite Cultures* 12 (1): 11–45.
- Bourdieu, Pierre. 1998. *Vom Gebrauch der Wissenschaft. Für eine klinische Soziologie des wissenschaftlichen Feldes*. Konstanz: Universitätsverlag.
- Campanario, Juan Miguel. 1996. Have referees rejected some of the most-cited articles of all times? *Journal of the American Society for Information Science* 47:302–310.
- Connelly, Catherine E., und Daniel G. Gallagher. 2007. Making „the list“: Business school rankings and the commodification of business research. *Journal of Curriculum Theorizing* 23 (Summer): 103–116.
- Edward, Hutchins. 1985. *Surely you're joking, Mr. Feynman!: Adventures of a curious character*. New York: Norton.
- Ewing, John. 2006. Measuring journals. *Notices of the American Mathematical Society* 53:1049–1053.

- Ferraro, Fabrizio, Jeffrey Pfeffer, und Robert I. Sutton. 2005. Economic language and assumptions: How theories can become self-fulfilling. *Academy of Management Review* 30:8–24.
- Fischer, Klaus. 2003. Soziale und kognitive Aspekte des Peer Review Verfahrens. In *Evaluation wissenschaftlicher Institutionen*, Hrsg. Klaus Fischer und Heinrich Parthey, 23–62. Wissenschaftsforschung Jahrbuch Berlin: Gesellschaft für Wissenschaftsforschung.
- Frank, Robert H., und Philip J. Cook. 1995. *The winner-take-all society*. New York: The Free Press.
- Frey, Bruno S. 2003. Publishing as prostitution – Choosing between one's own ideas and academic success. *Public Choice* 116:205–223.
- Frey, Bruno S. 2004. Plädoyer für eine positive Ökonomik. In *Perspektiven der Wirtschaftspolitik*, Hrsg. Christof A. Schaltegger und Stefan C. Schaltegger, 759–766. Zürich: vdf Hochschulverlag.
- Frey, Bruno S., und R. Jegen. 2001. Motivation crowding theory: A survey of empirical evidence. *Journal of Economic Surveys* 15:589–611.
- Frey, Bruno S., und Margit Osterloh. 2005. Yes, managers should be paid like bureaucrats. *Journal of Management Inquiry* 14:96–111.
- Gans, Joshua S., und George B. Shepherd. 1994. How are the mighty fallen: Rejected classic articles by leading economists. *Journal of Economic Perspectives* 8 (1): 165–180.
- Garfield, Eugene. 2006. The history and meaning of the journal impact factor. *Journal of the American Medical Association* 295 (1): 90–93.
- Gottfredson, S. D. 1978. Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgements. *American Psychologist* 33:920–934.
- Grey, Christopher. 2010. Organizing studies: Publications, politics and polemic. *Organization Studies* 31:677–694.
- Hartmann, Michael. 2007. Elite und Masse – die Aufspaltung der deutschen Universitätslandschaft in Forschungs- und Ausbildungsuniversitäten. In *Bildungspolitik und Bildungsforschung: Herausforderungen und Perspektiven für Gesellschaft und Gewerkschaften in Deutschland*, Hrsg. H. Sünker und I. Miethe, 87–99. Frankfurt a. M.: Peter Lang.
- Hartmann, Michael. 2010. Die Exzellenzinitiative – ein Paradigmenwechsel in der deutschen Hochschulpolitik. In *Die bedingte Universität. Die Institution der Wissenschaft zwischen „Sachzwang“ und „Bildungsauftrag“*, Hrsg. C. Adam, J. Müller, R. Thun, und W. Warnecke. 20–42. Stuttgart: Schmetterling Verlag.
- Hennig-Thurau, Thorsten, Gianfranco Walsh, und Ulf Schrader. 2004. VHB-JOURNAL: Ein Ranking von betriebswirtschaftlich-relevanten Zeitschriften auf der Grundlage von Expertenurteilen. *Zeitschrift für Betriebswirtschaftliche Forschung* 56:520–543.
- Hogler, Ramond, und Michael A. Gross. 2009. Journal rankings and academic research: Two discourses of faculty work. *Management Communication Quarterly* 23:107–126.
- Horrobin, D. F. 2001. Something rotten at the core of science? *Trends in Pharmacological Sciences* 22 (2): 51–52.
- Horstkotte, Hermann. 2010. Professoren wollen höheres Grundgehalt http://www.fr-online.de/in_und_ausland/wissen_und_bildung/aktuell/?em_cnt=1389358&em_loc=1739. Zugegriffen: 9. Juli 2010.
- Irlenbusch, B., und D. Sliwka. 2003. Steigern variable Löhne die Leistung? Arbeitspapier, Bonn.
- Judge, Timothy A. et al. 2007. What causes a management article to be cited – article, author, or journal? *Academy of Management Journal* 50:491–506.
- Kirp, David. 2004. *Shakespeare, Einstein, and the bottom line. The marketing of higher education*. Cambridge: Harvard University Press.
- Latzer, Barry. 2004. *The hollow core. Failure of the general education curriculum*. Washington: American Council of Trustees and Alumni.
- Lepper, Mark R., David Greene, und Richard E. Nisbett. 1973. Undermining children's intrinsic interest with extrinsic reward. *Journal of Personality and Social Psychology* 28:129–137.

- Luhmann, Niklas. 1998. *Die Wissenschaft der Gesellschaft*. Frankfurt a. M.: Suhrkamp.
- Maasen, Sabine, und Peter Weingart. 2007. Unternehmerische Universität und neue Wissenskulturskultur. In *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen*, Hrsg. Hildegard Matthies und Dagmar Simon. Wiesbaden: VS Verlag. (Leviathan Sonderheft 24;141–160).
- Macdonald, Stuart, und Jacqueline Kam. 2007. Ring a ring o' roses: Quality journals and gamesmanship in management studies. *Journal of Management Studies* 44:640–655.
- March, James G. 2004. Parochialism in the evolution of a research community: The case of organization studies. *Management and Organization Review* 1:5–22.
- Merton, Robert K. 1985. Der Matthäus-Effekt in der Wissenschaft. in *Entwicklung und Wandel von Forschungsinteressen*, Hrsg. Robert K. Merton, 147–171. Frankfurt a. M.: Suhrkamp.
- Miller, C. Chet. 2006. Peer Review in the organizational and management sciences: Prevalence and effects of reviewer hostility, bias, and dissensus. *Academy of Management Journal* 49:425–431.
- Müller-Böling, Detlef. 2000. *Die entfesselte Hochschule*. Gütersloh: Bertelsmann Stiftung.
- Neuhaus, Christoph, Werner Marx, und Hans D. Daniel. 2009. The publication and citation impact profiles of *Angewandte Chemie* and the *Journal of the American Chemical Society* based on the sections of Chemical Abstracts: A case study on the limitations of the Journal Impact Factor. *Journal of the American Society for Information Science & Technology* 60 (1): 176–183.
- Newman, Frank, Lara Couturier, und Jamie Scurry. 2004. The future of higher education. Rhetoric, reality, and the risks of the market. San Francisco: Jossey-Bass.
- o. Verf. 1999. Streit um „faule Professoren“. Dokumentation, *Forschung & Lehre* 6:452–453.
- o. Verf. 2009. Handelsblatt Betriebswirte-Ranking 2009: Top 200 – Lebenswerk <http://www.handelsblatt.com/bwl-lebenswerk/>. Zugegriffen: 9. Juli 2010.
- o. Verf. 2010a. Principles governing research at Harvard <http://www.fas.harvard.edu/~research/greybook/principles.html>. Zugegriffen: 9. Juli 2010.
- o. Verf. 2010b. Der Junge, der alles richtig macht <http://www.handelsblatt.com/politik/nachrichten/der-junge-der-alles-richtig-macht;2367977;0>. Zugegriffen: 9. Juli 2010.
- Osterloh, Margit, und Bruno S. Frey. 2008. Anreize im Wissenschaftssystem. Arbeitspapier, Zürich.
- Palan, Dietmar, und Thomas Werres. 2009. Der große Bluff. *Manager Magazin* 12:55–61.
- Peters, D., und S. Ceci. 1982. Peer review practices of psychological journals: the fate of published articles submitted again. *Behavioral and Brain Sciences* 165:187–195.
- Plumpe, Werner. 2010. Qualitätsmessung: Stellungnahme zum Rating des Wissenschaftsrates aus Sicht des Historikerverbandes. <http://hsozkult.geschichte.hu-berlin.de/index.asp?type=diskussionen&id=1101&view=pdf&pn=forum>. Zugegriffen: 9. Juli 2010.
- Podsakoff, Philip M. et al. 2008. Scholarly influence in the field of management: A bibliometric analysis of the determinants of university and author impact in the management literature in the past quarter century. *Journal of Management* 34 (4): 641–720.
- Reeson, A. F., und J. G. Tisdell. 2008. Institutions, motivations and public goods: An experimental test of motivational crowding. *Journal of Economic Behavior & Organization* 68:273–281.
- Sauder, Michael und Wendy N. Espeland. 2009. The discipline of rankings: Tight coupling and organizational change. *American Sociological Review* 74:63–82.
- Schiller, Friedrich. 1789. Was heißt und zu welchem Ende studiert man Universalgeschichte? In *Der Deutsche Merkur*, Hrsg. C. M. Wieland, 105–135. Weimar: Hofmann.
- Singh, Gangaram, Kamal M. Haddad, und Chee W. Chow. 2007. Are articles in „top“ management journals necessarily of higher quality. *Journal of Management Inquiry* 16:319–331.
- Smith, R. 1997. Journal accused of manipulating impact factor. *British Medical Journal* 314:461–463.
- Starbuck, William H. 2005. How much better are the most prestigious journals? The statistics of academic publication. *Organization Science* 16:180–200.

- Süß, Stefan. 2007. Grenzen einer leistungsorientierten Anreizgestaltung für Hochschullehrer. *Hochschulmanagement* 2 (3): 68–72.
- Svensson, Göran. 2010. SSCI and its impact factors: A „prisoner’s dilemma“? *European Journal of Marketing* 44:23.
- Voswinkel, Stephan. 2005. Die Organisation der Vermarktlichung von Organisationen – Das Beispiel erfolgsbezogenen Entgelts. In *Organisationsgesellschaft. Facetten und Perspektiven*, Hrsg. Wieland Jäger und Uwe Schimank, 287–312. Wiesbaden: VS Verlag.
- Weber, Max. 1994. Wissenschaft als Beruf, 1917/1919 – Politik als Beruf 1919. In *Werke*, Hrsg. von Wolfgang J. Mommsen und Wolfgang Schluchter in Zusammenarbeit mit Birgitt Morgenbrod. Tübingen: Mohr-Siebeck.
- Weingart, Peter. 2005. Das Ritual der Evaluierung und die Verführung der Zahlen. In *Die Wissenschaft der Öffentlichkeit. Essays zum Verhältnis von Wissenschaft, Medien und Öffentlichkeit*, Hrsg. Peter Weingart. 102–122. Weilerswist: Velbrück.
- Winnacker, Ernst-Ludwig. 2006. Im Wettbewerb um neues Wissen: Exzellenz zählt, *Forschung – Das Magazin Der Deutschen Forschungsgemeinschaft* 2006 (2): 5–11.