

Riskante Abwägungen zwischen Neugier und Nutzen

Evaluationen gehören zu den großen wissenschaftspolitischen Herausforderungen unserer Zeit, weil sie unterschiedliche Wertmaßstäbe verbinden müssen, um die nachhaltige Entwicklung des Wissenschaftssystems fördern zu können.

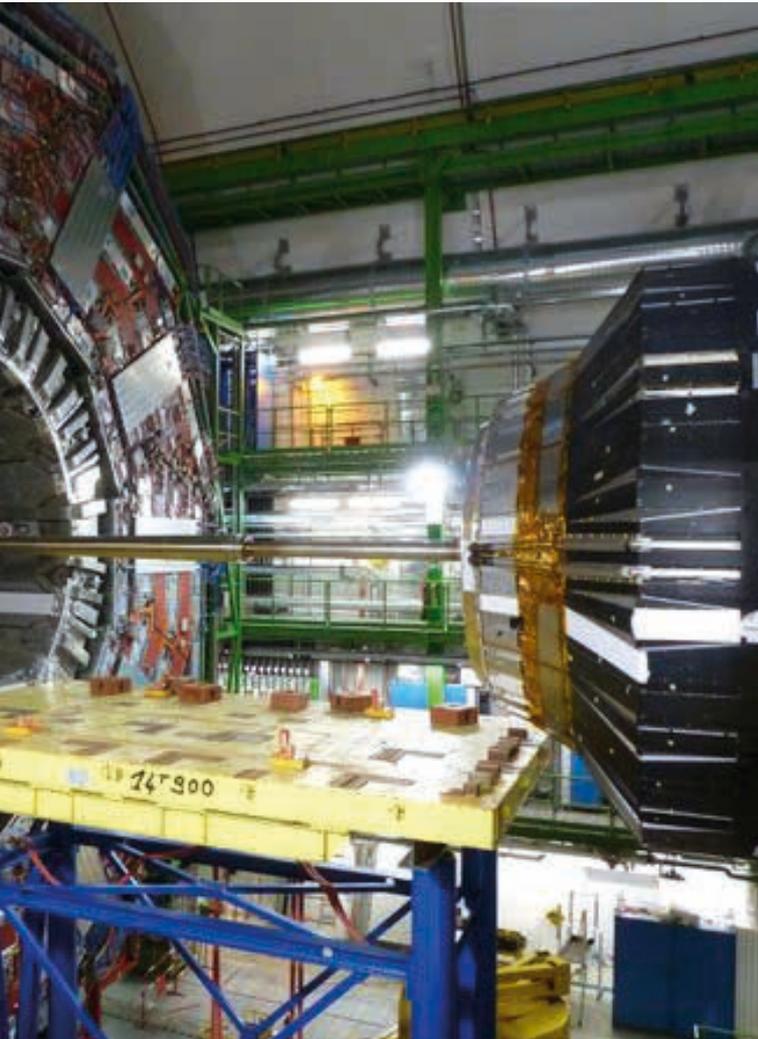
VON JÖRG HACKER UND STEFAN ARTMANN

EINE DER WICHTIGSTEN Voraussetzungen des wissenschaftlichen Fortschritts besteht darin, dass es die Forscher selbst sind, die über den Erkenntniswert von Beobachtungen, Hypothesen und Theorien entscheiden. Indem sie sich zu Experten für ein Forschungsgebiet ausbilden, erwerben sie die Kompetenz, im Allgemeinen am besten zu beurteilen, ob eine neue Information zum wissenschaftlich anerkannten Wissen gezählt werden sollte. Folgerichtig gehört es zum Selbstverständnis der sich selbst organisierenden Wissenschaft, dass die Forscher gleichfalls darüber entscheiden sollten, welche Personen, Projekte und Institutionen erfolgreich gearbeitet haben und entsprechend gute Chancen besitzen, zukünftig finanziert zu werden. So förderte beispielsweise die Deutsche Forschungsgemeinschaft (DFG) im Jahr 2012 laut ihrem Jahresbericht rund 31.000 Projekte mit einer Summe von beinahe 2,7 Mrd. Euro. Dementsprechend legt sie als Selbstverwaltungsorganisation der deutschen Wissenschaft einen überaus großen Wert auf die transparente Evaluierung ihrer Förderprogramme, die sich an klaren Bewertungsstandards ausrichtet.

Wenn es um die Verteilung einer Ressource wie Geld geht, treten auch im Wissenschaftssystem Konflikte auf, die analog in anderen gesellschaftlichen Bereichen auf der Tagesordnung stehen. Wie wichtig ist es, dass ein Forschungsprojekt die Neugier der Experten befriedigt, im Verhältnis zu dem Nutzen, den dasselbe Projekt für die Lösung drängender gesellschaftlicher Herausforderungen gebracht hat? Wie sollten die Grundlagen- und die angewandte Forschung hinsichtlich ihrer längerfristigen Konsequenzen für das Gemeinwohl bewertet werden? Welche Wertmaßstäbe sollten überhaupt bei Evaluationen eine Rolle spielen und wie sollten sie zueinander in Beziehung gesetzt werden?

Dass solche Fragen schon auf Grund des Finanzvolumens, um das es geht, gesamtgesellschaftliche Relevanz besitzen, zeigt sich an einem wichtigen Indikator: Laut einer Erhebung des Stifterverbands für die Deutsche Wissenschaft vom Dezember 2013 hat der Anteil der Ausgaben für Forschung und Entwicklung am Bruttoinlandsprodukt (BIP) in Deutschland mit 2,98 Prozent im Jahr 2012 fast das Drei-Prozent-Ziel der Lissabon-Strategie des Europäischen Rats erreicht. Dabei machten die Aufwendungen der öffentlichen Hand 0,96 Prozent des BIP aus – das sind über 25,5 Mrd. Euro. Diese Zahlen verdeutlichen die Notwendigkeit, forschungs- und innovationspolitische Maßnahmen evidenzbasiert zu evaluieren, wie jüngst auch die Expertenkommission Forschung und Innovation (EFI) in ihrem diesjährigen Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands betonte.





wird, zu einer differenzierten Bilanzierung des Erkenntnisfortschritts führen: Wenn es darum geht zu bewerten, welche Forschungsprojekte erfolgreich gewesen sind, könnte es zum Beispiel angeraten sein, diejenigen Ergebnisse nicht überzubewerten, die direkt zum gegenwärtigen Wissensstand beigetragen, aber womöglich nur geringe neue Forschungsaktivitäten angeregt haben. Ungewöhnliche Beobachtungen, gewagte Hypothesen und kühne Theorien, die zwischenzeitlich bereits falsifiziert worden sind, könnten eine nachhaltige, aber indirekte positive Rolle gespielt haben – etwa wenn ihre Falsifikation zu bedeutsamen Resultaten geführt hat, die ansonsten vermutlich erst später bekannt geworden wären.

Lässt sich wissenschaftliche Relevanz messen? Impact und Reputation

Der Anspruch, den relativen Wert bewährter und falsifizierter Theorien für den Erkenntnisfortschritt differenziert abzuschätzen, führt zu dem Problem, ob die innerwissenschaftliche Relevanz von Forschungsergebnissen eine messbare Größe ist. Diese Frage hat zur Entwicklung von heute weitverbreiteten Indikatoren wie dem Journal

Bewährt oder falsifiziert – oft dauert es Jahrzehnte, bis theoretische Annahmen bestätigt werden können, etwa die Existenz des Higgs-Teilchens. Im Bild: der Teilchenbeschleuniger LHC des Kernforschungszentrums CERN bei Genf, wo Physiker 2012 vermutlich das neue Elementarteilchen aufspürten.

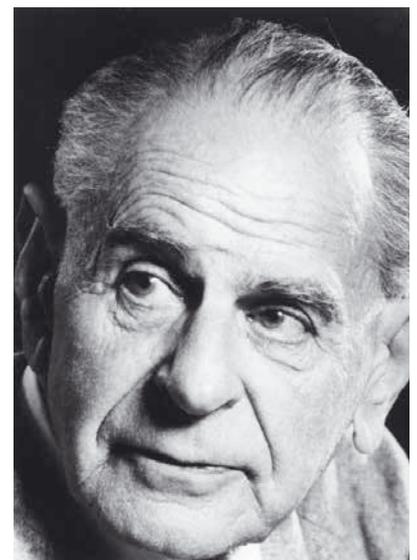
Kühne Hypothesen können die Wissenschaft oftmals am weitesten voranbringen: Karl R. Popper (1902–1994).

Wahrheit als grundlegender wissenschaftlicher Wertmaßstab

Auf welchen Wertmaßstab bezieht sich wissenschaftliches Handeln? Die Antwort scheint auf der Hand zu liegen: Wissenschaftler orientieren sich, wenn sie eine neue Theorie beurteilen, an der Unterscheidung zwischen wahr und falsch – oder, wie es Karl R. Popper in seinem Werk „Logik der Forschung“ formulierte, an der Unterscheidung zwischen bewährt und falsifiziert. Sagt eine Theorie zum Beispiel experimentelle Beobachtungen korrekt voraus, dann hat sie sich bewährt; widersprechen ihr diese Beobachtungen aber, dann gilt sie als falsifiziert. Dabei richtet sich an die Begründungen solcher Urteile wiederum die Forderung, mittels desselben Wertmaßstabs einschätzbar zu sein. So muss die Falsifizierung einer Theorie durch die Angabe einer Beobachtung, die ihr widerspricht, selbst falsifizierbar sein.

Diese Eigenart des wissenschaftlichen Wissens sollte, wenn sie in Evaluationen ernstgenommen

Impact Factor (JIF) angeregt. Er beansprucht, die Relevanz einer Zeitschrift zu messen; sein Wert ist, vereinfacht ausgedrückt, gleich der durchschnittlichen Anzahl von Zitationen, die ein in der Zeitschrift veröffentlichter Beitrag während der jeweils vorhergehenden zwei Jahre in anderen Zeitschriften aufweist. Die Erfolgsgeschichte des JIF hat mit einem Bedürfnis nach Evaluation begonnen: Amerikanische Hochschulbibliothekare wollten die Bedeutung von Zeitschriften für den wissenschaftlichen Diskurs quantifizieren, um ihre Entscheidungen über Subskriptionen objektiv begründen zu können.



Bei der Forschungsevaluation werden verschiedene quantitative Indikatoren eingesetzt, beispielsweise der Hirsch-Faktor, der den Einfluss eines Wissenschaftlers bibliometrisch erfassen soll, oder die Summe der Drittmittelinwerbungen eines Instituts in einem bestimmten Zeitraum. Unabhängig von der konkreten Ausgestaltung der Rankings und Ratings, die auf Indikatoren-Clustern beruhen, tragen sie alle dazu bei, wissenschaftliche Reputation als berechenbare Größe erscheinen zu lassen. Reputation bezeichnet, so der Soziologe Niklas Luhmann in seinem Werk „Die Wissenschaft der Gesellschaft“, „die Leistung der Erstkommunikation von Wissen“ und macht es einfacher, die Relevanz von Forschungsergebnissen einzuschätzen: Je höher die Reputation eines Wissenschaftlers ist, desto stärker haben seine bisherigen Publikationen im Urteil der Fachkollegen die Wissenschaft vorangebracht, so dass seine aktuellen Resultate eine entsprechende Aufmerksamkeit verdienen.

Gegenwärtig findet weltweit eine Debatte über Sinn und Unsinn der Objektivierung von Reputation durch quantitative Evaluationsindikatoren statt. Im Bereich der Zitationsindices hat diese Debatte Anfang 2013 zur San Francisco Declaration On Research Assessment (DORA) geführt, die von der American Society for Cell Biology initiiert worden ist (am.ascb.org/dora). Sie schlägt Maßnahmen gegen die Überbewertung des JIF vor und wurde bis Mitte März 2014 von fast 11.000 Wissenschaftlerinnen und Wissenschaftlern sowie Organisationen unterzeichnet. Die wichtigsten Argumente wider den überzogenen Einsatz solcher Reputationsmaße warnen davor, dass die Wirkungen, die sie auf das Verhalten der evaluierten Wissenschaftler ausüben, den Nutzen des Reputationsbegriffs in der wissenschaftlichen Praxis konterkarieren können. Die Ökonomin Margit Osterloh weist darauf hin, dass die Zerstörung von Reputation als sinnvoller Wertmaßstab für Forschungsleistungen und damit die praktische Falsifikation von quantitativen Reputationsindikatoren der Logik des „Paradoxes der Leistungsmessung“ folgt: Wenn Wissenschaftler sich mit einer Forschungsfrage vor allem oder ausschließlich deshalb beschäftigen, weil sie hoffen, dass sie ihre Ergebnisse in den führenden Zeitschriften veröffentlichen können, dann überlagert die extrinsische Motivation des Reputationsgewinns die intrinsische Motivation der Wahrheitssuche, deren Erfolg eigentlich gemessen werden sollte.

Die Objektivierung von Reputation zu Bewertungszwecken droht das evaluierte Handeln zu pervertieren, falls die Evaluation „[...] den Fokus auf den Leistungsindikator legt und nicht auf

das, was er messen soll. [...] Die einzige Methode, um diesem Paradox zu entrinnen und dennoch Leistungsindikatoren beizubehalten, wären deren ständige Veränderungen und Anpassungen durch die betroffenen Fachleute“ (M. Osterloh, Das Paradox der Leistungsmessung und die Nachhaltigkeit der Forschung, in: J. Hacker (Hrsg.), Nachhaltigkeit in der Wissenschaft. Leopoldina-Workshop am 12. November 2012 in Berlin, Halle (Saale) und Stuttgart, 2013, S. 104).

Im Sinne einer nachhaltigen Evaluationskultur sollten quantitative Reputationsmaße daher immer wieder innerhalb der verschiedenen Fachdisziplinen mittels einer differenzierten Beurteilung der Relevanz relativiert werden, welche die Forschungsergebnisse der betreffenden Wissenschaftler für den Erkenntnisfortschritt besitzen. In diese Richtung zielt zum Beispiel das so genannte „Informed Peer Review“ im Forschungsrating des Wissenschaftsrats. Solche anspruchsvolleren Verfahren sind für die evaluierenden Gremien gewiss in dem Sinne riskant, dass sich die Bewertung der Produktivität eines falsifizierbaren Forschungsergebnisses jederzeit ändern kann – aber dieses Risiko bewusst auf sich zu nehmen, macht die Verantwortung der evaluierenden gegenüber den evaluierten Wissenschaftlern aus.

Nachhaltigkeit als Wertmaßstab für Forschung

Die Frage nach der Relevanz von Forschungsergebnissen wird vermehrt mit Blick auf die Beziehungen zwischen dem Wissenschaftssystem und anderen gesellschaftlichen Bereichen gestellt. Dann tritt neben die Bedeutung wissenschaftlicher Resultate für den Erkenntnisfortschritt ihr außerwissenschaftlicher Nutzen. Im umfassendsten Sinne geschieht dies, wenn gefragt wird, welchen Beitrag die Wissenschaft für die nachhaltige Entwicklung der Menschheit leistet. Wer bestreite ernsthaft, dass die Wissenschaften ganz wesentlich dabei helfen können, globale Herausforderungen wie Klimawandel und Seuchenausbreitung zu meistern? Dieser Konsens allein reicht aber nicht aus, wenn sich Politik, Zivilgesellschaft und Wissenschaft darüber verständigen wollen, wie wissenschaftliche Erkenntnisprozesse, öffentliche Meinungsbildung und demokratische Entscheidungsfindung im Sinne der Nachhaltigkeit am besten miteinander zu verknüpfen sind. Denn die Beteiligten folgen durchaus unterschiedlichen Wertmaßstäben – selbst wenn es darum geht, ein gemeinsames Ziel zu verwirklichen.



Die allermeisten Wissenschaftler zeigen sich skeptisch, wenn Interessenvertreter aus Politik und Zivilgesellschaft im Namen des Gemeinwohls fordern, direkte Relevanz für nachhaltige Entwicklung zu einem neuen allgemeinen Kriterium der Forschungsevaluation zu erheben. Würde damit nicht eine Entwicklung einsetzen, an deren Ende die Unterordnung der Wissenschaft unter ihre unmittelbare gesellschaftliche Nützlichkeit stünde? Das erste Opfer wäre gewiss die Grundlagenforschung, bei der niemand vorherzusagen weiß, für welche gesellschaftlichen Probleme ihre Ergebnisse einmal relevant werden könnten. Nachhaltige Entwicklung, die Generationen überspannt, und kurzsichtige Wissenschaftsevaluation, die auf den Nachweis eines unmittelbaren Nutzens fixiert ist, passen nicht zusammen. Daher ist es eine der größten wissenschaftspolitischen Herausforderungen unserer Zeit, das Verhältnis zwischen den Wertmaßstäben der Wahrheit, der Reputation und der Nachhaltigkeit immer wieder neu auszutarieren.

Nützliche Neugier und Neugier auf den Nutzen

An Themen, zu denen momentan solche komplexen Abwägungen stattfinden, mangelt es wahrlich nicht. Dabei lassen sich zwei wichtige Strategien erkennen. Erstens ist es notwendig, die Politik und die Zivilgesellschaft, aber auch die Wissenschaft kontinuierlich darauf hinzuweisen, dass die von Neugier getriebene Forschung nachhaltigen Nutzen stiften kann. Dies ist etwa bei der Entwicklung dringend notwendiger neuer Antibiotika der Fall: Wie lässt sie sich durch

eine Koordination gesundheitspolitisch gesetzter Anreize, ergebnisoffener Grundlagenforschung (zum Beispiel in der Synthetischen Biologie) und ökonomischer Verwertung fördern? Ein weiteres Beispiel ist die langfristige Bedeutung von Disziplinen wie der Taxonomie, deren kontinuierliche Forschungsarbeit zwar nur selten Schlagzeilen produziert, die aber für Fortschritte in den Lebenswissenschaften (z. B. Biodiversitätsforschung) und der Medizin (z. B. Personalisierte Medizin) unabdingbar ist. Hierzu hat eine Arbeitsgruppe der Leopoldina im Juni dieses Jahres eine Stellungnahme vorgelegt.

Zweitens sollten Wissenschaftler neugieriger auf die Herausforderung gemacht werden, nutzenstiftende Anwendungen für ihre Forschungsergebnisse zu entwickeln und schädlichen Anwendungen entgegenzuarbeiten. Hier ist beispielsweise an das Stichwort „Dual use“ zu denken: Wie können Wissenschaftler Forschungsvorhaben etwa zu hochpathogenen Grippeviren, die potentiell sowohl nützliche als auch schädliche Anwendungen nahelegen, so durchführen, dass sie die Gefahr des Missbrauchs minimieren? Eine Arbeitsgruppe der DFG (federführend) und der Leopoldina hat jüngst Leitlinien zum Umgang mit Forschungsfreiheit und Risiken in der Wissenschaft erarbeitet. Auch für gesamtgesellschaftliche Vorhaben wie die Energiewende ist es von großer Relevanz, dass sich Wissenschaftler damit beschäftigen, realistische Handlungsalternativen zu beschreiben und systematisch zu erforschen – nicht zuletzt, um vor scheinbar vielversprechenden Irrwegen zu warnen. Letzteres ist etwa bei der Nutzung der Bioenergie der Fall, wie die Leopoldina 2013 in der Stellungnahme „Bioenergie – Möglichkeiten und Grenzen“ erläuterte.

Die Strategien der nützlichen Neugier und der Neugier auf den Nutzen sollen dazu beitragen, dass gerade angesichts zunehmender Forderungen nach mehr Nachhaltigkeit in der Wissenschaft keine unüberbrückbaren Spannungen zwischen den Erwartungen entstehen, die aus dem Wissenschaftssystem und den anderen Bereichen der Gesellschaft auf die Forschung gerichtet werden. Differenzierte und risikobewusste Evaluationen können bei dem notwendigen Interessensausgleich sehr hilfreich sein, indem sie verdeutlichen, dass Nutzen und Neugier einander nicht widersprechen: Sie entdecken den potentiellen Nutzen in der von Neugier getriebenen Forschung und wecken die Neugier auf nutzbringende Anwendungen wissenschaftlicher Expertise.

Forschungsevaluation mit Gemeinwohlorientierung: die Stellungnahme „Antibiotika-Forschung: Probleme und Perspektiven“ (2013) der Akademie der Wissenschaften in Hamburg und der Nationalen Akademie der Wissenschaften Leopoldina.

DIE AUTOREN

Prof. Dr. Jörg Hacker ist Mikrobiologe und seit 2010 Präsident der Deutschen Akademie der Naturforscher – Nationale Akademie der Wissenschaften Leopoldina. Er wurde 2006 als ordentliches Mitglied in die Bayerische Akademie der Wissenschaften gewählt und ist seit 2008 korrespondierendes Mitglied.

PD Dr. Stefan Artmann ist Philosoph und seit 2014 Leiter des Präsidialbüros der Nationalen Akademie der Wissenschaften Leopoldina.