

# IST GESELLSCHAFTLICHE RELEVANZ VON FORSCHUNG BEWERTBAR, UND WENN JA, WIE?

JULIAN HAMANN, DAVID KALDEWEY, JULIA SCHUBERT

## 1. EINLEITUNG

Ja, die gesellschaftliche Relevanz von Forschung ist zweifellos bewertbar. Exkursionen in die aktuelle Wissenschaftsforschung und Wissenschaftspolitik bringen vielfältige Indikatoren, Theorien und Methoden ans Tageslicht, die zudem durch eine Vielzahl von Akteuren aus Wissenschaft und Praxis mit breiter Expertise und zunehmender Professionalität laufend weiterentwickelt werden. Wir schlagen vor, die vielfältigen heute zur Anwendung kommenden Bewertungsverfahren auf drei idealtypische Varianten des „Wie“ zurückzuführen. Diese drei Bewertungsmodi stehen dann für grundsätzlich verschiedene Möglichkeiten der Bewertung gesellschaftlicher Relevanz. Schwieriger wird eine Antwort auf die Leitfrage dieses Essays, wenn wir neben der prinzipiellen Möglichkeit auch die Konsequenzen der Bewertung mitbedenken. Je nachdem, wie wir diese Konsequenzen einschätzen, stellt sich die normative Frage nach der Erwünschtheit oder gar Notwendigkeit von Relevanzurteilen: *Soll* Forschung überhaupt hinsichtlich ihrer gesellschaftlichen Relevanz bewertet werden oder wird damit die Autonomie der Wissenschaft beeinträchtigt?

Am Ende unseres Essays möchten wir deshalb die Voraussetzungen eines Bewertungssystems diskutieren, welches dieser Sollens-Frage standhält. Wir werden zeigen, dass eine über das triviale „Ja“ hinausgehende Antwort auf die Preisfrage in einer Pluralisierung der Bewertungsverfahren selbst liegt. Während die Diversität wissenschaftlicher Forschung von keinem einzelnen Bewertungsverfahren angemessen erfasst werden kann, würde eine Vielfalt nebeneinander eingesetzter Verfahren die gesellschaftliche Relevanz der Wissenschaft genau dadurch erhöhen, dass sich Forschung dann auf unterschiedliche Formen gesellschaftlicher Relevanz berufen könnte. So würde die Autonomie wissenschaftlicher Forschung gestärkt – und zwar nicht als bloß negative „Freiheit von“, sondern als positive „Freiheit zu“.

## 2. WIE KANN DIE GESELLSCHAFTLICHE RELEVANZ VON FORSCHUNG BEWERTET WERDEN?

Das Bedürfnis, die gesellschaftliche Relevanz wissenschaftlicher Forschung – oder allgemeiner: die Relevanz von nicht unmittelbar auf die Praxis bezogener erkenntnisorientierter Tätigkeit – festzustellen, ist so alt wie eben diese Tätigkeiten selbst. Schon die griechische Antike bietet hier anschauliche Beispiele. Einer alten Anekdote zufolge war Thales, weil er den Blick auf die Sterne gerichtet hatte, in einen Brunnen gefallen und wurde daraufhin von einer thrakischen Magd ausgelacht – dieses Lachen steht in der Philosophiegeschichte idealtypisch für ein von der Gesellschaft gefälltes negatives Relevanzurteil.<sup>1</sup> Andererseits weiß schon Aristoteles, wie man umgekehrt die Relevanz theoretischer Weltbetrachtung kommuniziert: Thales habe mit Hilfe der Astronomie eine ergiebige

Olivenernte vorausgesehen – und daraufhin für wenig Geld alle verfügbaren Olivenpressen aufgekauft, um diese dann für einen viel höheren Preis wieder zu verpachten. Die Bewertung der Nützlichkeit wissenschaftlicher Erkenntnis über den Indikator des Geldes, das man mit ihr verdienen kann, ist also nicht erst im 20. Jahrhundert üblich geworden, auch wenn heute, beispielsweise über die Anzahl der Patente und sich daraus ergebende quantifizierbare Einnahmen, die Operationalisierung dieses Kriteriums auf soliderer Basis steht. In jüngerer Zeit lassen sich nun zunehmend Versuche beobachten, die gesellschaftliche Relevanz von Forschung umfassender zu bestimmen. Wir schlagen vor, die diversen Methoden, Theorien und Konzepte idealtypisch in drei Modi einzuteilen. Urteile über die gesellschaftliche Relevanz von Forschung lassen sich demnach (a) historisch-narrativ, (b) standardisiert-administrativ oder (c) demokratisch-partizipativ fundieren.

(a) Im *historisch-narrativen* Bewertungsmodus wird grundsätzlich *ex post* über die Relevanz von Forschung entschieden. Das Argument dahinter ist, dass die gesellschaftliche Anschlussfähigkeit von Forschungsprogrammen erst dann bewertbar ist, wenn sich gezeigt hat, dass und warum sie sich in der gesellschaftlichen Praxis als einflussreich erwiesen haben. Das Kriterium für Relevanz ist hier die historische Realität, die immer erst im Rückblick zugänglich ist. Konsequenterweise weist dieser Bewertungsmodus schon die Möglichkeit einer *ex ante*-Bewertung von sich. Eine klassische Anekdote aus der Wissenschaftsgeschichte erzählt in diesem Sinne, wie Benjamin Franklin, der 1783 bei den ersten bemannten Flügen von Heißluftballons in Paris anwesend war, die Frage nach dem Nutzen solcher Erfindungen mit der rhetorischen Gegenfrage beantwortete: „What is the good of a newborn baby?“<sup>2</sup> Gemäß der Logik des historisch-narrativen Bewertungsmodus entscheidet also immer erst die Zukunft über die Relevanz der Forschung der Gegenwart: Heute können wir sagen, dass damals die Geschichte der Luftfahrt begann. Natürlich müssen dazu nicht immer Jahrzehnte oder gar Jahrhunderte vergehen; die Relevanz der Atombombe beispielsweise war den Zeitgenoss/inn/en schon einsichtig, bevor die erste Anwendung erfolgte. Doch auch in diesem Fall kann festgehalten werden, dass einige der physikalischen Grundlagen – etwa Albert Einsteins Erkenntnis der in der Masse enthaltenen und potenziell freisetzbaren Energie, Ernest Rutherfords Entwicklung des Kern-Hülle-Modells des Atoms oder Henri Becquerels Entdeckung der radioaktiven Strahlung des Urans – drei bis vier Jahrzehnte vor dem Manhattan Projekt gelegt wurden.

Am Beispiel der Luftfahrt und der Atombombe sehen wir, dass retrospektive Argumentationen die gesellschaftliche Relevanz von Forschung im Einzelfall sehr überzeugend begründen können. Die Grenzen des historisch-narrativen Bewertungsmodus werden jedoch sichtbar, sobald versucht wird, über anekdotische Evidenz hinaus die gesellschaftliche Relevanz von Grundlagenforschung systematisch zu erheben und methodisch kontrollierbare Indikatoren zu entwickeln. Sichtbar wurden diese Schwierigkeiten beispielsweise in einer vielbeachteten Debatte in der U.S.-amerikanischen Wissenschaftspolitik der 1960er Jahre. Zwischen 1963 und 1967 hatte das U.S. Department of Defense eine Studie mit dem Titel „Project Hindsight“ durchgeführt und dabei nachgewiesen, dass die Entwicklung wichtiger high-tech Waffensysteme nur zu einem sehr geringen Anteil auf Erkenntnissen der Grundlagenforschung aufbaut.<sup>3</sup> Darauf antwortete die National Science Foundation (NSF) mit der eigenen, 1969 publizierten Studie TRACES („Technology in Retrospect and Critical Events in Science“), die aufzeigte, welche Bedeutung die Grundlagenforschung für eine Reihe von wichtigen technologischen Innovationen im zivilen Bereich hatte.<sup>4</sup> Dabei kam die Studie zu Schlussfolgerungen, die denen des Project Hindsight konträr gegenüberstanden: 70% der für eine erfolgreiche Innovation notwendigen wissenschaftlichen Durchbrüche seien auf anwendungsferne Grundlagenforschung zurückzuführen. Der Vergleich der Studien und ihrer entgegengesetzten wissenschaftspolitischen Empfehlungen verweist zum einen auf das bekannte Problem, dass Forschungsergebnisse mit vorgefassten Meinungen der Auftraggeber korrelieren. Zum anderen, und das ist der interessantere Punkt, zeigen sich die unvermeidbaren methodischen Schwierigkeiten retrospektiver Argumentation. So erklärt sich die Differenz der beiden Studien unter anderem dadurch, dass die Hindsight-Studie einen Zeitraum von etwa 20 Jahren, die TRACES-Studie dagegen einen Zeitraum von bis zu 50 Jahren ansetzte. Für den historisch-narrativen Bewertungsmodus können wir daher festhalten, dass seine Resultate stark davon abhängen, was genau unter Grundlagenforschung verstanden wird, für welche Güter bzw. Konzepte

gesellschaftliche Relevanz beansprucht wird, und welcher historische Zeitraum für die Bewertung herangezogen wird.

(b) Auch im *standardisiert-administrativen* Bewertungsmodus wird *ex post* über die Relevanz von Forschung entschieden. An die Stelle historischer Narrative treten nun allerdings standardisierte Bewertungsverfahren, die auf einheitlich verwendbaren Relevanzkriterien aufbauen. Impliziert ist in diesem Bewertungsmodus ein Bild von Wissenschaft als Gegenstand gesellschaftlicher Investitionen und Governance. Erwartet werden messbare Erträge, sei es in Form von ökonomischen Pay-Offs oder allgemeiner in Form von positiven Effekten in anderen gesellschaftlichen Bereichen.

Einschlägige Beispiele sind nationale Bewertungsregime, wie die 2014 eingeführte Impact-Messung im Rahmen des britischen Research Excellence Framework (REF) und das in Australien jüngst angelaufene Engagement and Impact Assessment (EI). Sie zielen auf die turnusmäßige Bewertung von Forschung nach ihrem gesellschaftlichen Impact. In beiden Fällen basiert die Bewertung zwar auf *peer review*, wird aber zentral verwaltet und in ein standardisiertes Notensystem überführt. Während nationale Assessments fallstudienbasiert sind und institutionelle Einheiten wie universitäre Fakultäten oder Departments evaluieren, finden sich im Bereich der „Altmetrics“ indikatorbasierte Bewertungen, die stärker an Kennwerten orientiert sind. Hier wird die gesellschaftliche Relevanz von Forschung anhand der generierten Downloadzahlen oder anhand ihres Fußabdrucks in den sozialen Medien bewertet. Im Unterschied zu hochschulpolitisch verantworteten nationalen Assessments werden Altmetrics in der Regel von privatwirtschaftlichen Dienstleistern und Verlagen angeboten und sind nicht an nationale Hochschulsysteme, sondern an einzelne Publikationen oder Publikationsorte gekoppelt. Gemeinsam ist Altmetrics und nationalen Assessments jedoch, dass die Relevanz wissenschaftlicher Forschung und der mit ihr zutage geförderten Erkenntnis vermessen und ausgezählt, verglichen und gerankt wird.

Durch ihre systematische Anwendung versetzen Verfahren des standardisiert-administrativen Modus eine Vielzahl von Forschungskulturen in eine Prüfungssituation, in der diese ihre gesellschaftliche Relevanz in einem einheitlichen Format eindeutig nachweisen müssen. Dass sich ein unmittelbarer gesellschaftlicher Impact dann für bestimmte Arten von Forschung plausibler darstellen lässt als für andere, liegt nahe. Es überrascht daher nicht, wenn etwa in der jüngsten Runde des britischen REF die Medizin und die Gesundheitswissenschaften ihren Impact recht mühelos unter Beweis stellen konnten. Im Bereich der Genetik verweisen Forscher/innen beispielsweise darauf, die genetischen Grundlagen von Brustkrebs, Taubheit oder Insulinresistenz zu untersuchen und bringen damit die gesellschaftliche Relevanz ihrer Forschung durch den Bezug auf den allgemein anerkannten Wert menschlicher Gesundheit *en passant* auf den Punkt. Andere Forschungsfelder können ihren Impact nicht in der gleichen Unmittelbarkeit behaupten. So argumentieren beispielsweise Altphilolog/inn/en im REF, dass die Auseinandersetzung mit den griechischen Komödien dem zeitgenössischen Theaterbetrieb zugutekommt, während Philosoph/inn/en die gesellschaftliche Bedeutung ihrer Forschung in der Beantwortung der Frage sehen, ob wir in einer Computersimulation leben.<sup>5</sup> Es ist wohl kein Zufall, dass sich bei der Einführung des Impact-Kriteriums in Großbritannien vor allem die Fachvertretungen zweier theoretisch fundierter Disziplinen besorgt zeigten: Die London Mathematical Society verwies in einem Statement darauf, dass die Mathematik benachteiligt sei bei der Bewertung ihres gesellschaftlichen Impacts, weil ihre Forschungsfragen der allgemeinen Öffentlichkeit nur schwer vermittelbar seien; die British Philosophical Association erinnerte daran, dass die gesellschaftliche Bedeutung von Philosophie langfristig angelegt, unvorhersehbar und deshalb schwer zu quantifizieren sei.<sup>6</sup> Solche Warnungen verweisen auf ein weiteres Charakteristikum des standardisiert-administrativen Bewertungsmodus: Es geht hier nicht nur darum, einen einheitlichen Relevanzbegriff flächendeckend und systematisch auf verschiedene Forschungskulturen anzuwenden, sondern auch um eine im Vergleich zum historisch-narrativen Bewertungsmodus deutlich reduzierte Reichweite des Rückblicks. Die gesellschaftliche Relevanz muss innerhalb eines konkret definierten und kürzeren Zeithorizonts nachgewiesen werden: Im britischen REF sind es fünf Jahre, im australischen EI sogar nur drei Jahre. Würde man hier stattdessen 3.000 Jahre ansetzen, könnte kaum jemand sinnvoll an der gesellschaftlichen Relevanz der Mathematik oder der Philosophie zweifeln.

(c) Im Gegensatz zu sowohl dem historisch-narrativen wie auch dem standardisiert-administrativen Bewertungsmodus, welche die gesellschaftliche Relevanz wissenschaftlicher Forschung jeweils *ex post* feststellen, integriert der *demokratisch-partizipative* Modus die Bewertung bereits *ex ante* in den Forschungsprozess. Dieser Modus beruht auf dem Bild einer offenen Wissenschaft, die Wissen vor allem für außerwissenschaftliche Akteure, insbesondere für die Zivilgesellschaft produziert. Die gezielte Einbeziehung von Repräsentant/inn/en verschiedener gesellschaftlicher Gruppierungen soll die Relevanz von zunehmend spezialisierter Forschung mittels Transparenz und sozialer Inklusion sicherstellen. Dieser Modus bemüht somit einen Relevanzbegriff, der sich in erster Linie auf die gesellschaftliche Einbettung der Wissenschaft bezieht und nicht, wie etwa beim standardisiert-administrativen Modus, über sachliche Indikatoren vermessen werden kann. Gesellschaftliche Relevanz von Forschung soll dabei nicht einfach im Nachhinein beurteilt, sondern über kollaborative und partizipatorische Verfahren unmittelbar im Forschungsprozess hergestellt werden. Eine nachträgliche Bewertung ist dann im Idealfall gar nicht mehr nötig, da externe Erwartungen direkt in den Forschungsprozess integriert sind. Gesellschaftlich relevante Forschung ist dann gesellschaftlich eingebettete Forschung.

Über Konzepte wie „Responsible Research and Innovation“ (RRI), „Citizen Science“ oder „Open Science“ hat der demokratisch-partizipative Bewertungsmodus gerade in jüngerer Vergangenheit starken wissenschaftspolitischen Widerhall gefunden. In diesem Zuge haben sich zahlreiche und inhaltlich nicht immer scharf voneinander abgrenzbare Bewertungsverfahren etabliert, die von „Consensus Conferences“, über „Hybrid Forums“, „Trading Zones“ bis hin zu „Collective Learning“ reichen. Sie haben das Ziel, die Perspektiven und Belange der demokratischen Öffentlichkeit in den Forschungsprozess zu integrieren und damit eine Wissenschaft „mit der und für die Gesellschaft“ zu stärken.<sup>7</sup> Eine solche Wissenschaft, so die Hoffnung, wäre dann *per definitionem* relevant.

Diese gesellschaftliche Öffnung kann im demokratisch-partizipativen Modus – je nach konkretem Bewertungsverfahren – auf verschiedenen Ebenen und zu unterschiedlichen Zeitpunkten des Forschungsprozesses ansetzen. Bewertungsverfahren können erstens bereits auf eine Demokratisierung in der Ausarbeitung von Forschungsagenden abzielen. Die Demokratisierung der Forschung soll hier über die Festlegung der Forschungsfragen erfolgen – eine Forderung, die auch in der Wissenschaftsphilosophie Anklang findet.<sup>8</sup> Ein historisch frühes Beispiel für die Öffnung des Forschungsprozesses auf Ebene des Agenda-Settings sind bürgerschaftlich getragene Vereine des 19. Jahrhunderts, in denen etwa Natur- oder Heimatkunde betrieben wurde. Derartig strukturierte Forschung konnte gesellschaftliche Relevanz beanspruchen, weil sie in ihren Fragestellungen auch den Problemdefinitionen von Nicht-Wissenschaftler/inne/n folgte. In jüngerer Zeit spricht man diesbezüglich von „Public Engagement“-Strategien. Das EU-geförderte NanOpinion-Projekt beispielsweise hat nach eigenen Angaben in einem Verbund von elf europäischen Ländern über 27.000 Menschen erreicht, um deren Einschätzungen und Bedürfnisse in Bezug auf Nanotechnologie zu bündeln und in zukünftiger Forschung auf diesem Gebiet zu berücksichtigen. Ähnliche Ziele verfolgt das ebenfalls EU-geförderte VOICES-Projekt, in dem die Perspektiven und Belange von insgesamt 1.000 europäischen Bürger/inne/n in die Strukturierung einer gesellschaftlich relevanten Forschungsagenda zum Thema „städtisches Abfallmanagement“ integriert wurden. Bewertungsverfahren des demokratisch-partizipativen Modus können zweitens auf eine Öffnung des Forschungsprozesses auf Ebene der Datenerhebung setzen. Gerade jüngere technologische Entwicklungen haben die Möglichkeiten der Partizipation in diesem Bereich potenziert, so dass etwa mit Hilfe des Internets große Datenmengen in Zusammenarbeit von Bürger/inne/n und Forscher/inne/n transkribiert, klassifiziert und verarbeitet werden können. Seit fast 20 Jahren können Bürger/innen im Projekt SETI@home beispielsweise zur wissenschaftlichen Suche nach außerirdischer Intelligenz beitragen. Auch verschiedene Bereiche anwendungsorientierter Forschung entdecken und nutzen zunehmend die besondere Expertise von Stakeholdern bei der Datenerhebung. In Portugal setzen beispielsweise mehrere Projekte der Calouste Gulbenkian Foundation, des Portuguese Institute of the Sea and Atmosphere oder der ArtesanalPesca-Initiative auf die Integration von Beobachtungen und Einsichten aus dem lokalen Fischereibetrieb für Studien der marinen Ökosysteme. Einige demokratisch-partizipative Bewertungsverfahren öffnen den Forschungsprozess drittens auf Ebene der Ergebniskommunikation. Im Vordergrund steht hier der Wissenstransfer zwischen akademischer Forschung und gesellschaftlicher Öffentlichkeit. Häufig geht es dabei um eine Informierung der Öffentlichkeit zu Themen, die von

unmittelbarer Relevanz für den Alltag sind. Das dänische Projekt PULSE beispielsweise zielt seit 2015 mit Hilfe von „Science Exhibitions“ darauf ab, sozialstrukturell benachteiligte Familien mit Kindern über gesunde Lebensstile aufzuklären. Das Projekt verbindet ein laufendes Forschungsprojekt mit der kollaborativen Vermittlung der Forschungserkenntnisse.

Zusammenfassend unterscheiden sich die drei skizzierten Modi darin, wie sie einerseits ihre Bewertung legitimieren, und welche Relevanzbegriffe sie dabei andererseits mobilisieren. Die historische Rückschau bemüht einen Relevanzbegriff, der sich erst im Zeitverlauf entwickeln muss und sich nur über die Distanz zwischen zwei Zeitpunkten entfalten kann. Standardisierte Bewertungen definieren Relevanz als den unmittelbaren Impact von Forschungsergebnissen, die auf konkrete gesellschaftliche Probleme und Interessen bezogen sind. Dieser Impact, so die Prämisse, kann in Fallstudien eindeutig identifiziert oder im Falle der Altmetrics durch elaborierte Technologien algorithmisch berechnet werden. Demokratischen Bewertungsverfahren schließlich liegt ein Relevanzbegriff zugrunde, der auf kommunikative Offenheit, Transparenz und soziale Inklusion abstellt. In kritischer Perspektive zeigt sich hier allerdings das Problem der Repräsentativität der integrierten „Öffentlichkeit“. Im Versuch, eine möglichst breite Öffentlichkeit in den Forschungsprozess zu integrieren, wird diese meist auf demographische Kriterien reduziert. Auch eventuelle geographische Verzerrungen in der vermeintlichen Abbildung der Weltöffentlichkeit scheinen insbesondere für die Adressierung globaler Herausforderungen, wie Klimawandel und Naturschutz, ein ernsthaftes Problem darzustellen. Das Problem eines *ex ante* festgelegten Relevanzbegriffes wird dann lediglich verschoben auf das Problem, wer die Gesellschaft prozessual vertritt. Im Vergleich erkennen wir also ein diachrones Verständnis von Relevanz im Fall des historischen Bewertungsmodus, einen ergebnisorientierten Begriff von Relevanz im Fall des standardisierten Bewertungsmodus und einen prozeduralen Relevanzbegriff im Fall des demokratischen Bewertungsmodus.

Eine weitere Achse der hier vorgeschlagenen Systematik verläuft entlang der unterschiedlich begründeten Legitimität der gefällten Relevanzurteile. So suggeriert der historisch-narrative Bewertungsmodus, dass bereits der historische Rückblick die gesellschaftliche Relevanz der Wissenschaft offenbart – die Retrospektive legitimiert ihr Urteil unabhängig vom Betrachter. Die Autorität ihres Urteils offenbart sich scheinbar objektiv und wie von selbst. Dass das in der Praxis nicht so eindeutig ist, zeigt schon die erwähnte Debatte um die Project Hindsight- und TRACES-Studien. Dagegen liegt die Legitimität des standardisiert-administrativen Bewertungsmodus in der systematischen Anwendung von Fallstudien oder Indikatoren. Der demokratisch-partizipative Bewertungsmodus legitimiert sich schließlich *ex negativo* gerade daraus, dass Relevanzurteile nicht von Expert/inn/en gefällt werden, sondern dass die Gesellschaft selbst zu Worte kommt – sei es in Form einer breiten demokratischen Öffentlichkeit oder in Form der lokalen und alternativen Expertisen von Laien. Die drei Modi zapfen für ihre Relevanzurteile also je unterschiedliche Quellen der Autorität an. Sie werden legitimiert über den historischen Rückblick, ihre Systematizität oder ihre Nähe zu außerwissenschaftlichen Belangen.

### **3. SOLL DIE GESELLSCHAFTLICHE RELEVANZ VON FORSCHUNG BEWERTET WERDEN?**

Die gesellschaftliche Relevanz der Wissenschaft ist also in verschiedener Hinsicht bewertbar. Sie lässt sich im anekdotischen Rückblick belegen (historisch-narrativer Modus), sie lässt sich vermessen und in Form standardisierter Assessments vergleichbar machen (standardisiert-administrativer Modus) und sie lässt sich deliberativ fundieren, indem die außerwissenschaftliche Öffentlichkeit in den Forschungsprozess integriert wird (demokratisch-partizipativer Modus). Es bleibt die Frage: wozu eigentlich?

Das gängigste Argument für eine Relevanzbewertung lautet wahrscheinlich, dass die Wissenschaft der sie tragenden Gesellschaft gegenüber in der Pflicht stehe und deshalb auf das Gemeinwohl hin orientiert sein solle. Gerade für öffentlich finanzierte Forschung liegt eine Rechenschaftspflicht gegenüber der Gesellschaft auf der

Hand. Viele Relevanzbewertungen gehen deshalb einher mit einer einfachen wissenschaftspolitischen Strategie: Wenn klar ist, welche Forschung gesellschaftlich relevante Ergebnisse zeitigt, bestehen gute Gründe, diese Forschung öffentlich zu finanzieren. Davon, so die Annahme, profitiere dann im Gegenzug auch die Wissenschaft selbst, da sie sich auf diese Weise laufend ihres gesellschaftlichen Rückhaltes versichere und sich nicht in selbstzweckhaften Glasperlenspielen verliere. Transparente Nachweise der gesellschaftlichen Relevanz von Forschung stärken demnach das Vertrauen der Öffentlichkeit in die Wissenschaft – und nebenbei auch das Vertrauen in die Wissenschaftspolitik: Sie legitimieren die Milliardensummen, die in allen Industrienationen in die Förderung von Forschung und Entwicklung investiert werden.

Rückenwind erhalten Befürworter/innen der Relevanzbewertung auch durch ein sich seit den 1990er Jahren abzeichnendes gesamtgesellschaftliches Klima, in dem Evaluation, Transparenz und *accountability* in immer mehr gesellschaftlichen Teilbereichen wirksam werden. Zugleich ist es eben diese allgemeine Evaluationseuphorie in der von Michael Power beschriebenen „Audit Society“,<sup>9</sup> die auch Kritiker/innen der Relevanzbewertung auf den Plan ruft. Diese führen zwei grundsätzliche Argumente gegen die systematische Evaluation und Kontrolle der gesellschaftlichen Relevanz von Forschung ins Feld. Das erste Argument zielt darauf, wissenschaftliches Erkenntnisstreben als Selbstzweck und damit letztlich als eine Kulturleistung hervorzuheben.<sup>10</sup> Die systematische Ausrichtung auf gesellschaftliche Nützlichkeit, so die damit zusammenhängende Befürchtung, würde diesen essentiellen Kern der Wissenschaft und damit auch den Wert wissenschaftlichen Wissens an sich korrumpieren. Das zweite Argument zielt auf forschungspraktische und methodologische Fragen und verweist auf grundsätzliche Grenzen der Bewertbarkeit: Unklar sei zunächst, wie gesellschaftliche Relevanz definiert werden könne und wer hier die Deutungshoheit habe. Zudem seien die zeitlichen Rhythmen der langfristigen wissenschaftlichen Erkenntnisproduktion nicht kompatibel mit kurzfristigen Ansprüchen auf verwertbare Erkenntnisse.<sup>11</sup> Beide Argumente vertrauen darauf, die Qualitätskontrolle weiterhin der wissenschaftlichen Gemeinschaft zu überlassen, die hierfür immerhin seit Jahrhunderten über die Praxis des *peer review* verfügt. Sie warnen davor, dass externe Relevanzbewertung die nachhaltige Erfolgsgeschichte der Wissenschaft und damit auch ihren Beitrag zu gesellschaftlichem Fortschritt gefährde.

Unseres Erachtens gilt es diese grundsätzlichen Einwände immer mitzubedenken. Gleichzeitig meinen wir jedoch, dass sie der Komplexität, Größe und Ausdifferenzierung des heutigen Wissenschaftssystems nicht gerecht werden. Zum einen zeigen historische wie gegenwärtige Selbstbeschreibungen der Wissenschaft, dass und wie sich Wissenschaftler/innen schon immer bewusst im Spannungsfeld zwischen der autonomen Erkenntnisuche und der gesellschaftlichen Relevanz positioniert haben.<sup>12</sup> Zum anderen mag es in „postfaktischen“ Zeiten, in denen sich immer mehr Menschen von der Wissenschaft ab- und alternativen Fakten zuwenden, nicht ratsam sein, sich Fragen nach der gesellschaftlichen Relevanz von Forschung völlig zu entziehen. Natürlich muss es immer auch geschützte Räume für ein zurückgezogenes und allein an Wahrheitsfragen orientiertes Forschen geben. Aber so notwendig diese Art von Forschung ist, so wenig kann von ihr auf eine vollständige Isolierung der Wissenschaft von ihrer gesellschaftlichen Bedeutung geschlossen werden. Nun steht jedoch neben diesen beiden grundsätzlichen Argumenten gegen die Bewertung der gesellschaftlichen Relevanz von Forschung noch ein drittes Argument, welches die Komplexität des Wissenschaftssystems nicht ignoriert, sondern gerade hier seinen Ausgangspunkt hat. Auf dieses dritte Argument wollen wir uns im Folgenden konzentrieren. Es bezieht sich darauf, dass jedes Bewertungsverfahren, vor allem wenn es unmittelbar mit der Zuteilung von Ressourcen verknüpft ist, auch nicht-intendierte Effekte hat.

#### **4. NICHT-INTENDIERTE EFFEKTE VON RELEVANZBEWERTUNGEN**

Wenn Bewertungsverfahren in der Breite institutionalisiert werden, wenn also das Wissenschaftssystem gewissermaßen systematisch unter Aufsicht gestellt wird, lassen sich auf der Ebene der institutionellen Strukturen ebenso wie auf der Ebene der Forschungspraxis intendierte und nicht-intendierte Effekte beobachten. Intendierte Effekte können als positive Effekte beschrieben werden, die zu den gewünschten Resultaten führen; ihre

Antizipation motiviert ja zuallererst die Einführung von entsprechenden Bewertungsverfahren. Die intendierten Effekte von Relevanzbewertungen sind oben schon angeschnitten worden und zielen im weitesten Sinn auf die Gemeinwohlorientierung der Wissenschaft. Konkrete Ziele, die mit der Institutionalisierung von Bewertungsverfahren verfolgt werden, sind etwa eine Ausrichtung von Forschungsprogrammen an Problemen, die nicht nur wissenschaftsintern, sondern auch gesellschaftlich – etwa wirtschaftlich oder politisch – als wichtig erachtet werden (Agenda Setting), die Stärkung regionaler, nationaler und/oder supranationaler Forschungs- und Innovationssysteme durch engere Kopplungen von Wissenschaft, Industrie, Staat und Zivilgesellschaft (institutionelle Profilbildung), eine effektive und leistungsgerechte Verteilung von Ressourcen (Wettbewerb), die Förderung von Spitzenforschung als nationales Prestigeprojekt (Exzellenz), oder auch die Übersetzung von Forschungsergebnissen in eine für die gesamte Bevölkerung verständliche Sprache (Wissenschaftskommunikation).

Während intendierte Effekte geplant, aber deswegen keineswegs vor Kritik gefeit sind, erscheinen nicht-intendierte Effekte mehr oder weniger überraschend. Es handelt sich in der Regel (aber nicht notwendig) um negative bzw. unerwünschte oder zumindest nicht erwartete Nebenwirkungen. Bei den nicht-intendierten Effekten der Kopplung von Bewertungsverfahren und Forschungsförderung unterscheiden wir zwei Varianten: Zum einen kann es vorkommen, dass die mit den Bewertungsverfahren verknüpften Hoffnungen auf eine relevantere Wissenschaft nur scheinbar erfüllt werden und die Gesamtleistung des „kontrollierten“ Wissenschaftssystems dieselbe bleibt – oder sich sogar verschlechtert. Die Soziolog/inn/en Michael Sauder und Wendy Espeland haben hier auf Mechanismen der „Reaktivität“ verwiesen, d.h. auf das in vielen gesellschaftlichen Bereichen beobachtbare Phänomen, dass jede Messung von Leistungen bei der Erbringerin der Leistung dazu führt, die jeweils geltenden Kriterien zu antizipieren und sich ihnen strategisch – oft heißt das: oberflächlich – anzupassen.<sup>13</sup> Das kann im Falle der Wissenschaft bedeuten, dass Universitäten, Forschungsinstitute oder auch einzelne Forscher/innen ihre jeweiligen Forschungsprofile primär auf bewertungsrelevante Kriterien und Maßzahlen hin ausrichten. Im Extremfall kann es dabei zu einem „gaming the system“ kommen, zu einer versteckten strategischen Manipulation der eigenen Kennziffern mit dem einzigen Zweck, von außen positiv evaluiert zu werden.

Interessanter und in der Praxis bedeutsamer als solche gewissermaßen korrumpierten Reaktionen auf Bewertungsverfahren erscheint uns jedoch die zweite Variante nicht-intendierter Effekte. Hier geht es um diejenigen Wirkungen, die auch und gerade dann zu beobachten sind, wenn Wissenschaftler/innen die Bewertungssysteme nicht absichtlich bzw. aus strategischen Gründen zu manipulieren suchen und sich nur oberflächlich auf gesellschaftliche Erwartungen einlassen, sondern wenn sie sich tatsächlich mit besten Absichten an etablierten Relevanzkriterien orientieren. In diesem Fall wäre ja zu hoffen, dass das Ziel einer gesellschaftlich relevanten Wissenschaft näher rückt und nachhaltig institutionalisiert werden kann. Doch auch hier lassen sich nicht-intendierte Effekte ausmachen – Effekte also, die auch dann eintreten, wenn die Wissenschaft im besten Sinne des Wortes reaktiv ist und etwa aus eigenem Antrieb heraus gesellschaftlich relevante Probleme adressiert und die von der Öffentlichkeit zur Verfügung gestellten Ressourcen möglichst optimal und effektiv im Sinne des Gemeinwohls nutzt. Man könnte hier entsprechend von den nicht-intendierten Effekten der intendierten Effekte sprechen. Wir konzentrieren uns im Folgenden auf einen solchen spezifischen Effekt, nämlich das Problem des Diversitätsverlustes, welches sich in den drei Bewertungsmodi jeweils verschieden darstellt:

(a) Da sie immer nur auf vergangene Forschung bezogen sind, wirken historisch-narrativ fundierte Relevanzurteile auf den ersten Blick nicht unmittelbar auf die Forschung zurück. Im Prinzip können historisch-narrative Bewertungen dennoch Folgen für die Wahrnehmung gegenwärtiger Forschung haben, etwa wenn diese als Fortsetzung vergangener Erfolge betrachtet wird und daraus ihre Legitimität herleitet, oder wenn in gegenwärtigen Projekten weit in der Zukunft liegende Bewertungen antizipiert werden – man kann sich beispielsweise eine Wissenschaftlerin vorstellen, die eine in der Gegenwart unpopuläre Forschungsfrage mit hoher Konsequenz verfolgt, weil sie überzeugt davon ist, dass die Zukunft ihr in der Relevanzeinschätzung recht geben wird. Üblicherweise ist das retrospektive Relevanzurteil aber abgekoppelt von der gegenwärtigen

Forschungspraxis. Wie wirkt sich nun vor diesem Hintergrund der historisch-narrative Bewertungsmodus auf die Diversität des Wissenschaftssystems aus? Problematisch scheint in diesem Zusammenhang, dass historisch-narrative Relevanzbewertungen meist auf besonders einschlägige, bahnbrechende und einer allgemeinen Öffentlichkeit einsichtige Fälle zurückgreifen – etwa die Luftfahrt oder die Atombombe – und so, wenn überhaupt, nur die Spitze des Eisbergs erfolgreicher und relevanter wissenschaftlicher Erkenntnisse in den Blick gerät. Die normalwissenschaftliche Forschung, die Fortschritte im Detail, die langsamen und auf den ersten Blick unscheinbaren Veränderungen, die die Wissenschaft in der Gesellschaft wirksam werden lassen, geraten hier kaum in den Blick.<sup>14</sup> Natürlich kann es wünschenswert sein, wenn sich ein Wissenschaftler im Alltag von den großen Erfolgsgeschichten und revolutionären Paradigmenwechseln motivieren lässt, aber es besteht zugleich die Gefahr, die Realität alltäglicher Forschung, die zwingende Notwendigkeit, sich auf sehr kleine, von außen belanglos erscheinende kleine Puzzleteile zu konzentrieren, aus den Augen zu verlieren.<sup>15</sup> Im Kontext des historisch-narrativen Bewertungsmodus sind es am Ende sehr wenige große Namen, die erinnert werden, während die zehntausenden Wissenschaftler/innen, die ebenfalls am Gesamtfortschritt beteiligt sind, oft unsichtbar bleiben. Wenn nun die Orientierung an historischen Musterbeispielen Anlass dazu geben sollte, zu große Sprünge machen zu wollen und zugleich die Arbeitsteilung und fortgeschrittene Spezialisierung in immer weiter ausdifferenzierte Forschungsgebiete mit einem negativen Stigma zu versehen, dann würde dem Gesamtsystem die notwendige Wissensbasis entzogen: Ein Eisberg, der nur noch aus einer Spitze besteht, würde immer weiter im Wasser versinken. Es liegt auf der Hand, dass die Produktivität einer zunehmend interdisziplinären und transdisziplinären Wissenschaft auf lange Sicht auf den Erhalt der Diversität unzähliger kleiner Forschungsgebiete angewiesen ist.

(b) Auch standardisiert-administrative Bewertungsverfahren konstatieren gesellschaftliche Relevanz *ex post* und greifen damit nicht unmittelbar in den Forschungsprozess ein. Sofern sie aber turnusmäßig oder dauerhaft stattfinden, können die verwendeten Relevanzkriterien auf die Forschung zurückwirken. Sobald Forschung auf Grundlage einer systematischen Evaluation finanziell und symbolisch prämiert wird, lohnt es sich für Universitäten und Forschungsinstitute, aber auch für einzelne Forscher/innen, sich auf zukünftige Assessments einzustellen. Je standardisierter und transparenter die Bewertungssysteme, desto schneller lernen alle beteiligten Akteure, welche Art von Forschung, welche Ergebnisse oder welche Präsentationsform ihrer Ergebnisse in den Bewertungsverfahren gut abschneiden. Sie handeln nur rational, wenn sie ihre Forschung an den Relevanzkriterien ausrichten, die zukünftig an sie angelegt werden – was zurückführt zum Problem der Reaktivität, das wir oben bereits angeschnitten haben. Wissenschaftler/innen sind dann unter Umständen gut beraten, das Forschungsfeld zu wechseln oder systematisch auf aktuell nachgefragte Forschungsthemen zu setzen. Solche im Einzelfall gut begründeten Strategieentscheidungen können sich auf institutioneller Ebene jedoch zu einer systematischen Ungleichbehandlung von Forschungsfeldern aggregieren – und damit die Diversität der Forschung insgesamt reduzieren. Sollte das britische REF beispielsweise die Medizin und die Gesundheitswissenschaften in der Tat systematisch bevorzugen, so könnte dies zu einer Unterrepräsentation insbesondere von sozial- und geisteswissenschaftlicher Forschung führen; Universitäten könnten durch die Antizipation von Relevanzbewertungen zur Streichung gesellschaftlich „nicht relevanter“ Fächer verführt werden – die dann im Extremfall ganz von der Landkarte verschwinden. Nicht jedes Fach steht, wie etwa die Islamwissenschaft, durch politische Entwicklungen plötzlich wieder im Fokus öffentlicher Aufmerksamkeit.

(c) Demokratisch-partizipative Bewertungsverfahren bewerten nicht retrospektiv, sondern zielen darauf ab, die gesellschaftliche Relevanz von Forschung prozessual, d.h. bereits während der Durchführung herzustellen. Gesellschaftlich relevante Forschung antwortet auf die Interessen der „Öffentlichkeit“. Vor dem Hintergrund der Diversitätsfrage zeigt sich jedoch schnell die Gefahr, dass solche gut gemeinten Verfahren zu Lasten der spezialisierten Tiefenschärfe von Forschung gehen. Sofern die wissenschaftliche Wissensproduktion auf der Ebene der Forschungspraxis für das bürgerliche Engagement von Laien geöffnet wird, entsteht unweigerlich ein Spannungsverhältnis mit spezialisierten Expert/inn/endiskursen, die einerseits den traditionellen Kern wissenschaftlichen Engagements ausmachen, andererseits aber Laien *per definitionem* ausschließen. Die repräsentative Integration einer demokratischen Öffentlichkeit in den Forschungsprozess erschwert nicht nur die

Verwendung der unabdingbar anspruchsvollen Fachsprachen und Forschungsmethoden, sondern schränkt auch die Aussagekraft und Genauigkeit gesammelter Daten ein. In der Folge setzen sich Projekte durch, welche den sehr spezifischen Anforderungen demokratischer Öffnung gerecht werden. Wenn zudem Projekte mit leicht kommunizierbaren Zielsetzungen systematisch gefördert werden, kann das den Forschungsprozess auch für populistische Tendenzen öffnen. Insgesamt dürfte es daher kein Zufall sein, dass aus Projekten wie NanOpinion oder VOICES weniger einschlägige wissenschaftliche Publikationen als vielmehr *policy briefs* und *booklets* hervorgehen. Die Diversität von Forschung ist in diesem Fall reduziert durch die eingeschränkten Anwendungsbereiche demokratisch-partizipativ organisierter Forschung.

In der Summe ergibt sich der beschriebene Diversitätsverlust also dadurch, dass die Menge an Fragestellungen, Gegenständen, Theorien und Methoden in den sich immer weiter ausdifferenzierenden Disziplinen und Forschungsfeldern notwendigerweise eine viel größere Vielfalt an Forschung repräsentiert als es der jeweils aktuelle Stand an Bewertungsmodi, Bewertungsverfahren und Relevanzkriterien je abzubilden vermag. Genau genommen handelt es sich bei der Evaluationsforschung selbst um nur eines von tausenden Forschungsfeldern, so dass notwendig ein Komplexitätsgefälle besteht zwischen der Vielfalt der Wissenschaften und der Vielfalt der Methoden, die diese Wissenschaften in ihrer Relevanz bewerten. Es ist dann nicht sinnvoll erwartbar, dass ein noch so ausgefeiltes Bewertungssystem jedem Forschungsprojekt und jeder disziplinären Perspektive gleichermaßen gerecht wird.

## **5. WIE SOLL DIE GESELLSCHAFTLICHE RELEVANZ VON FORSCHUNG BEWERTET WERDEN?**

Die Frage, ob die gesellschaftliche Relevanz der Forschung bewertet werden soll, kann mit gutem Gewissen nur dann beantwortet werden, wenn zugleich eine Art Nebenfolgenabschätzung vorgenommen wird. Eine solche Nebenfolgenabschätzung wiederum erfordert zunächst eine Systematisierung der existierenden Konzepte und Methoden der Relevanzbewertung. Wir haben deshalb vorgeschlagen, einen historisch-narrativen, einen standardisiert-administrativen und einen demokratisch-partizipativen Bewertungsmodus zu unterscheiden. Die drei Bewertungsmodi fundieren ihr Urteil auf unterschiedliche Weise und stützen sich auf je eigene Relevanzbegriffe. Eine Gemeinsamkeit besteht jedoch darin, dass jeder Bewertungsmodus, sobald er als universeller und verbindlicher Modus konzipiert wird und damit die flächendeckende Institutionalisierung entsprechender Bewertungsverfahren anleitet, auf die zu bewertende Forschung zurückwirkt und ihre Vielfalt einschränkt. Unsere Diskussion dieser zentralen Problematik des Diversitätsverlusts schließt nicht aus, dass es weitere problematische Nebenwirkungen gibt, die ebenfalls berücksichtigt werden müssten. Nun muss aus all dem jedoch nicht der Schluss gezogen werden, dass man vorsichtshalber ganz auf Relevanzbewertungen verzichten sollte. Im Gegenteil, es gibt, wie wir gezeigt haben, gute Gründe für die Bewertung der gesellschaftlichen Relevanz von Forschung. Doch die Vielfalt der Antworten auf die „Wie“-Frage zwingt uns dazu, die Diskussion auf eine andere Ebene zu heben. Wir kommen abschließend also auf die Frage zurück, *wie* die gesellschaftliche Relevanz von Forschung bewertet werden soll und möchten vorschlagen, die Pluralisierung der Bewertungsverfahren nicht als Problem, sondern als Teil der Lösung zu betrachten.

Auf die Relevanzfrage gibt es eine Vielzahl möglicher Antworten. Jede Bewertung gesellschaftlicher Relevanz geht von eigenen Relevanzbegriffen aus, folgt eigenen Perspektiven und findet in je eigenen Grenzen statt. Folgt man dieser Perspektive, wird es insbesondere darauf ankommen, dass jeder Bewertungsmodus und jedes Bewertungsverfahren diese eingebaute Partialität innerhalb des gesamten Wissenschaftssystems berücksichtigt. Sinnvoll erscheint vor diesem Hintergrund ein Bewertungssystem, das, ähnlich den Wissenschaften selbst, eine Pluralität von Perspektiven in sich abbildet. Die von uns unterschiedenen drei Bewertungsmodi und die zahlreichen Ausformungen dieser Idealtypen zeigen, dass wir dazu bereits heute über ein vielfältiges Instrumentarium verfügen: So kann die Relevanz von Forschung vor dem Hintergrund langfristiger historischer

Entwicklungen eingeschätzt, systematisch durch Patente, Spin-Offs oder andere Indikatoren gemessen, oder über den Einbezug der Öffentlichkeit in den Forschungsprozess hergestellt werden. Alles in allem entspricht der verfügbare *pool* an Bewertungsinstrumenten aber in keiner Weise der für einzelne Beobachter/innen unüberschaubaren Komplexität des modernen Wissenschaftssystems. So ist es eine immer nur im Einzelfall beantwortbare Frage, welche Bewertungshorizonte für welche Art von Forschung geeignet sind. Um der Vielfalt und Komplexität von Forschung gerecht zu werden, müssten Indikatoren so breit differenziert werden, dass dies ihre flächendeckende Anwendung konterkarieren würde.

Von dieser Diagnose eines Komplexitätsdefizits der Bewertungslandschaft ist es nur ein kurzer Weg hin zu den oben bereits aufgeführten Argumenten gegen jegliche externe Bewertung der Wissenschaft: Verfechter/innen einer möglichst hohen Forschungsautonomie postulieren, dass nur Spezialist/inn/en in der Lage seien, über die Forschung in ihrem jeweiligen Feld zu urteilen. Weil aber dieses klassische Autonomieideal wenig Raum für die faktisch gegebenen, unzähligen Inspirationen und Irritationen lässt, die die Forschung ihren gesellschaftlichen Anlehnungskontexten verdankt, erscheint es uns nicht als zielführend, wissenschaftliche Autonomie gegen eine dann als heteronom wahrgenommene Relevanzbewertung auszuspielen. Vielmehr wäre daran zu arbeiten, die drei idealtypischen Bewertungsmodi intern weiter auszudifferenzieren, um so eine möglichst breite Palette von konkreten Bewertungsverfahren zur Verfügung zu haben. Weiter ginge es darum, jede universelle und allgemeinverbindliche Anwendung von Bewertungsmodi und Bewertungsverfahren zu unterlaufen, um stattdessen fallabhängig nur die jeweils angemessenen Relevanzbegriffe und -kriterien produktiv werden zu lassen. In einer idealen Welt hätten Wissenschaftler/innen dann die Möglichkeit, die eigene Forschung vor dem Horizont einer Vielfalt von Bewertungsmodi und Bewertungsverfahren zu spiegeln und zu verantworten. Zugleich hätten sie die Freiheit, Relevanzbegriffe und Relevanzkriterien immer dann zu ignorieren, wenn sie sich nicht als produktiv für die eigene Forschung erweisen. Die Autonomie der Wissenschaft wäre so keine bloß negative „Freiheit von“, sondern eine positive „Freiheit zu“, d.h. eine Freiheit, sich immer dann auf Relevanzdiskurse einzulassen, wenn diese einen Gewinn für die Forschung und eine Horizonterweiterung bedeuten. Autonomie wäre dann, kurz gesagt, die Freiheit zur Heteronomie. Gesellschaftliche Relevanz wäre dementsprechend nicht über standardisierte Bewertungsverfahren zu diagnostizieren, sondern als sich selbst immer weiter ausdifferenzierender Horizont von Relevanzperspektiven zu erschließen, die jeweils in ganz spezifischen Situationen einen produktiven und vielleicht nur momenthaften Anschluss wissenschaftlicher Erkenntnisse an gesellschaftliche Problemlagen ermöglichen. Die Frage wäre dann nicht mehr, ob die gesellschaftliche Relevanz von Forschung bewertbar ist „und wenn ja, wie“, sondern: „und wenn ja, wie viele“.

## ANMERKUNGEN

- <sup>1</sup> Hans Blumenberg, *Das Lachen der Thrakerin. Eine Urgeschichte der Theorie*, Frankfurt am Main 1987; Christina Schües, „Das Lachen der thrakischen Magd. Über die ‚Weltfremdheit‘ der Philosophie“, *Bochumer Philosophisches Jahrbuch für Antike und Mittelalter* 13 (2008), 15–31.
- <sup>2</sup> Seymour L. Chapin, „A Legendary Bon Mot? Franklin’s ‚What Is The Good of a Newborn Baby?‘“, *Proceedings of the American Philosophical Society* 129 (1985), 278–290; Bernard I. Cohen, „Faraday and Franklin’s ‚Newborn Baby‘“, *Proceedings of the American Philosophical Society* 131 (1987), 177–182.
- <sup>3</sup> C. W. Sherwin, R. S. Isenson, „Project Hindsight: A Defense Department Study of the Utility of Research“, *Science* 156 (1967), 1571–1577.
- <sup>4</sup> Peter Thompson, „TRACES: Basic Research Links to Technology Appraised“, *Science* 163 (1969), 374–375.
- <sup>5</sup> Alle im Rahmen des REF evaluierten Impact-Fallstudien sind online abrufbar unter der URL: <http://impact.ref.ac.uk/CaseStudies/>
- <sup>6</sup> London Mathematical Society, *Impact in the Mathematical Sciences in REF2014*, A discussion paper of the London Mathematical Society, London 2011; British Philosophical Association, *Impact in the Research Excellence Framework*, London 2009.
- <sup>7</sup> „Wissenschaft mit der und für die Gesellschaft“ (*science with and for society*) beschreibt ein Einzelziel des EU „Horizon 2020“ Programms. URL: <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/scienceand-society>; letzter Zugriff 14.08.2018.
- <sup>8</sup> Zum Beispiel bei Philipp Kitcher, *Science, Truth, and Democracy*, Oxford 2001.
- <sup>9</sup> Michael Power, *The Audit Society. Rituals of Verification*, Oxford 1997.

- <sup>10</sup> Für eine prominente und historisch einflussreiche Version dieses Standpunktes siehe Michael Polanyi, *The Logic of Liberty. Reflections and Rejoinders*, London 1951.
- <sup>11</sup> Siehe z.B. Richard Münch, *Die akademische Elite. Zur sozialen Konstruktion wissenschaftlicher Exzellenz*, Frankfurt am Main 2007.
- <sup>12</sup> David Kaldewey, *Wahrheit und Nützlichkeit. Selbstbeschreibungen der Wissenschaft zwischen Autonomie und gesellschaftlicher Relevanz*, Bielefeld 2013.
- <sup>13</sup> Wendy Espeland, Michael Sauder, „Rankings and Reactivity. How Public Measures Recreate Social Worlds“, *American Journal of Sociology* 113 (2007), 1–40.
- <sup>14</sup> Siehe dazu auch Steven Shapin, „Invisible Science“, *The Hedgehog Review* 18(3) (2016).
- <sup>15</sup> Auf die Bedeutung des puzzle solving für den Fortgang der Wissenschaft verweist bereits Thomas S. Kuhn, *The Structure of Scientific Revolutions*, Second edition, enlarged, Chicago/London 1970.

**Die Autor/inn/en:**

Dr. Julian Hamann | Leibniz Center for Science and Society (LCSS), Universität Hannover

Prof. Dr. David Kaldewey | Forum Internationale Wissenschaft (FIW), Universität Bonn

Julia Schubert | Forum Internationale Wissenschaft (FIW), Universität Bonn