

3. Messen und Ermessen: Vertrauen in Zahlen oder Expertise für technische Sicherheit und Wissenschaftsförderung

Karin Zachmann, Stefan Esselborn, Ruth Müller und Kay Felder

Messen und Ermessen beschreiben Praktiken des Beweisens und Begründens, zwischen denen sich in (spät-)modernen Gesellschaften ein spannungsreiches Verhältnis entwickelt. Vom Sprachbild her unterscheiden sich Messen und Ermessen nur durch eine kleine Vorsilbe mit zwei Buchstaben. Aber dieser kleine Unterschied auf der Ebene der Zeichen zeigt eine große Differenz auf der semantischen Ebene an. Messen operiert mit Zahlen und Metriken und ist mit der Erwartung eindeutiger Ergebnisse verbunden.¹ Ermessen hingegen mobilisiert Erfahrung, die Unterschiedlichkeit impliziert, weil sie im Individuum verankert ist.² Es sind nicht allein die konkreten Verfahrensschritte, sondern vor allem die Modi des Autorisierens, die beide Vorgehensweisen unterscheiden. Messen ist die numerische Operation in Quantifizierungsprozeduren, mit deren Hilfe das Argumentieren für Übereinkünfte, also die Erzeugung von Evidenz als Modus des Beweisens und Begründens, auf das Medium der Zahlen verlagert wird. Im Unterschied dazu ist die Praxis des Ermessens an die Autorität von Expert_innen gebunden. Hier geht es darum, Bewertungen und Entscheidungen auf der Grundlage von Erfahrungen mittels Abwägung vorhandenen Wissens bzw. vorliegender Fakten zu treffen. In der Rechtswissenschaft ist das Ermessen ein fest eingeführter Fachbegriff, der einen Handlungsspielraum in der Rechtsanwendung erfasst. Ermessensurteile bzw. die Validierung von Daten und Fakten auf der Grundlage der Ab- oder Einschätzung, Beurteilung und Gewichtung durch Expert_innen sind jedoch keineswegs auf den Rechtsbereich begrenzt. Sie spielen vor allem

-
- 1 Die dabei erzielte Eindeutigkeit ist dabei aber nur um den Preis von Vereinfachungen zu haben, mit denen Komplexität reduziert wird. Alles, was sich nicht in die numerische Sprache übertragen lässt, wird ausgeblendet. Vgl. dazu auch die Diskussion der Arbeiten von Heintz und Desrosières im folgenden Abschnitt.
 - 2 Infolge der Personengebundenheit sind Ermessensbegründungen immer perspektivisch und damit ist ihre Eindeutigkeit von vornherein begrenzt. Als Klassiker zur Personengebundenheit von Erfahrungswissen vgl. Michael Polanyi, *Personal Knowledge. Towards a Post-Critical Philosophy*, Chicago 1974.

dort eine Rolle, wo das Vertrauen in Expert_innen hoch ist oder wo Messen nicht möglich ist bzw. dessen Ergebnisse nicht überzeugend sind.

Das Kapitel *Messen und Ermessen* verfolgt das Anliegen, beide Verfahren als spezifische Evidenzpraktiken in (spät-)modernen Wissensgesellschaften zu charakterisieren, die dabei entstehenden Arrangements aus Daten, Modellen, Verfahren, Institutionen, Expert_innen, Computern, Rechtsnormen etc. offenzulegen und das Verhältnis beider Praktiken zueinander zu untersuchen. Dabei gilt es zu prüfen, inwieweit die seit den 1960er Jahren rasant zunehmende und durch Computerisierungsprozesse beförderte Entwicklung und Verwendung numerischer Verfahren, Modelle und quantitativer Indikatoren in Wechselwirkung mit gesellschaftlichen Veränderungen erfolgt, die quantitative Evidenz auf Kosten der Erfahrung von Expert_innen favorisiert. Dafür spricht die These, dass in partizipativ-demokratisch verfassten Gesellschaften die Autorität von Expert_innen zunehmend hinterfragt wird³ und die Transparenz regelgebundener Verfahren sowie die Selbstreferentialität numerischer Daten Partizipationsmöglichkeiten zumindest verspricht. Zudem haben sich die Techniken der Datengewinnung und -verarbeitung seit dem Zweiten Weltkrieg rasant verbessert, so dass die Praktiken des Messens in immer mehr Bereichen Anwendung finden. Es ist jedoch auch zu vermuten, dass die steigende Quantifizierung von Evidenz und die Flut an möglichen Indikatoren für ihre Bewertung die Hinwendung zum Ermessensurteil der Expert_innen verstärkt. Und die den Quantifizierungsprozessen vorausgehenden Ermessensurteile treten wieder ins Zentrum der Öffentlichkeit, wenn Zahlen keine Glaubwürdigkeit erlangen. Das konnte am Jahreswechsel zu 2019 gut in der Debatte um Stickoxidgrenzwerte beobachtet werden, als die dem Zahlenwert zugrunde liegenden Annahmen strittig wurden.⁴ Hingegen manifestiert sich in den Ansätzen zur Gewichtung von Evidenz (*weight of evidence*), die im Zusammenhang mit der Forderung nach evidenzbasierten Entscheidungen in immer mehr Bereichen eingeführt werden, eine Tendenz, jegliche Evidenz, also auch Ermessensurteile von Expert_innen, quantitativ zu bewerten.⁵ Zu betrachten, was gemessen und was ermes-

3 Harry Collins u. Robert Evans, *Rethinking Expertise*, Chicago 2007.

4 Christiane Köllner, Köhler-Papier erweist der Stickoxid-Debatte einen Bärendienst, Onlineartikel 7.2.2019: <https://www.springerprofessional.de/stickstoffoxide/emissionen/koehler-papier-erweist-der-stickoxid-debatte-einen-baerendienst/16443526> [Stand: 4.3.2019].

5 Vgl. Julie E. Goodman, Lisa A. Bailey et al., A Survey of Frameworks for Best Practices in Weight-of-Evidence Analyses, in: *Critical Reviews in Toxicology*, 43, H. 9, 2013, S. 753–784: doi.org/10.3109/10408444.2013.832727.

wird, und wie die Praktiken konkret zusammenwirken, um überzeugende Argumente zur Auflösung gesellschaftlicher Konfliktlagen zu gewinnen, wird dazu beitragen, die Rolle des Wissens als umkämpfte Ressource für politische und gesellschaftliche Entscheidungsprozesse besser zu verstehen.

Empirische Grundlage für die Analyse des Messens und Ermessens sind zwei Fallstudien: eine historische, die Evidenz für Sicherheit, und eine soziologische, die Evidenz für Exzellenz zum Gegenstand hat. In diesem Kapitel gilt es zu untersuchen, wie, unter welchen konkreten soziopolitischen Bedingungen und mit welchen Konsequenzen die Frage verhandelt bzw. entschieden wurde, ob Evidenz für Sicherheit oder Leistung besser über Verfahren der Messung und Berechnung oder aber im Vertrauen auf implizites und erfahrungsbasiertes Wissen von Expert_innen zu gewinnen sei – oder in ihrer Kombination. Sind Messen und Ermessen distinkte Evidenzpraktiken, die nicht nur verschieden wirken, sondern scheinbar auch konkurrieren? Oder werden Messen und Ermessen vor allem als zwei unterschiedliche Praktiken inszeniert, um den wachsenden Herausforderungen an Geltungsansprüche für Wissen und Entscheidungen in unterschiedlichen Kontexten zu genügen? Diesen Fragen wollen wir anhand unserer beiden Beispiele nachgehen. Wir beginnen mit der Referenz auf jene Forschungsliteratur, die uns besonders inspiriert hat.

Die rhetorische Macht quantitativer Evidenz im Verhältnis zum Ermessensurteil von Expert_innen

Quantifizierung ist in der zweiten Hälfte des 20. Jahrhunderts zu einem omnipräsenten Phänomen geworden. Das Digitale Wörterbuch der deutschen Sprache hat das „gemessen“, als Häufigkeit der Wortverwendung im Zeitverlauf.⁶ Diese ist seit dem Ende der 1950er und nochmal seit dem Ende der 1980er Jahre rasant angestiegen. Das verweist auf eine voranschreitende Umstellung gesellschaftlicher Kommunikation und Aushandlungsprozesse auf das Medium der Zahlen. Die Anfänge dieser Transformation gehen schon auf die Herausbildung der amtlichen Statistik im 18. und 19. Jahrhundert zurück, aber erst in der zweiten Hälfte des 20. Jahrhunderts wird sie als Alltagsphänomen signifikant. Was kann diesen Befund erklä-

6 DWDS-Wortverlaufskurve für „Quantifizierung“, erstellt durch das Digitale Wörterbuch der deutschen Sprache: <https://www.dwds.de/r/plot?view=1&corpus=zeitungen&norm=date%2Bclass&smooth=spline&genres=0&grand=1&slice=1&prune=0&window=3&wbase=0&logavg=0&logscale=0&xrange=1945%3A2017&q1=Quantifizierung> [Stand: 19.2.2019].

ren? Und welche Konsequenzen hat diese Umstellung für die Modi des Wissens und Begründens? Ansätze zur Beantwortung dieser Fragen finden sich in soziologischen und historischen Studien zur Quantifizierung, deren Erkenntnisse Ausgangspunkt für unsere Überlegungen zu den Evidenzpraktiken des Messens und Ermessens sind.⁷

Grundlage für die voranschreitende Quantifizierung ist erfolgreiche Kommunikation im Medium der Zahlen. Deren Unterschied zu anderen Kommunikationsformen kann die rhetorische Macht quantitativer Evidenz verständlich machen. Die Soziologin Bettina Heintz hat diese Besonderheit als „numerische Differenz“ charakterisiert und an vier Merkmalen festgemacht.⁸ Ein erstes Merkmal folgt aus den Erzeugungsregeln der numerischen Sprache. Sie ist unter der Voraussetzung von Zahlen- und Rechenkompetenz in hohem Maße selbstexplikativ und nicht auf externe Kommentare angewiesen. Zugleich unterbindet die hochgradige Normierung der numerischen Sprache Kontingenz und erleichtert damit die Zustimmung. Als zweites Merkmal wird Selbstreferentialität genannt. Die Grundkategorien der Mathematik sind innermathematisch begründet und Zahlen werden intern bestätigt, wenn sie nur noch mit anderen Zahlen ins Verhältnis gesetzt werden. Drittens ist die numerische Sprache praktisch universell, muss nicht übersetzt werden und ist so anschlussfähig in ganz unterschiedlichen kulturellen Kontexten. Damit ist die Kommunikation über numerische Sprache per se intersubjektiv, während interlinguale Kommunikation wegen der Polysemie der Sprachen nur einen pragmatischen und nie einen umfassenden Konsens ermöglicht. Viertens präsentieren sich die strikt regelgebundenen Verfahren der Mathematik, wie z.B. der formale Beweis, als logisch stringent und entsubjektiviert. Diese spezifischen Eigenschaften der Kommunikation im Medium der numerischen Sprache macht diese, so Heintz, besonders glaubwürdig und geeignet, Evidenz als Zustimmung herzustellen.

Aber wie operiert Quantifizierung, so dass die der numerischen Sprache zugeschriebene besondere Glaubwürdigkeit auch wirken kann? Ausführlich hat dazu der Soziologe Alain Desrosières gearbeitet.⁹ Quantifizierung

7 Die Soziologie entdeckte die Quantifizierung unlängst als neues Forschungsfeld. Rainer Diaz-Bone u. Emmanuel Didier (Hg.), *Conventions and Quantification – Transdisciplinary Perspectives on Statistics and Classifications*, Special Issue, in: *Historical Social Research* 41, H. 2, 2016.

8 Bettina Heintz, *Numerische Differenz. Überlegungen zu einer Soziologie des (quantitativen) Vergleichs*, in: *Zeitschrift für Soziologie* 39, 2010, S. 162–181.

9 Alain Desrosières, *Die Politik der großen Zahlen. Eine Geschichte der statistischen Denkweise*, Berlin 2005.

erfolgt in zwei Schritten, der Konvention bzw. dem Übereinkommen und der Messung. Im ersten Schritt werden neue Objekte erzeugt, numerische Ausdrücke wie z.B. die Arbeitslosenquote, der Intelligenzquotient, das Risiko (als Produkt aus Eintrittswahrscheinlichkeit eines Schadensereignisses und Schadenshöhe) oder der Abgaswert. Für diesen Schritt sind Ermessensurteile von Experten ausschlaggebend. Danach erfolgt die Messung als regelgebundene Implementierung der Konvention. Die Erzeugung der neuen Objekte setzt die Entwicklung hypothetischer Äquivalenzen voraus, die Vergleiche, Kompromisse, Übersetzungen, Registrierungen, Kodierungen und Berechnungen einschließen. So war z.B. für die Bestimmung des Kernkraftrisikos zu entscheiden, welche Schadensereignisse berücksichtigt werden sollten und wie man die Schadenshöhe definiert. Und bei der Fixierung von Leistungsindikatoren für Wissenschaftler_innen war zu klären, welche Faktoren dafür sowohl signifikant als auch messbar sind.

Wenn die Quantifizierungsprozeduren aber etabliert und zur Routine geworden sind, dann werden die quantifizierten Objekte scheinbar naturalisiert und die ursprünglichen Übereinkommen sind vergessen, solange nicht besondere Ereignisse verlangen, die Blackboxes, zu denen die Zahlenwerte werden, wieder zu öffnen. Mit Desrosières lässt sich Objektivierung also auch als Kodifizierung quantitativer Objekte verstehen, die Inter-subjektivität ermöglichen und Menschen zusammenbringen, weil mit den Objekten eine gemeinsame Sprache entstanden ist, um über sie zu sprechen. Damit lassen sich die neuen Objekte als *boundary objects*¹⁰ begreifen, die eine *trading zone* etablieren, und die gemeinsame Sprache (mit Galison) als Interimssprache.¹¹ Sie ermöglicht Verständigung nicht auf der Grundlage vollständiger Übereinstimmung (*thickness of consensus*), sondern pragmatischer Interpretation (*thinness of interpretation*). In den *trading zones* können Modelle, Graphiken, Simulationen, Wahrscheinlichkeitsaussagen etc. als Objekte von Interimssprachen zirkulieren. Sie ermöglichen die Kommunikation und damit nicht nur die Übertragung von Wissen von den Orten der Erzeugung zu den Orten der Anwendung, sondern auch das Überspielen von Wissensdifferenzen.

10 Zum Konzept der *boundary objects* vgl. Susan Leigh Star u. James R. Griesemer, Institutional Ecology, 'Translations' and Boundary Objects. Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39, in: Social Studies of Science 19, 1989, S. 387–420.

11 Peter Galison, Trading with the Enemy, in: Michael E. Gorman (Hg.), Trading Zones and Interactional Expertise. Creating New Kinds of Collaboration, Cambridge, MA 2010, S. 25–52.

An anderer Stelle ist der Modus der Evidenzgenerierung über quantifizierte Objekte als Abkürzung charakterisiert worden. Sie operiert als Transformation komplexer Zusammenhänge in Einfaches, das unmittelbar einleuchtet, und generiert Eindeutigkeit in einer abgekürzten Darstellung, die ursprüngliche Übereinkommen ausblendet.¹² Dabei spielen auch ästhetische Aspekte eine Rolle, wenn Mathematiker_innen beispielsweise Zwischenschritte in Beweisverfahren im Interesse einer stringenteren Darstellung herauskürzen. Das gilt auch für die Visualisierung der neuen Objekte, seien es Tabellen oder auch Graphen. In den Debatten um Kernkrafttrisiken lässt sich das, wie Stefan Esselborn weiter unten in diesem Kapitel herausgearbeitet hat, sehr gut an den sogenannten „Farmerkurven“ verfolgen. Sie sollen unmittelbar einsichtig machen, was sicher ist und was nicht. Die neuen Objekte und ihre Darstellungsformen, so lässt sich schließen, haben eine starke rhetorische Funktion. Es geht nicht in erster Linie darum, die „Wahrheit“ zu beweisen, sondern darum, zu überzeugen. Erinert sei hier an Richard Rorty: „Unsere Gewissheit wird eine Funktion des Miteinanders von Personen sein, nicht ihrer Interaktion mit einer nicht-menschlichen Realität. Wir werden nicht nach einem unerschütterlichen Fundament Ausschau halten, sondern nach unanfechtbaren Argumenten.“¹³

Unter welchen konkreten soziopolitischen Bedingungen aber war (bzw. ist) Quantifizierung als Evidenzpraktik erfolgreich? Sie war nach Auffassung amerikanischer Wissenschaftshistoriker dort besonders wirkungsmächtig, wo demokratische Strukturen die Autorität von Expert_innen marginalisierten. Das war im 20. Jahrhundert zuerst in den USA der Fall und kann, so Gerd Gigerenzer, erklären, warum gerade die USA eine Führungsrolle in der Entwicklung statistischer und stochastischer Methoden und insbesondere ihrer Anwendung übernommen haben.

This prominence of statistical tools reflects not so much an implicit faith in experts as the weakness of alternative authorities, the absence of strong traditional elites. Where values clash and consensus is elusive, numbers and the techniques that manipulate them are esteemed

12 Michael Cuntz, Barbara Nitsche, Isabell Otto u. Marc Spaniol, Die Listen der Evidenz. Einleitende Überlegungen, in: dies. (Hg.), Die Listen der Evidenz, Köln 2006, S. 22f.

13 Richard Rorty, Der Spiegel der Natur. Eine Kritik der Philosophie, Frankfurt am Main 1987, S. 176.

for their ostensible neutrality. With statistics, hotly debated issues can seemingly be turned into problems to be solved.¹⁴

Gerade die Ausbreitung stochastischer Verfahren in immer neue Anwendungsbereiche markiert eindrücklich die zunehmende Autorität quantitativer Evidenz. Insbesondere im subjektivistischen Zweig der Stochastik, dem Bayesianismus, wird deutlich, dass die Verwendung von Zahlen Gewissheit vermitteln soll, selbst wenn diese sich auf subjektive Annahmen und nicht auf Fakten und Daten gründet. Wahrscheinlichkeiten werden als „Grad vernünftiger Glaubwürdigkeit“ akzeptabel sowohl im modernen Wissenschaftsverständnis, das hypothetische Aussagen als wissenschaftliches Wissen einbezieht,¹⁵ als auch in gesellschaftlichen Aushandlungs- und Entscheidungsprozessen, wo die Stringenz stochastischer Verfahren Vertrauen in das Ergebnis gewährt.¹⁶

Auch Theodore Porter fand für die rhetorische Macht quantitativer Evidenz besondere Gründe in den soziopolitischen Traditionen der USA, wobei sein Fokus auf der Anwendung von Messergebnissen im Bereich politischen Entscheidens liegt. Das starke Misstrauen gegenüber autoritativem Expert_innenwissen in einer politischen Kultur demokratischer Offenheit artikulierte sich hier in wirkungsmächtigen Forderungen nach Transparenz des Regierungshandelns. Nicht verantwortliche Ermessensentscheidungen, sondern das Befolgen von Regeln entsprach diesem Bedürfnis, und das war am besten mit quantifizierbaren Kriterien zu erreichen. Wie Vertrauen in Zahlen das Misstrauen in die Regierung kompensiert, das hat Porter anhand der Genese der Kosten-Nutzen-Analyse in den USA überzeugend demonstriert.¹⁷ Mit der Begrenzung des Sagbaren auf das Berechenbare konnten Zweifel und Komplikationen als potentielle Konfliktquellen ausgeblendet werden. Die standardisierten Quantifizierungstechniken ersparten den verantwortlichen Funktionären Begründungen und simulierten quasi-automatisierte Verfahren der Entscheidungsfindung, in-

14 Gerd Gigerenzer et al., *The Empire of Chance. How Probability Changed Science and Everyday Life*, Cambridge 1989, S. 236.

15 Gregor Schiemann, *Wahrheitsgewissheitsverlust. Hermann von Helmholtz' Mechanismus im Aufbruch der Moderne. Eine Studie zum Übergang von klassischer zur modernen Naturphilosophie*, Darmstadt 1997, S. 138.

16 Gigerenzer (1989).

17 Theodore Porter, *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life*, Princeton 1995, bes. Kapitel 7 und ders., *Objectivity as Standardization. The Rhetoric of Impersonality in Measurement, Statistics, and Cost-Benefit Analysis*, in: Ellan Megill (Hg.), *Rethinking Objectivity*, Durham u. London 1994, S. 197–236.

dem sie die eigentlichen Ermessensentscheidungen verdeckten. Die aber hatten im Vorfeld stattgefunden, als die Parameter für die Berechnungen festgelegt wurden.

Das Misstrauen gegenüber dem Staat, das in den USA das Vertrauen in Zahlen so wirkungsvoll befördert hat, fand im späten 20. Jahrhundert eine Neuauflage im Aufschwung des Neoliberalismus in der ganzen westlichen Welt. Kern dieser Entwicklung ist der Umbau von Staat und Gesellschaft nach dem Muster privater Unternehmen und das Regieren im Modus des New Public Management. Das ist verbunden mit der extensiven Nutzung von Benchmarks und Indikatoren zur Leistungsmessung, gemäß der Maxime von Beratungsfirmen: „If you can’t measure it, you can’t manage it.“¹⁸ Es mündete in eine Entwicklung, die Jerry Muller als Tyrannei der Metriken für die Bereiche der Universitäten und Hochschulen, die Schulen, die Medizin, die Polizei, das Militär, das Wohltätigkeitswesen und den Unternehmens- und Finanzsektor in den USA analysiert und bewertet hat. Er arbeitet primär die nicht intendierten negativen Folgen forciert Quantifizierung heraus.¹⁹

Richard Rottenburg und Koautor_innen haben die Proliferation von quantitativer Evidenz ganz grundsätzlich als Kernelement der neoliberalen Transformation des Regierens charakterisiert.²⁰ Die Studien untersuchen an konkreten Einzelfällen, wie die Techniken der Quantifizierung gesellschaftliche Aushandlungs- und politische Entscheidungsprozesse verändern und damit zur Neugestaltung des Verhältnisses von Markt, Staat und Gesellschaft beitragen. Hieran knüpfen unsere Fallstudien an. Dabei stellen wir die Frage in den Mittelpunkt, inwieweit die in unseren konkreten Fällen zu beobachtenden Praktiken des Messens und Ermessens und ihre Wechselwirkung Hinweise geben auf Veränderungen a) im Verhältnis von Experten und Gesellschaft und b) in den Strategien zur Herstellung von Übereinstimmung in gesellschaftlichen Konfliktlagen über die Glaubwürdigkeit von Wissen.

18 Jerry Z. Muller, *The Tyranny of Metrics*, Princeton 2018, S. 37. Diese Behauptung wird in der Management-Literatur von Peter Drucker zum Glaubensgrundsatz erhoben. Vgl. Peter Drucker, *The Practice of Management*, London 1955.

19 Vgl. dazu auch die schon vor Muller erschienene und die deutschsprachige Literatur einbeziehende Untersuchung von Steffen Mau, *Das metrische Wir. Über die Quantifizierung des Sozialen*, Berlin 2017.

20 Richard Rottenburg et al. (Hg.), *A World of Indicators. The Making of Governmental Knowledge through Quantification*, Cambridge 2015.

Risikozahlen und Grenzwerte als Evidenz für Sicherheit

Sicherheit avancierte im letzten Drittel des 20. Jahrhunderts zu einem Kernthema gesellschaftlicher Auseinandersetzungen. Vor dem Hintergrund sich ausweitender Krisenerfahrungen in so verschiedenen Bereichen wie der Energieversorgung, dem Umweltschutz, der Wirtschaft, dem Finanzwesen und in den internationalen Beziehungen innerhalb und zwischen den Machtblöcken wurde das Sicherheitsversprechen des Staates umfassender interpretiert und aufgerufen und erlangte so wachsende Bedeutung für die Loyalität der Bevölkerung gegenüber ihrer jeweiligen Regierung. Aber wie konnte Sicherheit nachgewiesen und ihre Einhaltung durchgesetzt werden? Ein Bereich, in dem die gesellschaftlichen Auseinandersetzungen um Sicherheit in den 1970er und 1980er Jahren eskalierten, war der im Kontext der Ölkrise in vielen Ländern forcierte Kernkraftwerksbau. Der Konflikt um die Kernenergietechnik entwickelte sich zum Katalysator für eine zunächst kontroverse, aber langfristig höchst erfolgreich etablierte, quantifizierbare Vorstellung von Sicherheit: das Risiko. An der Entwicklung der probabilistischen Risikoanalyse (PRA) lässt sich zeigen, dass sich die Messung von Sicherheit über die Quantifizierung und den Vergleich von Risiken ganz strategisch auf die Autorität der Zahlen stützte, um über die Eindeutigkeit der numerischen Sprache die Relativierung von Sicherheitsansprüchen gesellschaftlich durchzusetzen.

Das enorme Gefahrenpotential kerntechnischer Anlagen machte den Nachweis ihrer Sicherheit von Beginn an sowohl dringend erforderlich als auch außergewöhnlich schwierig. Da Realtests von Kernreaktoren nicht praktikabel waren, mussten Wege gefunden werden, ausschließlich „auf dem Papier“ Evidenz für nukleare Sicherheit zu erbringen, bevor die betreffenden Anlagen in Betrieb gingen. Zu diesem Zweck entwickelten die U.S.-amerikanischen Behörden in den 1950er Jahren das Konzept des „Maximum Credible Accident“ (MCA), das kurz darauf auch in der Bundesrepublik als „Größter anzunehmender Unfall“ (GAU) übernommen wurde. Prospektive Betreiber mussten nachweisen, dass ihr Reaktor den schwersten Unfall, dessen Eintreten noch als „glaubhaft“ („credible“) eingestuft werden konnte, ohne katastrophale Folgen überstehen würde. In der Praxis wurde dieser schwerste zu beherrschende Unfall für die in der BRD eingesetzten Leistungsreakortypen als Komplettabriss der Hauptleitung des primären Kühlkreislaufs definiert.²¹

21 David Okrent, *Nuclear Reactor Safety. On the History of the Regulatory Process*, Madison, WI 1981, S. 32–34; für die BRD Paul Laufs, *Reaktorsicherheit für Leis-*

Der MCA/GAU lieferte zwar ein eindeutiges Ergebnis („sicher“ oder „nicht sicher“) für das Genehmigungsverfahren, war unter Kerntechnikern aber von Anfang an umstritten. Nicht nur war angesichts der großen Zahl möglicher Unfallverläufe eine einseitige Konzentration auf einen bestimmten Unfalltyp durchaus problematisch. Mindestens ebenso heikel war die Einteilung möglicher Unfälle in „glaubwürdige“ Ereignisse, gegen die Vorsorge getroffen werden musste, und rein „hypothetische“, die man vernachlässigen konnte. Da keine eindeutigen Kriterien existierten, blieb dies letztlich der relativ willkürlichen Ermessensentscheidung der zuständigen Experten überlassen. Als Alternative empfahl eine Gruppe internationaler Reaktorsicherheitsexperten eine probabilistische Herangehensweise. Da absolute Sicherheit sowieso nie erreichbar sei, so der Kanadier Ernest Sidall 1959, müsse man versuchen, das Risiko der Kerntechnik quantitativ zu erfassen und so mit anderen Risiken vergleichbar zu machen.²² Sein britischer Kollege Frank Farmer schlug 1967 einen festen Risikogrenzwert für Kernreaktoren vor, der beispielsweise auf 0,01 durchschnittlich erwartete Todesfälle pro Betriebsjahr festgesetzt werden könne. Rechnete man diesen Wert entsprechend um, ließ er sich sogar graphisch darstellen. In Farmers Risikokurve, die die erwartete Häufigkeit verschiedener Unfalltypen mit der dabei jeweils zu erwartenden Freisetzung von Spaltprodukten korrelierte, erschien die Grenze zwischen dem zulässigen und dem unzulässigen Risikobereich nun als klare Linie.²³

In der Bundesrepublik standen die führenden Reaktortechniker solchen Vorschlägen zunächst eher skeptisch gegenüber. Wie eine 1966 in Jülich durchgeführte Fachtagung zum Thema ergab, hielt die Mehrzahl der deutschen Experten eine ausreichend zuverlässige Risikoquantifizierung angesichts des Mangels an Daten und realistischen Rechenmodellen bis auf Weiteres für praktisch nicht durchführbar. Nicht wenige bezweifelten zudem, ob eine probabilistische Berechnung des Risikos von Kernreaktoren überhaupt sinnvoll möglich sei, da sich zentrale Einflussgrößen wie die Zuverlässigkeit einzelgefertigter Bauteile oder das Verhalten von Bedien-

tungskernkraftwerke. Die Entwicklung im politischen und technischen Umfeld der Bundesrepublik Deutschland, Berlin 2013, S. 258–269.

22 Ernest Sidall, Statistical Analysis of Reactor Safety Standards, in: *Nucleonics* 17, H. 2, 1959, S. 64–69.

23 Frank R. Farmer, Siting Criteria. A New Approach, in: International Atomic Energy Agency (Hg.), *Containment and Siting of Nuclear Power Plants. Proceedings of a Symposium on the Containment and Siting of Nuclear Power Plants, Held by the International Atomic Energy Agency in Vienna, 3–7 April 1967*, Wien 1967, S. 303–329.

mannschaften quantitativ-statistisch kaum adäquat erfassen ließen. Manchen war auch einfach unwohl bei dem Gedanken, die Idee absoluter Sicherheit aufzugeben und mit dem Tod von Menschen zu rechnen, wie es der probabilistische Ansatz erforderte. Sie weigerten sich, „als einzige Sicherung gegen schwere Störfälle nur eine geringe Eintrittswahrscheinlichkeit zu dulden“,²⁴ wie die Fachzeitschrift *Atomwirtschaft* schrieb. Noch 1970 war in Fachkreisen die Rede von „Probabilisten“ und „Deterministen“ als „zwei feindliche[n] Lager[n]“, die sich in der bundesdeutschen Expertencommunity gegenüberstünden.²⁵

Trotz der erheblichen Bedenken von Expert_innenseite erschien die Idee eines „objektiven“, numerischen Nachweises der (relativen) Sicherheit von Kernkraftwerken vor dem Hintergrund der eskalierenden öffentlichen Kontroverse zu Beginn der 1970er Jahre nicht nur der Atomindustrie, sondern auch der zuständigen Ministerialbürokratie zunehmend attraktiv. Ab Ende 1972 war im Bundesforschungsministerium die Rede von einer umfassenden „Risikostrategie“ für Kernkraftwerke, die sowohl eine schnellstmögliche Risikoquantifizierung als auch die Weiterverwendung der produzierten Zahlen in der öffentlichen Debatte wie im Genehmigungsverfahren vorbereiten sollte.²⁶ Als 1975 in den USA mit der so genannten „Rasmussen-Studie“ die weltweit erste umfassende Risikoanalyse eines Kernkraftwerks erschien, lagen den zuständigen bundesdeutschen Ministerien bereits detaillierte Pläne für eine eigene Forschungsinitiative zur Risikoquantifizierung vor.²⁷

Diese bestand aus zwei miteinander verbundenen Teilen. Flaggschiff des ersten, technisch-quantifizierenden Teils, der sich auf eine möglichst

24 Heinz G. Seipel, Diskussion über das Konzept des Größten Anzunehmenden Unfalls. Bericht über das Fachgespräch des IRS in Jülich, in: *Atomwirtschaft – Atomtechnik* 12, H. 33, 1967, S. 147f.

25 A. Birkhofer, W. Braun, E. H. Koch, O. Kellermann, K.-H. Lindackers u. D. Smidt, Reaktorsicherheit in der Bundesrepublik Deutschland, in: *Atomwirtschaft – Atomtechnik* 15, H. 9/10, 1970, S. 441–448.

26 Bundesarchiv Koblenz, B 106/52527: Entwurf Ergebnisprotokoll. Quantitative Methoden der Sicherheitsbeurteilung, 19.12.1972.

27 Bundesarchiv Koblenz, B 106/52530: Rahmenprogramm für die Entwicklung von quantitativen Methoden zur Sicherheitsbeurteilung von Kernkraftwerken, P. Kafka, MRR – I – 16, März 1974; Technische Risiken kerntechnischer Anlagen und ihre Perzeption. Arbeitsvorschlag des Battelle Instituts, Juli 1974. Zur Rasmussen-Studie und ihren Vorläufern vgl. Thomas R. Wellock, A Figure of Merit. Quantifying the Probability of a Nuclear Reactor Accident, in: *Technology and Culture* 58, 2017, S. 678–721; Rodney P. Carlisle, Probabilistic Risk Assessment in Nuclear Reactors. Engineering Success, Public Relations Failure, in: *Technology and Culture* 38, 1997, S. 920–941.

genaue Berechnung des der Kernenergie inhärenten Risikos konzentrierte, war die 1976 begonnene „Deutsche Risikostudie“ (DRS). Die Studie sollte anfangs nur die Ergebnisse der Rasmussen-Studie für die BRD nachvollziehen, entwickelte sich aber in der Folge zu einem eigenständigen und durchaus umfangreichen Risikoanalyse-Projekt mit einem Finanzrahmen von über 7 Millionen DM, das sich in zwei Phasen letztlich bis zum Ende der 1980er Jahre hinziehen sollte.²⁸

Parallel zu den Anstrengungen zur technischen Risikoquantifizierung lief gleichzeitig eine zweite Untersuchungsreihe zur „nicht-technischen“ Seite der Risikostrategie an, die den Titel „Untersuchungen über technische, organisatorische und gesellschaftliche Voraussetzungen für Risikostrategien“ trug. Ziel des mit 3,4 Millionen DM ausgestatteten Forschungsprogramms war es, mit psychologischen und sozialwissenschaftlichen Methoden zu untersuchen, „wie Risikozahlen aus Risikostudien als Sachargument in der Sicherheitsdiskussion mit der Bevölkerung verwendet werden können“, wie es in einem internen Memorandum hieß.²⁹ Die beteiligten Sozialwissenschaftler_innen verstanden ihre Aufgabe allerdings keineswegs nur als reine Werbekampagne zugunsten der Atomindustrie. Die fertige Studie kritisierte 1980 explizit die Verwendung von Risikozahlen gegenüber der Öffentlichkeit, da dies zu einer künstlichen Verkürzung der gesellschaftlichen Aushandlungsprozesse um die Kerntechnik führe. Mit dem Fokus auf die quantitative Dimension des Risikos werde ausgeblendet, was nicht quantifizierbar sei. Gleichzeitig würden die subjektiven Komponenten der Risikobestimmung unterschlagen und die Diskussion auf ein Feld verschoben, wo spezielle Fachkompetenzen Voraussetzung

28 Vgl. Gesellschaft für Reaktorsicherheit (GRS), Deutsche Risikostudie Kernkraftwerke. Eine Untersuchung zu dem durch Störfälle in Kernkraftwerken verursachten Risiko, Köln 1979, sowie Gesellschaft für Reaktorsicherheit (GRS), Deutsche Risikostudie Kernkraftwerke, Phase B. Eine Untersuchung im Auftrag des Bundesministers für Forschung und Technologie, Köln 1990. Dennoch stellte die DRS, die sich ausschließlich mit dem operativen Risiko von Leichtwasserreaktoren beschäftigte, eigentlich nur einen relativ kleinen Teil eines sehr viel ambitionierteren Plans dar. Um sinnvoll vergleichen zu können, war ursprünglich nicht nur eine umfassende Quantifizierung des Risikos des kompletten Brennstoffkreislaufs für unterschiedliche Reaktortypen vorgesehen, sondern auch verschiedener nicht-nuklearer Industriezweige sowie einiger allgemeiner Lebensrisiken.

29 Bundesarchiv Koblenz, B 106/89178: Ergebnisbericht 2. Sitzung Projektbeirat Risiko, 18.9.1978.

zum Kommunizieren seien.³⁰ Mit der subjektiven Seite der Risikobestimmung sprachen die Autoren die Risikowahrnehmung an. Ebenso bedeutungsvoll aber war die subjektive Komponente in der eigentlichen quantitativen Bestimmung von Kernkrafttrisiken. Sie wird mit Hilfe des Bayesischen Ansatzes der Wahrscheinlichkeitsrechnung eingebracht, der es ermöglicht, das Problem unvollständiger Informationen als mathematischen Ausdruck zu rahmen. Das erlaubt es nun, das fehlende Wissen durch Ermessensannahmen zu ersetzen, um quantifizieren zu können.³¹ Messen und Ermessen sind hier also in engster Weise miteinander verflochten. Problematisch wird diese Verflechtung dann, wenn Ermessensurteile mit den Messverfahren verdeckt und die numerische Differenz strategisch eingesetzt wird.

Die Kernkraftgegner_innen teilten die Kritik an der (aus ihrer Sicht) mangelnden Objektivität der Risikoberechnungen. Dennoch zeigten sich viele von ihnen zumindest anfangs durchaus bereit, sich grundsätzlich auf die Idee der Risikoquantifizierung einzulassen.³² Das lag nicht zuletzt daran, dass sich die dabei produzierten Zahlen – speziell die errechneten Unfallfolgen – bei entsprechender Interpretation auch als Evidenz für die *Unsicherheit* von Kernkraftwerken nutzen ließen. Damit konterkarierte die Anti-AKW-Bewegung das Eindeutigkeitsversprechen der numerischen Sprache. Dies tat sie so erfolgreich, dass die Veröffentlichung des ersten Bandes der DRS 1979 von Kernkraftbefürwortern vornehmlich als Anlass

30 Jobst Conrad u. Camilla Krebsbach-Gnath, Technologische Risiken und gesellschaftliche Konflikte. Politische Risikostrategien im Bereich der Kernenergie. Bericht für das BMI, Dezember 1980, S. 100–104.

31 Im ersten Heft des ab 1981 erscheinenden Journals *Risk Analysis* begründeten zwei renommierte US-amerikanische Praktiker in der privatwirtschaftlichen Risikobewertung von Kernkraftwerken die Verwendung subjektivistischer Wahrscheinlichkeiten in der Risikoquantifizierung. Stanley Kaplan u. B. John Gerrick, On the Quantitative Definition of Risk, in: *Risk Analysis* 1, 1981, S. 11–27. Aber der Bayesianische Ansatz war damals selbst unter den Experten noch höchst umstritten. Rasmussen hatte ihn in seiner Studie zwar extensiv verwendet, aber vorsichtigerweise nur an einer Stelle im Anhang erwähnt. Sharon Bertsch McGrayne, *The Theory That Would Not Die. How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*, Yale 2011, S. 179.

32 Ein Beispiel dafür ist die ausführliche Kritik des Bremer Arbeits- und Umweltschutz-Zentrums an der DRS, das zwar die errechneten Werte, nicht aber den Ansatz an sich angriff; vgl. Hans-Jürgen Danzmann, Kritik der „Stellungnahme zur Deutschen Risikostudie der GRS“ des Bremer Arbeits- und Umweltschutzzentrums (GRS-36), Köln 1981.

zur Sorge wahrgenommen wurde.³³ Auf einer Fachtagung der Gesellschaft für Sicherheitswissenschaft Anfang der 1980er Jahre wandten sich Vertreter der Atomindustrie praktisch unisono gegen eine Ausweitung des Einsatzes von Risikoanalysen – unter anderem mit dem Argument, dass nicht angängig sei, die „Verantwortung der Ermessensausübung durch Gutachter und Genehmigungsbehörden in Genehmigungsverfahren abzuladen auf anonyme Rechenprogramme“.³⁴

Tatsächlich spielt die Risikoanalyse – im Gegensatz zu anderen Staaten wie etwa den USA, die seit den 1980er Jahren quantitative „safety goals“ einführten – in Deutschland bis heute offiziell nur eine untergeordnete Rolle im atomrechtlichen Genehmigungsverfahren.³⁵ Dennoch hat sich die PRA sowohl in der BRD als auch international spätestens seit den 1990er Jahren als Standard etabliert und ist aus der nuklearen Sicherheitstechnik nicht mehr wegzudenken. In der technischen Literatur wird dabei mittlerweile immer auch auf die Grenzen der Aussagekraft von PRA und die in ihnen enthaltenen Ungewissheiten hingewiesen.³⁶ Trotz allem hat sich auch die Verwendung von Risikozahlen in der politischen Diskussion um die Kernenergie als merkwürdig resistent gegenüber aller Kritik erwiesen und überstand sogar Großkatastrophen wie die Reaktorunglücke von Tschernobyl oder Fukushima.³⁷

Obwohl die Risikoquantifizierung durch PRA also nicht – wie ursprünglich erhofft – zu einer Wiederherstellung des gesellschaftlichen Konsenses über die Nutzung der Kernenergie führte, hatte sie doch eine erhebliche Langzeitwirkung. Die neue Evidenzpraxis wirkte als *boundary object* und konstituierte einen neuen Verhandlungsraum, in dem in der

33 Vgl. etwa die Broschüre „argumente“ des Reaktorbauers Kraftwerk Union (KWU), 22.8.1979, H. 2–3, Universitätsarchiv Stuttgart 200/1560.

34 Probabilistische Risikoanalyse, Sicherheitswissenschaftliche Monographien, Bd. 4, hg. v. Wolfgang Braun im Auftrag der Gesellschaft für Sicherheitswissenschaft, Offenbach 1984, S. 351.

35 U.S. Nuclear Regulatory Commission, Safety Goals for Nuclear Power Plant Operation (NUREG-0880), Washington, D.C. 1983. Nach der Katastrophe von Tschernobyl wurde die PRA allerdings auch in der BRD zum Teil der nun alle zehn Jahre vorgeschriebenen Periodischen Sicherheitsüberprüfungen (PSÜ); vgl. Laufs (2013), S. 166.

36 Vgl. etwa IAEA, Safety Assessment for Facilities and Activities. General Safety Requirements. IAEA Safety Standards Series, General Safety Requirements no. GSR Part 4 (Rev. 1), Wien 2016.

37 John Downer, Disowning Fukushima. Managing the Credibility of Nuclear Reliability Assessment in the Wake of Disaster, in: Regulation & Governance 39, H. 6, 2013, S. 287–309; Jacob Darwin Hamblin, Fukushima and the Motifs of Nuclear History, in: Environmental History 17, 2012, S. 285–299.

scheinbar objektiven Sprache der Zahlen Evidenz für die Aushandlung der Sicherheit von (Kern-)Technik generiert werden konnte. Zur Durchführung der sehr aufwändigen und kostenintensiven Analysen ist zudem eine stabile, wissenschaftlich-technische Infrastruktur entstanden, die ein beträchtliches Momentum erzeugt hat.³⁸ Über den Nuklearbereich hinaus ist die Risikoanalyse zu einem veritablen Zweig der boomenden Sicherheitsindustrie geworden. Mittlerweile werden in immer mehr Bereichen Risikozahlen als Grenzwerte oder Sicherheitsziele zur Darstellung und Regulierung von „Sicherheit“ verwendet. Das ist so lange eine unkritische Entwicklung, wie den einer Argumentation mit Zahlen inhärenten Verkürzungen dadurch begegnet wird, dass die dem Messen zugrunde liegenden Ermessensentscheidungen reflektiert werden.

Leistungsindikatoren als Evidenz für Exzellenz

Auf die Frage, was Exzellenz in der Wissenschaft eigentlich ist, beziehungsweise wie sie gemessen werden kann, antwortete die Wissenschaftsforscherin und ehemalige Präsidentin des Europäischen Forschungsrats (European Research Council, ERC) Helga Nowotny: „Ich sage ganz gerne, wenn ich gefragt werde, wie ich Exzellenz definiere: man erkennt sie, wenn man ihr begegnet.“³⁹ Mit dieser Beschreibung zeigt sich klar die Abgrenzung von Exzellenz und ihrem Erkennen zu quantifizierbaren Messinstrumenten von Leistung und Qualität. Exzellenz erscheint als „mehr“ als nur Messbares. Die strategische Verwendung des Begriffs der Exzellenz in wissenschaftspolitischen Diskursen erinnert stark an Versuche, das „Profane“ vom „Heiligen“ oder Besonderen zu trennen⁴⁰ und mit dem Begriff der Exzellenz ein Distinktionsmerkmal herzustellen. Der Begriff der Exzellenz ist vor allem durch seinen großen Einfluss auf wissenschaftspolitische Entscheidungen von Interesse. Exzellenz wird regelmäßig als ein Kerncharak-

38 Zum Konzept des technologischen Momentums vgl. Thomas P. Hughes, *The Evolution of Large Technological Systems*, in: Wiebe Bijker, Thomas Hughes u. Trevor Pinch (Hg.), *The Social Construction of Technological Systems*, 3. Auflage, Cambridge, MA 1993, S. 51–82. Zur Bedeutung von Infrastrukturen für den Erfolg und die Persistenz von Kennzahlen und Indikatoren vgl. Rottenburg et al. (2015), S. 4.

39 Helga Nowotny, Göttinger Universitätsrede „Auf der Suche nach Exzellenz – Wie viel Evaluierung verträgt das Wissenschaftssystem?“, Göttingen 2011: <https://www.youtube.com/watch?v=V9le2irvwk> [Stand: 3.3.2019].

40 Emile Durkheim, *Die elementaren Formen des religiösen Lebens*, Frankfurt am Main 1986 [1912].

teristikum wünschenswerter Wissenschaft und förderungswürdiger Wissenschaftler_innen postuliert und dabei von Wissenschaftler_innen, Politiker_innen und Policy-Makern gleichermaßen als wichtig angesehen.⁴¹

Ein Gremium, das nicht nur den Exzellenz-Begriff über die letzten Jahre weiter popularisiert hat, sondern sich auch bewusst als eine Institution positioniert, die mit ihrem Verständnis von Exzellenz einen Gegendiskurs zu oftmals als oppressiv wahrgenommenen Quantifizierungsdiskursen setzt, ist der ERC. Gegenläufig zu den, beispielsweise von Müller, oft beklagten Quantifizierungstrends in der Wissenschaft entwickelte der ERC als eine der zentralsten Forschungsförderungsinstrumente Europas eine, zumindest teilweise, widerständige Definition von Exzellenz: Evidenz für Exzellenz wird beim ERC nicht nur durch standardisierbare quantitative Indikatoren gemessen, sondern durch Expert_innenurteile, die über diese Faktoren hinaus Qualität erkennen können, „erkannt“. Förderungen durch den ERC gelten heute vermehrt als der Goldstandard für Exzellenz und besonders in der Nachwuchsförderung nimmt der ERC eine wichtige Rolle ein. So ist zum Beispiel die Einwerbung eines ERC Starting oder Consolidator Grants an vielen europäischen Universitäten eine Leistung, die mit großer Wahrscheinlichkeit zur Entfristung führt und vielerorts höher bewertet wird als ähnliche, nationale Förderformate. Als zentrales Qualitätskriterium des ERC wird immer wieder seine Unabhängigkeit von politischen und wirtschaftlichen Interessen genannt. Diese Unabhängigkeit ermöglicht es dem ERC auch im Begutachtungsprozess Ermessensspielräume zu gestalten, in denen Gutachter_innen auf Basis ihrer eigenen wissenschaftlichen Exzellenz wissenschaftliche Leistungen selbstständig bewerten und Qualität erkennen. Für das vorliegende Fallbeispiel möchten wir daher Ergebnisse aus einer größeren qualitativen Interviewstudie mit 20 ERC-Gutachter_innen vorstellen, die Einblicke liefern, wie diese De- und Re-Konstruktion der Messbarkeit von Leistung die Gutachter_innen in ihrer Arbeit beschäftigt und wie sie aber auch ihre eigene Identität als Wissenschaftler_innen und Gutachter_innen konsequenzenreich beeinflusst. Begutachtung im Kontext des ERC findet in der Form von sogenannten Expert_innen-Panels mit jeweils ca. zehn bis 15 Vertreter_innen eines Forschungsbereiches statt. Diese Panels treffen sich in einer annähernd gleichbleibenden Konstellation alle zwei Jahre für einen Zeitraum von ca. sechs bis acht Jahren. Diese Praxis, über einen längeren Zeitraum hinweg immer

41 Jack Stilgoe, Against Excellence, in: The Guardian, 2014: <https://www.theguardian.com/science/political-science/2014/dec/19/against-excellence> [Stand: 3.3.2019].

wieder zusammenzukommen, nimmt dabei zentralen Einfluss auf die Prozesse der Begutachtung und formt Vorstellungen von wissenschaftlicher Exzellenz und davon, wie diese erkannt und bewertet werden kann, auf entscheidende Weise. Peer-Reviewer_innen beim ERC treffen Wertzuschreibungen an bestimmte Lebensläufe und Projekte und definieren dabei innerhalb der Bewertungssituation und über die Jahre hinweg, was Exzellenz im gegenwärtigen Wissenschaftsbetrieb ausmacht.⁴²

Da die Förderungsquote sehr gering ist (über verschiedene Jahre hinweg liegt sie bei den Starting⁴³ und Consolidator Grants⁴⁴ schwankend ca. zwischen neun und 15 Prozent), ist es für die Peer-Reviewer_innen laut den Interviews manchmal herausfordernd, eine zufriedenstellende Auswahl an Kandidat_innen zu finden. Oftmals müssen in den Sitzungen dann auch pragmatische Entscheidungen getroffen werden, in denen Exzellenz wiederum auf relativ einfach kommunizierbare Faktoren, wie beispielsweise Publikationszahlen pro Jahr, heruntergebrochen werden muss. Allerdings überlässt der ERC den Gutachter_innen die Entscheidung formal ihrer eigenen freien Einschätzung und die ERC Gutachter_innen, beschreiben sich und ihre Position in den Interviews in erster Linie als unabhängige Wissenschaftler_innen. In den Interviews erzählen die ERC Peer-Reviewer_innen, dass sie vom ERC aufgrund ihrer eigenen hohen wissenschaftlichen Leistung ausgewählt wurden und diese sie zur Einschätzung von Kolleg_innen befähigt. Der ERC lässt den Gutachter_innen bewusst sehr viel Freiheit, indem sowohl in den Einweisungen als auch den Richtlinien immer wieder betont wird, dass die wissenschaftliche Exzellenz des Antrags das zentralste Bewertungskriterium ist und wirtschaftliche oder politische Aspekte auszuklammern seien. Die wissenschaftliche Qualität wird somit zum entscheidenden Faktor, welcher von einem vom ERC bestimmten, aber trotzdem unabhängigen Expert_innenpanel als exzellent anerkannt und prämiert werden kann. Quantitative Indikatoren spielen aber auch schon in der Vorselektion der Bewerber_innen eine Rolle. In den Richtlinien des ERC gibt es klare Voraussetzungen, wer einen Antrag stellen darf. Diese Richtlinien orientieren sich dabei an nur wenigen, klar messbaren

42 Kathia Serrano Velarde, *The Way We Ask for Money... The Emergence and Institutionalization of Grant Writing Practices in Academia*, in: *Minerva*, 56, H. 1, 2018, S. 85–107.

43 Bundesministerium für Bildung und Forschung, Informationen zum ERC Starting Grant: <https://www.eubuenro.de/erc-stg.htm> [Stand: 3.3.2019].

44 Bundesministerium für Bildung und Forschung, Informationen zum ERC Consolidator Grant: <https://www.eubuenro.de/erc-consolidator-grants.htm> [Stand: 3.3.2019].

und erkennbaren Kriterien, wie beispielsweise dem akademischen Alter oder einer Erstautor_innenschaft. Im Sinne der Idee, dass wissenschaftliche Exzellenz aber nun „mehr“ als nur das klar Messbare ist, bleibt das Erkennen der Exzellenz in einem nächsten Schritt des Begutachtungsverfahrens ein Ermessensurteil von unabhängigen Expert_innen.

Ein wichtiger Aspekt der Performance dieser Unabhängigkeit der Gutachter_innen zeigte sich beispielsweise darin, dass diese oft eine kritische Haltung gegenüber der Standardisierung von Bewertungspraktiken sowie dem Vokabular der Exzellenz des ERC selbst einnahmen. Gutachter_innen betonten in unseren Interviews immer wieder, dass quantitative Faktoren mehr der Vorselektion dienen und nur beschränkte Aussagekraft für Qualität haben. Begriffe wie Exzellenz, Outstanding etc. seien primär Worthül- sen, die über die tatsächliche wissenschaftliche Qualität nichts aussagen können. Dabei berufen sich die Gutachter_innen immer wieder auf ihre eigenen intrinsischen Fähigkeiten, Qualität abseits von Statistiken, Trends und Buzzwords erkennen zu können. Interessanterweise stimmen sie dabei aber in ihrer Definition von Qualität wiederum genau mit dem ERC und seiner Vorstellung vom „Erkennen“ von Exzellenz überein. Denn auch wenn sie den Begriff per se oftmals ablehnen, identifizieren sich die Gutachter_innen mit der seiner vagen Definition zugrunde liegenden Idee, dass zum Erkennen von Exzellenz implizite und nicht-quantifizierbare Ermessensurteile von Expert_innen unersetzbar bleiben. In diesem Begriff und dem damit verbundenen Verständnis von Begutachtung findet sich somit ein altes wissenschaftliches Ideal vom organisierten Skeptizismus der Wissenschaft wieder: Der/die gute Reviewer_in erkennt Qualität unabhängig von Systemen und „modischem“ Vokabular.⁴⁵ Darüber hinaus ermöglicht gerade die Freiheit, die der ERC seinen Gutachter_innen gibt, und die damit einhergehende kritische Haltung gegenüber dem Begutachtungssystem ein vermeintlich unabhängiges Erkennen von Exzellenz.

Damit liefert der ERC einen im wissenschaftlichen System vielerorts herbeigesehnten Gegenentwurf zu bürokratischen und quantifizierbaren Vermessungen von Leistung. Diese subjektive und situierte Definition von Exzellenz bringt aber wiederum andere Probleme und Fragen mit sich. Diese zeigen sich beispielsweise an zwei Themen: Geschlechter- und Nationalitäten-Bias. Die anhaltend schlechteren Erfolgchancen von Frauen und von Bewerber_innen aus ost- und südeuropäischen Ländern werden vom

45 Robert Merton, Wissenschaft und demokratische Sozialstruktur, in: Peter Weingart (Hg.), Wissenschaftssoziologie 1. Wissenschaft als sozialer Prozess, Frankfurt am Main 1972, S. 45-59.

ERC selbst thematisiert und statistisch erfasst.⁴⁶ Laut den Interviews beschäftigt die Gutachter_innen diese Problematik oftmals auch innerhalb von Panel-Diskussionen und bei gemeinsamen Abendessen. Da sich der Geschlechter- und Nationalitäten-Bias beim ERC trotz des Fokus auf „wissenschaftliche Qualität allein“ bislang nicht auflösen ließ, standen die Gutachter_innen diesem Phänomen in den Interviews oft sprachlos gegenüber. Quotensysteme oder Affirmative-Action-Programme⁴⁷ wurden meistens entschieden abgelehnt. Interessanterweise wären hier messbare Quoten und Zahlen eine Antwort auf die bestehenden Ungleichheiten und bislang unüberbrückbaren Benachteiligungen von manchen europäischen Regionen.

Unser Fallbeispiel muss damit auch in einem Kontext gelesen werden, in welchem der ERC mit seinem Fokus auf Ermessensurteile als eine Antwort auf Probleme mit der Quantifizierung interpretiert werden kann. Der ERC verdankt seinen Ruf, eine außergewöhnlich gute Begutachtungspraxis zu pflegen, zum Teil sicher auch der Tatsache, dass er in Zeiten der zunehmenden Quantifizierung eine Ausnahme darstellt. Er nutzt die, u.a. von Muller, postulierte Unzufriedenheit mit der Quantifizierung in der Wissenschaft, um einen vermeintlichen Gegenentwurf zu präsentieren. Gerade in der Wissenschaft wird die mit der Quantifizierung verbundene Simplifizierung oftmals als eine potentiell für die Wissenschaft gefährliche Reduktion von Komplexität gesehen.⁴⁸ Doch speziell bei der Frage nach der Exzellenz – die ja als etwas „Besonderes“ gesehen wird – erscheint diese Reduktion, die Espeland als eine Auslöschung von Narrativen bezeichnet, oftmals als unzureichend. Um das Besondere einer Person beziehungsweise eines Projektes beschreiben zu können, braucht es jene Narrative, um den Kontext herzustellen, in dem das „Besondere“ erkannt werden kann. Das Berufen auf Zahlen und Messungen alleine erscheint dabei als eine der Wissenschaft konträr gegenüberstehende, bürokratische und profane Praxis. Trotzdem spielen Zahlen und Quantifizierungen in der Urteilsfindung

46 Siehe beispielsweise: <https://erc.europa.eu/thematic-working-groups/working-group-gender-balance> [Stand: 22.10.2019].

47 Darunter werden gesellschaftspolitische Maßnahmen verstanden, die der gesellschaftlichen Benachteiligung mit gezielter Vorteilsgewährung entgegenwirken sollen: Thomas Bartscher u. Regina Nissen, Art. „Affirmative Action“, in: Gabler Wirtschaftslexikon: <https://wirtschaftslexikon.gabler.de/definition/affirmative-action-29900> [Stand: 8.5.2019].

48 Wendy Espeland, Narrating Numbers, in: Richard Rottenburg u. Sally Engle Merry (Hg.), *A World of Indicators: The Making of Governmental Knowledge through Quantification*, Cambridge 2015, S. 56–75.

eine nicht unbedeutende Rolle: In den Interviews betonten Gutachter_innen immer wieder, dass in einer ersten Runde der Auswahl oft jene aussortiert werden, deren Publikationsoutput viel zu gering ist oder deren Leistungen hinter den Erwartungen für ihr akademisches Alter liegen. Expert_innenurteile werden vor allem für jene Anträge relevant, die nicht ganz eindeutig als unzureichend oder herausragend identifiziert werden können. Wissenschaftler_innen, die in Zeiten knapper werdender Ressourcen und hohem Wettbewerb erfolgreich sein wollen, brauchen somit immer ausreichend gute Zahlen, um Teil des „Indicator Games“ sein zu können und so überhaupt wettbewerbsfähig zu sein.⁴⁹ Das Ermessensurteil beim ERC wird hier zur nächsten Stufe der Selektion in einem Prozess schärfer werdenden Wettbewerbs.

So zeigt das Beispiel des Messens und Ermessens von wissenschaftlicher Leistung, dass die Tyrannei der Zahlen nicht eindeutig einem Ideal von Freiheit und Inklusion gegenübersteht. Während im Selbstverständnis des ERC wissenschaftliche Freiheit nur dann gegeben ist, wenn Expert_innen unabhängig von außerwissenschaftlich festgelegten Kriterien wie dem H-Index urteilen, zeigen die reflexiven Gespräche in den Interviews auf, dass Entscheidungsprozesse in Peer-Review-Verfahren auch ohne vorgegebene Zahlen normative Vorstellungen von Wissenschaft hervorbringen und diese zur Urteilsfindung auch benötigen. Während in den Interviews oftmals sehr komplexe Vorstellungen von wissenschaftlicher Exzellenz, abseits von Publikationsleistungen und Geradlinigkeit der Lebensläufe, diskutiert werden, treffen die Expert_innen ihre Ermessensurteile sehr wohl im Sinne von normativen Vorstellungen über Wissenschaft, die von maximalem Leistungsoutput in möglichst kurzer Zeit ausgehen. Diese Beobachtung ist verknüpft mit Bestrebungen der Gutachter_innen, trotz der ihnen zugestandenen Freiheit faire und nachvollziehbare Entscheidungen zu treffen. Dabei kommt auch der Panelstruktur selbst eine besondere Rolle zu, in der sich die Gutachter_innen in einem interdisziplinären Setting gegenseitig von ihrer Einschätzung überzeugen müssen. Gerade in diesem Kontext braucht es auch leicht kommunizierbare und nachvollziehbare Argumente, die disziplinäre Grenzen überschreiten. Damit zeigt sich nicht nur wieder einmal die Verschränkung von Messen und Ermessen, sondern es wird auch deutlich, dass die Ideale von Freiheit und Gerechtigkeit in der Urteilsfindung nicht unbedingt Hand in Hand gehen und Quantifizierung

49 Alan Irwin, If the Indicator Game is the Answer, Then What is the Question?, in: Engaging Science, Technology, and Society 3, 2017, S. 64–72: <https://estsjournal.org/index.php/ests/article/view/110/79> [Stand: 3.3.2019].

und Oppression auch nicht unbedingt als logische Kopplung zu denken sind. Denn Quantifizierung ermöglicht ein Sichtbar- und Nachvollziehbar-machen von normativen Vorstellungen, die auch ohne ein explizites Ausweisen immer ein Teil von Ermessensurteilen sein müssen: So könnte beispielsweise eine prozentuale Quantifizierung von Frauen und Bewerber_innen aus bislang beim ERC erfolgloseren Ländern die Möglichkeit bieten, regulierend und steuernd auf scheinbar selbstläuferische Tendenzen zur Normierung von wissenschaftlicher Leistung einzuwirken.

In seiner Vision von Begutachtung greift der ERC das in der Wissenschaft vorherrschende Spannungsverhältnis zwischen Quantifizierung und Exzellenz auf und erzeugt ein Idealbild des besonderen ERC-Gutachters bzw. der besonderen Gutachterin, die als reflexive Expert_innen und exzellente Wissenschaftler_innen auftreten. Die beim ERC zur Schau gestellte Hervorhebung der Bedeutung von Ermessensurteilen schafft auf Seiten der Gutachter_innen ein hohes Vertrauen in das System, aber auch in ihre eigenen Entscheidungen und ihre Rolle als unabhängige Wissenschaftler_innen. Es bleibt jedoch die Frage, inwiefern die Vision von ‚guter Gutachter_innentätigkeit‘, die so in den Panels des ERC erzeugt wird, bestimmte Werteregister in den Vordergrund, andere jedoch in den Hintergrund drängt. Welche Aspekte der Peer-Review, wie beispielsweise Gender- oder Nationalitäten-Bias, werden in dieser Logik des Ermessens als unvermeidbares Übel hingenommen, und mit welchen Konsequenzen? Inwiefern erzeugen die Praxisdefinitionen von Exzellenz, die in den Panels konstruiert und substantiiert werden, bestimmte hegemoniale Vorstellungen von Qualität, und schließen andere aus? Und wie steht das zu dem reflexiven und kritischen Selbstverständnis der Gutachter_innen und ihrer Identitätsarbeit? Ermessensurteile im Kontext von Peer-Review in Forschungsförderungsinstitutionen stellen dabei immer auch Evidenzierungspraktiken für die Verteilung von Ressourcen dar und formen bestimmte Versionen von wissenschaftlicher Qualität und damit letztlich auch Evidenz.

Fazit

Wie lassen sich nun unsere Ergebnisse der beiden Fallstudien zur Quantifizierung von Evidenz für technische Sicherheit resp. für wissenschaftliche Exzellenz verknüpfen und verallgemeinern? In beiden Fällen intendieren Regierungsbehörden (auf nationaler bzw. auf europäischer Ebene) dazu, mittels quantitativer Indikatoren (a: zur Sicherheitsbewertung von Kernreaktoren mittels Risikozahlen bzw. b: zur Leistungsbewertung von Wissenschaftler_innen mittels Kennzahlen) die Allokation von Ressourcen zu len-

ken und dafür gesellschaftliche Zustimmung zu erlangen. In beiden Bereichen beruht die Entwicklung der Indikatoren auf komplexen Übereinkünften (in der Kernkraft sind das die vielstufigen Entscheidungen zum Aufbau von Risikoanalysen; in der Forschungsförderung ist es die Standardisierung von Evaluierungskriterien). In jedem der beiden Fälle ermöglicht die Umstellung der Bewertung auf das Medium der Zahlen Vergleichbarkeit zwischen verschiedenen Optionen und damit Entscheidungen nach dem Muster marktförmigen Verhaltens. Die Anwendung der quantitativen Indikatoren als Evidenz für Sicherheit oder Leistung traf auf beiden Untersuchungsfeldern jedoch keineswegs auf einhellige Zustimmung. So wird im Kernkraftbereich der Bundesrepublik Anfang der 1980er Jahre ausdrücklich davor gewarnt, probabilistische Risikoanalysen „für den Zweck eines öffentlichen Minimierungswettbewerbs errechneter Risiken verschiedener Technologien [zu missbrauchen] in dem utopischen Glauben, eine Gesellschaft mit technisch industriellem Nullrisiko erreichen zu können“.⁵⁰ In Evaluierungsverfahren ist die Angemessenheit von Messverfahren ebenfalls häufig ein Streitpunkt und es zeigt sich ein neuer Rekurs auf das geschulte Expert_innenurteil. Gleichwohl ist in beiden Bereichen der Quantifizierungstrend virulent, der einerseits auf Grundlage auch technisch erweiterter Vergleichs- und Korrelierungsmöglichkeiten neue Einsichten und Entscheidungsoptionen erzeugt, andererseits aber die Gefahr der Verkürzung auf das quantitativ „Sagbare“ mit sich bringt. Für Theodore Porter implizieren die Verkürzungen einen Trend zum „Thinning“, zur Ausdünnung des Wissens, dessen Ergebnisse er in einem Forschungsprojekt unter dem bezeichnenden Titel „Funny Numbers“ untersucht. Funny sind die Zahlen, deren Glaubwürdigkeit auf den technischen Routinen ihrer Erzeugung (durch Messen) beruht, wenn es nicht darum geht, dass sie wahr oder richtig, sondern dass sie scheinbar objektiv sind, und ihren subjektiven (Ermessens-)Charakter verbergen.⁵¹ Die Persistenz von Risikozahlen trotz katastrophaler Unfälle im Kernkraftbereich bekräftigt Porters Argument.

Gleichzeitig regt die Frage danach, wie Probleme der Verteilungsgerechtigkeit im Evaluationsmodell des ERC sinnvoll Berücksichtigung finden könnten, auch Überlegungen dahingehend an, wie Messen und Ermessen in Gremien der Entscheidungsfindung potentiell besser miteinander integriert werden könnten. Während im ERC Ermessen großgeschrieben wird,

50 Braun (1983), S. 351.

51 Theodore Porter, Funny Numbers, in: Culture Unbound 4, H. 4, 2012, S. 585–598: <http://www.cultureunbound.ep.liu.se> [Stand: 4.3.2019].

ist dieses Ermessen eben auch nicht gefeit davor, traditionelle Formen der Über- und Unterbewertung zu perpetuieren. In Kombination mit dem Bias, der schon auftritt, wenn es darum geht, die messbaren Attribute zu erwerben, die die Eintrittskarte zur Arena des Ermessens darstellen (wie etwa die zu Verfügung stehenden Ressourcen für die wissenschaftliche Arbeit im europäischen Süden, oder der bekannte Genderbias in der Peer-Review), entstehen hier signifikante Formen des sozialen Ausschlusses. Während Messen und Ermessen oft in Opposition zueinander gedacht werden, entsteht hier der Bedarf nach Formen der sorgsam Komplementarität, in der vielschichtige Formen des Messens und Ermessens miteinander kombiniert werden können. So zeigen beide Beispiele auf, dass Entscheidungsprozesse immer beide Praktiken beinhalten und es oftmals auch eine Frage der Inszenierung, des „Präsentierens“ oder „Stagings“ ist, welche Praktik in welcher Form und wie hervorgehoben wird. Dabei geht es vor allem um die Frage der Glaubwürdigkeit und darum, welche Formen des Begründens als besonders durchsetzungsstark in einem spezifischen Kontext vermutet werden. Sorgsame Komplementarität und ein Anerkennen der Bedingtheit, aber auch der Stärken und Schwächen beider Begründungspraktiken erscheint dabei als wichtiger und zentraler Schritt zur Entwicklung von vertrauenswürdigen und gleichzeitig nachvollziehbaren Evidenzierungspraktiken in Zeiten von Misstrauen und Vertrauensverlust in Wissenschaft und Expertise.

Literatur

- Bartscher, Thomas u. Nissen, Regina, Art. „Affirmative Action“, in: Gabler Wirtschaftslexikon: <https://wirtschaftslexikon.gabler.de/definition/affirmative-action-29900> [Stand: 8.5.2019].
- Birkhofer, A., Braun, W., Koch, E. H., Kellermann, O., Lindackers, K.-H. u. Smidt, D., Reaktorsicherheit in der Bundesrepublik Deutschland, in: Atomwirtschaft – Atomtechnik 15, H. 9/10, 1970, S. 441–448.
- Bundesministerium für Bildung und Forschung, Informationen zum ERC Consolidator Grant: <https://www.eubuenro.de/erc-consolidator-grants.htm> [Stand: 3.3.2019].
- Bundesministerium für Bildung und Forschung, Informationen zum ERC Starting Grant: <https://www.eubuenro.de/erc-stg.htm> [Stand: 3.3.2019].
- Carlisle, Rodney P., Probabilistic Risk Assessment in Nuclear Reactors. Engineering Success, Public Relations Failure, in: Technology and Culture 38, 1997, S. 920–941.
- Collins, Harry u. Evans, Robert, Rethinking Expertise, Chicago 2007.

- Conrad, Jobst u. Krebsbach-Gnath, Camilla, Technologische Risiken und gesellschaftliche Konflikte. Politische Risikostrategien im Bereich der Kernenergie. Bericht für das BMI, Dezember 1980.
- Cuntz, Michael, Nitsche, Barbara, Otto, Isabell u. Spaniol, Marc, Die Listen der Evidenz. Einleitende Überlegungen, in: dies. (Hg.), Die Listen der Evidenz, Köln 2006.
- Danzmann, Hans-Jürgen, Kritik der „Stellungnahme zur Deutschen Risikostudie der GRS“ des Bremer Arbeits- und Umweltschutzzentrums (GRS-36), Köln 1981.
- Desrosières, Alain, Die Politik der großen Zahlen. Eine Geschichte der statistischen Denkweise, Berlin 2005.
- Diaz-Bone, Rainer u. Didier, Emmanuel (Hg.), Conventions and Quantification – Transdisciplinary Perspectives on Statistics and Classifications, Special Issue, in: Historical Social Research 41, H. 2, 2016.
- Downer, John, Disowning Fukushima. Managing the Credibility of Nuclear Reliability Assessment in the Wake of Disaster, in: Regulation & Governance 39, H. 6, 2013, S. 287–309.
- Drucker, Peter, The Practice of Management, London 1955.
- Durkheim, Emile, Die elementaren Formen des religiösen Lebens, Frankfurt am Main 1986 [1912].
- Espeland, Wendy, Narrating Numbers, in: Richard Rottenburg u. Sally Engle Merry (Hg.), A World of Indicators. The Making of Governmental Knowledge through Quantification, Cambridge 2015, S. 56–75.
- Farmer, Frank R., Siting Criteria. A New Approach, in: International Atomic Energy Agency (Hg.), Containment and Siting of Nuclear Power Plants. Proceedings of a Symposium on the Containment and Siting of Nuclear Power Plants, Held by the International Atomic Energy Agency in Vienna, 3–7 April 1967, Wien 1967, S. 303–329.
- Galison, Peter, Trading with the Enemy, in: Michael E. Gorman (Hg.), Trading Zones and Interactional Expertise. Creating New Kinds of Collaboration, Cambridge, MA 2010, S. 25–52.
- Gesellschaft für Reaktorsicherheit (GRS), Deutsche Risikostudie Kernkraftwerke, Phase B. Eine Untersuchung im Auftrag des Bundesministers für Forschung und Technologie, Köln 1990.
- Gesellschaft für Reaktorsicherheit (GRS), Deutsche Risikostudie Kernkraftwerke. Eine Untersuchung zu dem durch Störfälle in Kernkraftwerken verursachten Risiko, Köln 1979.
- Gigerenzer, Gerd et al., The Empire of Chance. How Probability Changed Science and Everyday Life, Cambridge 1989.
- Goodman, Julie E., Bailey, Lisa A. et al., A Survey of Frameworks for Best Practices in Weight-of-Evidence Analyses, in: Critical Reviews in Toxicology, 43, H. 9, 2013, S. 753–784: doi.org/10.3109/10408444.2013.832727.
- Hamblin, Jacob Darwin, Fukushima and the Motifs of Nuclear History, in: Environmental History 17, 2012, S. 285–299.

- Heintz, Bettina, Numerische Differenz. Überlegungen zu einer Soziologie des (quantitativen) Vergleichs, in: Zeitschrift für Soziologie 39, 2010, S. 162–181.
- Hughes, Thomas P., The Evolution of Large Technological Systems, in: Wiebe Bijker, Thomas Hughes u. Trevor Pinch (Hg.), The Social Construction of Technological Systems, 3. Auflage, Cambridge, MA 1993, S. 51–82.
- IAEA, Safety Assessment for Facilities and Activities. General Safety Requirements. IAEA Safety Standards Series, General Safety Requirements no. GSR Part 4 (Rev. 1), Wien 2016.
- Irwin, Alan, If the Indicator Game is the Answer, Then What is the Question?, in: Engaging Science, Technology, and Society 3, 2017, S. 64–72: <https://estsjournal.org/index.php/ests/article/view/110/79>.
- Kaplan, Stanley u. Gerrick, B. John, On the Quantitative Definition of Risk, in: Risk Analysis 1, 1981, S. 11–27.
- Köllner, Christiane, Köhler-Papier erweist der Stickoxid-Debatte einen Bären-dienst, Onlineartikel 7.2.2019: <https://www.springerprofessional.de/stickstoffoxi-de/emissionen/koebler-papier-erweist-der-stickoxid-debatte-einen-baerendienst/16443526> [Stand: 4.3.2019].
- Laufs, Paul, Reaktorsicherheit für Leistungskernkraftwerke. Die Entwicklung im politischen und technischen Umfeld der Bundesrepublik Deutschland, Berlin 2013.
- Leigh Star, Susan u. Griesemer, James R., Institutional Ecology, ‘Translations’ and Boundary Objects. Amateurs and Professionals in Berkeley’s Museum of Vertebrate Zoology, 1907–39, in: Social Studies of Science 19, 1989, S. 387–420.
- Mau, Steffen, Das metrische Wir. Über die Quantifizierung des Sozialen, Berlin 2017.
- McGrayne, Sharon Bertsch, The Theory That Would Not Die. How Bayes’ Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy, Yale 2011.
- Robert Merton, Wissenschaft und demokratische Sozialstruktur, in: Peter Weingart (Hg.), Wissenschaftssoziologie 1. Wissenschaft als sozialer Prozess, Frankfurt am Main 1972, S. 45–59.
- Muller, Jerry Z., The Tyranny of Metrics, Princeton 2018.
- Okrent, David, Nuclear Reactor Safety. On the History of the Regulatory Process, Madison, WI 1981.
- Polanyi, Michael, Personal Knowledge. Towards a Post-Critical Philosophy, Chicago 1974.
- Porter, Theodore, Funny Numbers, in: Culture Unbound 4, H. 4, 2012, S. 585–598: <http://www.cultureunbound.ep.liu.se> [Stand: 4.3.2019].
- Porter, Theodore, Objectivity as Standardization. The Rhetoric of Impersonality in Measurement, Statistics, and Cost-Benefit Analysis, in: Ellan Megill (Hg.), Rethinking Objectivity, Durham u. London 1994, S. 197–236.
- Porter, Theodore, Trust in Numbers. The Pursuit of Objectivity in Science and Public Life, Princeton 1995.

- Probabilistische Risikoanalyse, Sicherheitswissenschaftliche Monographien, Bd. 4, hg. v. Wolfgang Braun im Auftrag der Gesellschaft für Sicherheitswissenschaft, Offenbach 1984.
- Rorty, Richard, *Der Spiegel der Natur. Eine Kritik der Philosophie*, Frankfurt am Main 1987.
- Rottenburg, Richard et al. (Hg.), *A World of Indicators. The Making of Governmental Knowledge through Quantification*, Cambridge 2015.
- Schiemann, Gregor, *Wahrheitsgewissheitsverlust. Hermann von Helmholtz' Mechanismus im Aufbruch der Moderne. Eine Studie zum Übergang von klassischer zur modernen Naturphilosophie*, Darmstadt 1997.
- Seipel, Heinz G., Diskussion über das Konzept des Größten Anzunehmenden Unfalls. Bericht über das Fachgespräch des IRS in Jülich, in: *Atomwirtschaft – Atomtechnik* 12, H. 33, 1967, S. 147f.
- Serrano Velarde, Kathia, *The Way We Ask for Money... The Emergence and Institutionalization of Grant Writing Practices in Academia*, in: *Minerva*, 56, H. 1, 2018, S. 85–107.
- Siddall, Ernest, *Statistical Analysis of Reactor Safety Standards*, in: *Nucleonics* 17, H. 2, 1959, S. 64–69.
- Stilgoe, Jack, *Against Excellence*, in: *The Guardian*, 2014: <https://www.theguardian.com/science/political-science/2014/dec/19/against-excellence> [Stand: 3.3.2019].
- U.S. Nuclear Regulatory Commission, *Safety Goals for Nuclear Power Plant Operation (NUREG-0880)*, Washington, D.C. 1983.
- Wellock, Thomas R., *A Figure of Merit. Quantifying the Probability of a Nuclear Reactor Accident*, in: *Technology and Culture* 58, 2017, S. 678–721.