

Gesellschaftliche Erwartungen an »Big Data« in der Wissenschaft

Zur Mär vom »Ende der Theorie«

Axel Gelfert

1. Gesellschaftliche Wahrnehmung von Wissenschaft und Technik

Trotz aller Klagen über eine zunehmende Wissenschaftsfeindlichkeit politischer Entscheidungsträger und einer lautstarken Minderheit ›querdenkender‹ Privatpersonen zeigen Erhebungen zum Vertrauen der Öffentlichkeit in die Institution Wissenschaft – von gelegentlichen Schwankungen und einigen, in der Regel eng umgrenzten Themenfeldern abgesehen – eine erstaunliche Konstanz, und zwar auf hohem Niveau. Vor allem in wissenschaftlich-technisch hochentwickelten Industrieländern ist eine deutliche Mehrheit der Bevölkerung der Ansicht, dass die Vorteile wissenschaftlicher Forschung (»benefits of scientific research«) ihre negativen Auswirkungen (»harmful results«) überwiegen. In Deutschland erhebt das von der Initiative *Wissenschaft im Dialog* herausgegebene »Wissenschaftsbarometer« seit 2014 jährlich Daten zur Einstellung der Bürgerinnen und Bürger gegenüber der Wissenschaft. Demnach fluktuierte der Anteil derer, die Wissenschaft und Forschung in Deutschland vertrauen, in den Jahren 2017 bis 2019 um die 50 %-Marke und stieg bis 2021 auf über 60 Prozent. Der beobachtete Anstieg lässt sich auch mit der ab 2020 grassierenden SARS-CoV-2-Pandemie in Verbindung bringen, zeigen doch Zwischenerhebungen, dass das Wissenschaftsvertrauen gerade in der Frühphase

der Pandemie auf Werte über 70 Prozent stieg. Auch in den Vereinigten Staaten liegt der Anteil derjenigen, für die die Vorteile wissenschaftlicher Forschung ihre Nachteile überwiegen, seit rund vierzig Jahren bei deutlich über zwei Dritteln der Befragten; ein signifikanter Wandel dieser allgemein positiven Einschätzung von Wissenschaft ist – trotz verhärteter Fronten in einigen Punkten, so der existentiellen Frage des anthropogenen Klimawandels – nicht festzustellen.¹

So begrenzt die Aussagekraft derartiger Erhebungen in ihrer Pauschalität ist – und so schwierig es nicht nur aufgrund sprachlicher Unterschiede in den verwendeten Formulierungen ist, diese länderübergreifend miteinander zu vergleichen – so legen sie dennoch nahe, dass die öffentliche Akzeptanz von Wissenschaft wenigstens indirekt an eine Nutzenerwartung geknüpft ist. Dazu passt, dass in einer Umfrage der American Academy of Arts & Sciences (AAAS) aus dem Jahr 2018 sowohl bei den am häufigsten mit »Scientific Research« assoziierten Gegenstandsbereichen als auch bei der Frage nach den zu setzenden Forschungsprioritäten *Biomedizin*, *Krankheitsbekämpfung* und *Gesundheitsforschung* an erster Stelle genannt werden.²

Trotz der konstant hohen Wertschätzung wissenschaftlicher Forschung und der damit einhergehenden Erwartungshaltung zeigen Umfragen jedoch auch wachsende Besorgnis darüber, dass die Wissenschaft zu immer schnelleren lebensweltlichen Veränderungen beiträgt (vgl. NSB 2018). Nicht zu unterschätzen ist auch die konstante Präsenz wissenschaftsskeptischer Strömungen, die sich – unter anderem befeuert durch die Möglichkeit zur Selbstorganisation über elektronische soziale Medien – in den letzten Jahren verfestigt haben und sich bisweilen, wie die sog. »Querdenker«-Szene im Falle der SARS-CoV-2-Pandemie, lautstark gegen das Primat wissenschaftlicher Erkenntnis stellen. So stimmten laut Wissenschaftsbarometer 39 Prozent der deutschen Bevölkerung im Jahr 2021 der Aussage eher oder voll zu, Wissen-

-
- 1 Vgl. hierzu das »Wissenschaftsbarometer 2021« sowie das »2018 Digest« der »Science and Engineering Indicators«, erhoben vom National Science Board der US-amerikanischen National Science Foundation (vgl. NSB 2018).
 - 2 Vgl. hierzu den Bericht »Perceptions of Science in America« (AAAS 2018: 3).

schaftler würden »uns nicht alles, was sie über das Coronavirus wissen«, sagen. 40 Prozent stimmten dieser Aussage eher oder gar nicht zu, 19 Prozent waren unentschieden, was das bereits erwähnte Phänomen einer punktuellen Polarisierung der gesellschaftlichen Wissenschaftswahrnehmung widerspiegeln mag.

Peter Weingart hat darauf hingewiesen, dass das Versprechen auf wissenschaftlichen Fortschritt und den sich daraus ergebenden gesellschaftlichen Nutzen notwendigerweise unter dem fundamentalen Vorbehalt steht, »dass der letztgültige Beweis für die Nützlichkeit erst im Nachhinein erbracht werden kann, dass der Zeitraum bis dahin ungewiss ist, da es sich bei Forschung um einen Vorstoß in Neuland handelt« (Weingart 2011: 45). Wissenschaft und Forschung sind nicht schon dadurch, dass sie keinen zeitnahen gesellschaftlichen Nutzen zeitigen, als fehlgeleitet oder mangelhaft erwiesen; ebenso wird die nur bedingte Kommunizierbarkeit aktueller, ergebnisoffener Wissenschaft an ein Laienpublikum in der Regel nicht einem Mangel an Interesse seitens des Wissenschaftsbetriebs zugeschrieben, sondern wird gesellschaftlich als Kehrseite des hohen Spezialisierungsgrades derselben akzeptiert. Ebenso herrscht weitgehend die Ansicht vor, »dass die Regeln, mit denen das nützliche Wissen zu produzieren ist, keiner Kontrolle und keines Eingriffs von außen zugänglich seien« (ebd.).

Die Autonomie von Wissenschaft und Forschung als einem erkenntnisgeleiteten, auf Förderung mit öffentlichen Mitteln angewiesenen Projekt, das gesellschaftlichen Nutzen verspricht, ihn jedoch nur zeitversetzt und oft über Umwege hervorzubringen im Stande ist, findet zwar weitgehende Akzeptanz, ist damit jedoch zugleich an einen permanenten gesellschaftlichen Vertrauensvorschuss gebunden. Diesen nicht zu verspielen, ist die systemische Herausforderung, der sich der Wissenschaftsbetrieb im Zusammenspiel mit anderen gesellschaftlichen Teilsystemen – etwa der Politik, aber zunehmend auch den Medien – stellen muss. Daraus ergibt sich ein latenter Rechtfertigungsdruck, der jedoch meist erst dann akut wird, wenn zum Beispiel durch wissenschaftspolitische Reformen der Forschungsförderung oder durch wissenschaftsrelevante Krisen wie die SARS-CoV-2-Pandemie

bestehende Verhältnisse – und damit auch etablierte »Arrangements von Vertrauen und Kontrolle« (ebd.: 46) – in Frage gestellt werden.

Die Gleichzeitigkeit von breiter gesellschaftlicher Akzeptanz von öffentlich geförderter wissenschaftlicher Forschung und diffusen Sorgen über den von Wissenschaft und Technik eingeschlagenen Weg spiegelt ein latentes Spannungsverhältnis wider zwischen der Aussicht auf wissenschaftlichen Fortschritt einerseits und der Art und Weise, in der sich dieser andererseits in der gesellschaftlichen Wirklichkeit niederschlägt. Während, wie erwähnt, ein Teil der Gesellschaft Bedenken gegenüber Forschung und Wissenschaft hegt, da diese das Tempo lebensweltlicher Veränderungen beschleunigten, macht sich in wissenschafts- und technikaffinen Kreisen, zumal an der Schnittstelle zwischen Wissenschafts- und Wirtschaftspolitik, Skepsis darüber breit, ob die traditionell enge Verbindung zwischen wissenschaftlichem Fortschritt und Produktivitätszuwachsen bzw. Innovationen und Wirtschaftswachstum auf Dauer fortgeschrieben werden kann. Auch hier kann die Diskussion in den USA als ein – durchaus nicht unproblematischer – Gradmesser gelten. Bestseller wie Tyler Cowens *The Great Stagnation: How America Ate All the Low-hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better* (2011) oder Essays wie Patrick Collisons und Michael Nielsens »Science Is Getting Less Bang for Its Buck« (2018) können zugleich als Ausdruck eines enttäuschten Amerikanischen Exzeptionismus und als Skepsis gegenüber der Idee eines unilinearen und irreversiblen wissenschaftlichen Fortschritts verstanden werden.

Der 2020 verstorbene Anthropologe David Graeber diagnostiziert eine Verlangsamung des wissenschaftlich-technischen Fortschritts in der zweiten Hälfte des 20. Jahrhunderts und beklagt eine »tief sitzende Ernüchterung«, die auf der »Erfahrung eines gebrochenen Versprechens« fußt – was Graeber selbstironisch an seinen eigenen enttäuschten Kindheitshoffnungen auf »fliegende Autos« festmacht. Dass »Teleportation« und »Kolonien auf dem Mars« (zwei weitere Beispiele Graebers) womöglich niemals mehr als Kindheitsfantasien und PR-Versprechungen von Technikapologeten waren, gibt Graeber zwar indirekt zu, doch hält er daran fest, dass manches von dem, was einem technikaffinen Publikum in der Jahrhundertmitte versprochen

worden war, durchaus in Reichweite gewesen wäre. Jedoch: »Es gibt noch immer keine Computer, mit denen man ein anregendes Gespräch führen kann, oder Roboter, die den Hund ausführen oder die Wäsche bügeln können.« (Graeber 2016: 129)

Als Ursache dafür, dass nur ein Bruchteil dessen, was prinzipiell in den letzten Jahrzehnten wissenschaftlich und technologisch erreichbar gewesen wäre, überhaupt systematisch weiterverfolgt wurde, macht Graeber die Assimilation »poetischer Technologien« – worunter er die »Nutzung rationaler, technischer, bürokratischer Mittel, um überschäumende, unrealistische Phantasien zum Leben zu erwecken« versteht – zu bloß »bürokratischen Technologien« verantwortlich: Administrative Zwänge und Notwendigkeiten, so Graeber, seien »von einem Mittel zum Zweck der technologischen Entwicklung geworden« (ebd.: 175). Als Beispiel führt Graeber die eigene Erfahrung einer zunehmend bürokratisierten Hochschullandschaft an, die sich in vauseilendem Gehorsam bürokratischen Managementtechniken und externen Erwartungshaltungen unterwirft. Die Universität sei von einem Ort der Exzentriz und (wenigstens potentieller) Brillanz zu einem »Ort professioneller Selbstvermarkter« (ebd.) geworden. Dass diesem Lamento etwas Nostalgisches anhaftet, mag dem essayistischen Format von Graebers Überlegungen geschuldet sein, passt jedoch zu ähnlichen Tendenzen innerhalb anderer Disziplinen.

In der Theoretischen Physik verbindet sich etwa beispielhaft eine – zum Teil auf nostalgischer Verklärung beruhende – kollektive Erinnerung an zurückliegende »Goldene Zeiten« des Erkenntnisfortschritts, insbesondere durch die theoretischen Durchbrüche der Relativitätstheorie und der Quantentheorie in den ersten Dekaden des 20. Jahrhunderts, mit einer Generalkritik am gegenwärtigen Zustand der Disziplin. Zwar, so die Theoretische Physikerin Sabine Hossenfelder, sei mit einer Verlangsamung des Erkenntnisfortschritts »auf jedem hochentwickelten Forschungsgebiet zu rechnen« (Hossenfelder 2018: 121). Doch gerade die Hochenergiephysik scheue sich in besonderem Maße, sich aktiv um die Verfolgung empirisch überprüfbarer Forschungsansätze zu kümmern. Dass sich insbesondere der theoretische Erkenntnisfortschritt verlangsamt habe, erfolge »weitgehend

aus denselben Gründen, aus denen neue Experimente nur schwer zu entwickeln sind: Die leichten Lösungen wurden bereits ausprobiert« (ebd.).

Jeder Versuch, den wissenschaftlich-technischen Fortschritt auf globaler Ebene zu kartieren, ist angesichts der Auffächerung der Wissenschaft in Einzeldisziplinen, der Zunahme interdisziplinärer Forschungsprojekte und der Schwierigkeit, eine Demarkation von Wissenschaft und Nicht- bzw. Pseudowissenschaft vorzunehmen, notwendig zum Scheitern verurteilt. Die wissenschaftliche Publikationstätigkeit ist womöglich der einzige quantifizierbare Anhaltspunkt für die Ausweitung wissenschaftlichen Wissens, kann jedoch aus offensichtlichen Gründen kaum mit dem qualitativen Kriterium wissenschaftlichen Erkenntnisfortschritts gleichgesetzt werden. Legt man dennoch diesen Maßstab an, so erweist sich das bereits Mitte des 20. Jahrhunderts von Derek de Solla Price festgestellte exponentielle Wachstum wissenschaftlicher Veröffentlichungen als bis in jüngste Zeit erstaunlich robust.³ Dass sich trotz vielfach beklagter Hyperspezialisierung und kontrovers diskutierten Problemen wie der *Replikationskrise* diese Kontinuität in der Wissenschaftsentwicklung fortschreibt, ist einigen Kommentatoren Anlass genug, von einem (scheinbaren) Paradoxon zu sprechen: »It appears paradoxical that science is producing outstanding new results and theories at a rapid rate at the same time that researchers are identifying serious problems in the practice of science that cause many reports to be irreproducible and invalid«, schreiben etwa Shiffrin et al. (2018: 2632).

Sichtbar werden die erzielten Fortschritte für den Laien – aber auch für Wissenschaftler selbst, zumindest wenn es sich um Entwicklungen außerhalb des eigenen Spezialgebiets handelt – oft erst durch technische Neuerungen, die Eingang in den Alltag finden: »computers, cell-phones, satellites, GPS, information availability, antibiotics, heart operations, increasing age of mortality, fertilizers, genetically improved crops, minimally invasive joint surgeries and joint replacements, contact lenses and laser surgery, magnetic resonance imaging, robotics,

3 Vgl. hierzu (Price 1963).

electric vehicles, air bags, solar energy, wind energy, better weather forecasting, and on and on« (Shiffrin et al. 2018: 2633). Für den einzelnen ist es nahezu unmöglich, daraus ein valides Gesamtbild wissenschaftlichen Fortschritts zu rekonstruieren, zumal gängige Interpretationsmuster oft anhand einer überdurchschnittlichen Betonung des Konsumaspektes eingeübt worden sind. So verzerrt der Fokus auf IT-Hardware und *consumer electronics* vielfach die gesellschaftliche Diskussion: Mögen die (oft nur inkrementellen) Hardware-Fortschritte – noch dünnere Handys, noch hellere Displays – auch ein sichtbares Zeichen technischen Fortschritt sein, so rücken durch deren Betonung die ungleich vielschichtigeren, dabei jedoch für den User typischerweise unsichtbaren, Fortschritte durch Vernetzung und Virtualisierung der Daten in den Hintergrund, selbst wenn es diese sind, die die Möglichkeit neuartiger technischer Anwendungen eröffnen und erfahrbar machen: Musik wird aus der Cloud gestreamt, soziale Netzwerke sondieren algorithmisch die Vorlieben ihrer Nutzer, Entscheidungsprozesse werden automatisiert.

Allenfalls dann, wenn Datensicherheit, Gefahren für die Privatsphäre und die Gefahr intransparenter Monopolbildungen durch die Konzentration der Internethoheit in den Händen weniger Digitalkonzerne wie Facebook und Google, diskutiert werden, werden Fragen der Datafizierung menschlicher Lebenswelten und die Kombination von *machine learning* und Künstlicher Intelligenz in der Öffentlichkeit diskutiert. Die Folgen von Datafizierung und Hinwendung zu datenintensiven Forschungsmethoden auf die Wissenschaft selbst werden dagegen nur selten thematisiert.

2. »Big Data« und wissenschaftliche Theoriebildung

Die tiefgreifenden Veränderungen, die dem Wissenschaftsbetrieb durch die zunehmende Verbreitung datengetriebener Methoden bevorstehen, werden in der breiteren Öffentlichkeit bestenfalls punktuell – und auch dann oft nur mit großem Zeitverzug – diskutiert. Zwar dürfte einer Mehrheit der (wissenschaftsinteressierten) Öffentlichkeit

klar sein, dass moderne Forschungsmethoden – von der Sequenzierung des Erbguts von Organismen bis hin zur Datensammlung durch die Großgeräte der Teilchenphysik – auf enorme Datenmengen angewiesen sind. Und auch Grundzüge neuartiger Forschungsansätze, wie etwa des maschinellen Lernens und der damit verbundenen Notwendigkeit, Algorithmen anhand großer Datenmengen anzulernen, sind mittlerweile mit einiger Regelmäßigkeit Gegenstand wissenschaftsjournalistischer Darstellungen. Die SARS-CoV-2-Pandemie ab dem Jahr 2020 und die damit einhergehenden öffentlichen Diskussionen zur Modellierung der Pandemieentwicklung, der Sequenzierung von Virusgenomen und allgemein zur Wissenschaftsentwicklung unter den Bedingungen von ›Fast Science‹ haben dazu in nicht geringem Maße beigetragen.

Inwieweit die zunehmende Verbreitung von Big Data eine Rückwirkung auf die öffentliche Wissenschaftswahrnehmung hat, ist bisher kaum untersucht worden: Empirische Untersuchungen, die etwa die öffentliche Meinung zur Rolle von Big Data in den Wissenschaften zum Gegenstand hätten, existieren praktisch nicht – und wären womöglich seitens einer breiten Öffentlichkeit aufgrund mangelnden Hintergrundwissens um intrawissenschaftliche Prozesse nur wenig erfolgversprechend. Dabei böten die spürbaren lebensweltlichen Veränderungen durch Digitalisierung und Big Data durchaus vielversprechendes Rohmaterial für eine gesellschaftliche Diskussion über die Rolle und die möglichen zukünftigen Entwicklungen von Wissenschaft und Technik selbst. Wie Sabina Leonelli anmerkt: »The blatantly transformative role of digitalization for all parts of society, and increasing public awareness of its shortcomings, provides a fertile terrain for dialogue.« (Leonelli 2021)

Innerhalb einer technikaffinen Teilöffentlichkeit ist der Einfluss von Big Data auf die Wissenschaftspraxis durchaus früh diskutiert worden. Der Leserschaft des Magazins WIRED, die vom Verlag als »globally-minded thought leaders, innovators, bloggers, and connectors [who are] constantly seeking new ideas« (WIRED Media-Kit 2017) beschrieben wurde, wurde bereits 2008 in einer Überschrift suggeriert, Big Data zöge »The End of Theory« nach sich. Dem Autor, Chris Anderson,

zufolge könne sich die Wissenschaft angesichts der sich im Petabyte-Bereich bewegenden Datenmengen fortan auf reine Datenanalyse zurückziehen, ohne dabei auf theoretische Modelle oder Hypothesen zurückgreifen zu müssen. Es reiche, so Anderson, Daten »in die größten Computercluster zu werfen, die die Welt je gesehen hat, und statistische Algorithmen-Muster finden zu lassen, die die Wissenschaft sonst« gar nicht auffinden könnte (Anderson 2008; Übers. d. Verf.). Die Steigerung verfügbarer Datenmengen eröffne gewissermaßen die Perspektive eines »Turbo-Empirismus«, der sich potentiell nicht nur auf die Totalität *aller* gemessenen Daten stützen könne, sondern noch dazu in der Lage sei, die Kategorien und Kriterien, nach denen Daten zu sichten und Korrelationen zu beurteilen sind, nahtlos und *bottom-up* zu generieren. Ließe sich darüber hinaus die Datensammlung automatisieren, wäre damit das Feld bereitet für einen Empirismus, der sich jeglicher anthropozentrischen Fixierung auf menschliche Vorannahmen und Vorurteile entledigt hätte.

Ein Einwand, der Hoffnungen entgegensteht, die Daten *für sich* sprechen zu lassen, ergibt sich daraus, dass hinreichend große Datensätze immer zu Scheinkorrelationen Anlass geben, deren Zahl zu allem Überfluss mit dem Umfang der Datensätze noch zunimmt (vgl. Calude & Longo 2016). Wenn aber überspitzt formuliert die meisten Korrelationen Scheinkorrelationen sind, relativiert sich die Hoffnung auf eine modell- und theoriefreie Wissenschaft, deren Aufgabe es ja gerade sein sollte, zwischen Scheinkorrelationen und solchen, die kausal oder strukturell begründet sind, zu unterscheiden. Vertreter eines wissenschaftlichen Realismus können zudem darauf verweisen, dass Datenmodelle – und seien sie noch so gut in der Lage, vorliegende Daten zu beschreiben – kategorial von Theorien zu unterscheiden sind, insoweit letztere gerade in der Lage sein müssen, zwischen bloßen Korrelationen und echten Kausalzusammenhängen zu differenzieren. Ein propagiertes »Ende der Theorie« wird nur diejenigen zufriedenstellen können, die bereit sind, die Hoffnung auf wahre Erkenntnis fundamentaler Zusammenhänge zugunsten des Kriteriums eines – fraglos nützlichen, aber eben bestenfalls als Anhaltspunkt dienenden – instrumentellen Erfolgs aufzugeben.

So verlockend die Aussicht auf einen ›turbo-empiristischen‹ Automatismus der Welterkenntnis auch sein mag, so fragwürdig ist die Annahme, die algorithmische Mustererkennung durch maschinelles Lernen könne ohne weitere Voraussetzungen eine menschlichen Erkenntnisinteressen gemäße Welterkenntnis ermöglichen. Zu Recht wird oft von der »Opazität« der algorithmischen Identifikation von Datenstrukturen gesprochen, d.h. von der Unmöglichkeit, aus dem instrumentellen Erfolg bei der Vorhersage oder Reproduktion von Daten theoretische Einsicht zu generieren. Die analytische Nichtnachvollziehbarkeit der Genese von *machine-learning*-basierten Datenmodellen lässt sich zwar formal der Beschränktheit unseres eigenen kognitiven Apparats anlasten; jedoch spiegelt sie zugleich die pragmatische Beschränktheit der entsprechenden Analysemethoden wider. Mögen die algorithmisch generierten Datenmodelle auch hinsichtlich ihrer empirischen Adäquatheit traditionellen Methoden objektiv überlegen sein, so müssen sie doch – darauf hat Mieke Boon hingewiesen – weiterhin an pragmatischen Kriterien wie Konsistenz, Kohärenz, Einfachheit, explanatorische Kraft, Reichweite, Relevanz und Intelligibilität gemessen werden, ohne deren Erfüllung sie weder handhabbar noch in vielen Kontexten nutzbar wären.⁴

Dem Big-Data-Ansatz gegenübergestellt wird bisweilen das Projekt einer datengetriebenen Forschung, die auf eine »hybride Kombination abduktiver, induktiver und deduktiver Zugänge mit dem Ziel, ein Phänomen zu verstehen« setzt (Kitchin 2014: 5). Das Ziel ist dabei weder ein bloßes Fortschreiben beobachteter Muster noch ein Subsumieren unter *ex ante* vorausgesetzte Gesetzmäßigkeiten oder theoretische Hypothesen. Vielmehr wird unter Berücksichtigung von relevantem Hintergrundwissen und Erwartungen darüber, welche Ansätze besonders vielversprechend sein könnten, eine Auswahl möglicher Erklärungen und Systematisierungen auf ihre Fruchtbarkeit hin verglichen – zwar

4 Vgl. hierzu Mieke Boon: »The crux of pragmatic criteria such as consistency, coherency, simplicity, explanatory power, scope, relevance and intelligibility in generating and accepting scientific knowledge is to render scientific knowledge manageable for humans in performing epistemic tasks.« (Boon 2020: 59)

ergebnisoffen, aber ohne den Anspruch, *alle* Daten auf *jedwede* Weise in die Auswertung miteinzubeziehen. Eine solche *data-driven science* wäre wissenschaftlicher Theoriebildung nicht antagonistisch gegenübergestellt, sondern könnte als datengetriebenes Äquivalent dessen gelten, was in den letzten Jahren als »explorative Wissenschaft« bzw. »exploratives Modellieren« diskutiert worden ist: d.h. ein systematisches Ausloten möglicher Zusammenhänge in einem Phänomenbereich, für den entweder noch keine Theorie im engeren Sinne zur Verfügung steht oder für den generell keine Aussicht auf eine geschlossene theoretische Beschreibung besteht.⁵

3. Wissenschaftsbetrieb im Zeitalter von »Big Data«

Mögen Berichte von einem angeblichen »Ende wissenschaftlicher Theorien« aus den im vorigen Abschnitt diskutierten Gründen auch verfrüht sein, so bedeutet dies nicht, dass datengetriebene Forschung ihrerseits nur eine Art Hilfsfunktion wahrnimmt. Vielmehr eröffnen Big Data und vergleichbare Forschungsansätze der Wissenschaft neue Perspektiven, selbst wenn auch hier viele der laufenden Entwicklungen noch nicht den Weg in die breitere Wissenschaftswahrnehmung durch die Öffentlichkeit gefunden haben. Die öffentliche Wahrnehmung von Big Data bleibt in erster Linie verbunden mit Fragen der Datensicherheit und der Transparenz in staatlichen und gewerblichen Kontexten, die sich daraus ergeben, dass Konsumenten und Nutzer im Internet zahlreiche Datenspuren hinterlassen, deren Zusammenführung mit Offline-Datenbanken weitreichende Aussagen über das Verhalten der Betroffenen erlauben. Diese könnte es zum Beispiel Versicherungen und Bonitätsprüfern erlauben, einzelne Bürgerinnen und Bürger von wichtigen Leistungen auszuschließen und damit de facto zu diskriminieren.

Am ehesten überlappen sich öffentliche Wahrnehmung von Big Data und wissenschaftlicher Ist-Zustand im Bereich der biomedizinischen Forschung. Im Zuge des Trends zur Selbstoptimierung sind

5 Vgl. hierzu (Gelfert 2016: insb. Kap. 4) sowie (Fisher, Gelfert & Steinle 2021).

Smart Watches und sog. *wearables* wie FitBit-Bänder beliebt; zudem zeichnen Apps Blutdruck- und Cholesterinwerte auf oder überwachen den Schlafrhythmus. Auf wissenschaftlicher Ebene laufen derweil Bemühungen, die anfallenden Mengen an Vitaldaten aufzubereiten und nach bislang unerkannten Signaturen möglicher Erkrankungen zu durchforsten. Auch hier divergieren die gesellschaftlichen und wissenschaftlichen Sorgen: Steht auf gesellschaftlicher Ebene die Befürchtung, Daten könnten unbefugt genutzt oder missbraucht werden, im Mittelpunkt, so stellt sich in der medizinischen Diagnostik die Frage, inwieweit den automatisierten Auswertungen durch Computer vertraut werden kann. Solange Behandlungen sich nicht auf überprüfte Hypothesen über Ursache-Wirkung-Beziehungen stützen können, scheint manchem Kritiker das Sich-Verlassen auf Big-Data-Diagnostik als eine Form blinden Vertrauens in nicht länger überprüfbare Algorithmen.

Unterdessen haben sich im Hintergrund längst Infrastrukturprojekte zur Vernetzung bestehender Datenbanken etabliert. Das »All of Us«-Programm der US-amerikanischen National Institutes of Health soll bestehende Gendatenbanken zusammenführen und die Genome von mehr als einer Million Bürgerinnen und Bürger für die Erforschung von Volkskrankheiten aufbereiten, in der Hoffnung, damit in Zukunft individualisierte Präzisionsmedizin zu ermöglichen. Ähnliches ist auf EU-Ebene mit dem »1+ Million Genomes«-Projekt geplant, zu dem auch die deutsche Genominitiative »genomDE« beitragen soll. Die noch immer populäre Vorstellung, für viele Krankheiten gäbe es »ein Gen«, ist dabei schon lange nicht mehr leitend; vielmehr hat sich die Erkenntnis durchgesetzt, dass sich hinter klinischen Diagnosen von Krankheitsbildern eine Vielzahl von Faktoren und Wirkmechanismen verbergen kann, deren Identifikation auch mit einer Revision des Krankheitsbildes selbst verbunden sein kann.

Nicht nur im Bereich der Medizin, sondern auch in Wissenschaftszweigen, die weniger öffentliche Aufmerksamkeit erfahren, wie z. B. der Pflanzenforschung, haben datenintensive Methoden große Verbreitung erlangt und traditionelle Vorgehensweisen revolutioniert. Aufgrund der Verbesserung von Techniken zur Gensequenzierung und des kommer-

ziellen Preisverfalls bei den für die Kartierung genetischer Marker nötigen Ingredienzien ist mittlerweile die Erfassung phänotypischer Merkmale – zumal, wenn es sich um komplexe, multi-faktorielle Merkmale wie den Ertrag oder die Schädlingsresistenz von Nutzpflanzen handelt – zum limitierenden Faktor bei der gentechnischen Züchtung neuer Pflanzensorten geworden. Um dem zu begegnen, werden zunehmend Werkzeuge zur Hochdurchsatz-Phänotypisierung eingesetzt, die die relevanten Merkmale quantitativ über eine Reihe von Anbaustandorten hinweg messen – zum Teil mittels selbstlernender Algorithmen und Verfahren der Fernerkundung. Wie dieses Beispiel zeigt, bewegen sich mittlerweile auch Felder wie die Pflanzenforschung, die in der Öffentlichkeit als eher traditionell wahrgenommen werden, bisweilen an vorderster Front bei der Integration datenintensiver Methoden, wie sie durch die Konvergenz von Gensequenzierung, High-Throughput-Phänotypisierung und GIS-Daten gegeben ist.

Die scheinbar unbegrenzte Verfügbarkeit von Daten ist einerseits die Quelle aller mit Big Data verbundenen Hoffnungen, wirft andererseits grundlegende Fragen auf, wie sich Daten verschiedenen Ursprungs auf valide Weise integrieren lassen. Wie Leonelli argumentiert, kommt es dabei vor allem auf die soziale Organisation, die beteiligten Institutionen und die zu Grunde liegenden Technologien an; nur im Zusammenspiel derselben ist die Integration, Kommunikation und Anwendung von Daten verschiedenen Ursprungs für die diversen verfolgten (epistemischen und nicht-epistemischen) Zwecke realistisch.⁶ Daten müssen »kuratiert« werden, was unter anderem hohe Ansprüche an den Mark-up mittels Metadaten stellt: »Where data and metadata professionals exist, they often operate as intermediaries, performing the articulation and liaison work necessary to represent complex and dynamic scientific research within standardized metadata products.« (Mayernik 2019: 736)

Überdies muss kritisch darauf hingewiesen werden, dass die in vielen Disziplinen aktiv propagierte Hinwendung zu datenintensiven Methoden und Ansätzen nicht ohne Verluste vonstatten geht. So werden

6 Vgl. hierzu (Leonelli 2016).

qualitative Herangehensweisen in der Konkurrenz mit datengetriebenen bzw. Big-Data-Ansätzen oft ausgebremst. Dabei geht es nicht nur um eine mögliche Bedrohung der methodologischen Autonomie ganzer Wissenschaftszweige, wie zum Beispiel der Geistes- und Kulturwissenschaften (etwa mit Blick auf das Feld der *Digital Humanities*), sondern vor allem auch um interdisziplinäre Felder wie die Gesundheitswissenschaften, bei denen die Gefahr besteht, dass beispielsweise ethnographische Ansätze oder auch Einzelfallstudien unter immer stärkeren Legitimationsdruck geraten. Selbst in den traditionell technikaffinen Ingenieurwissenschaften treibt einige Gemüter die Sorge um, die Dominanz von KI-Ansätzen würde zu einem signifikanten Verlust führen – und zwar an ›Freude an der Technik‹, da der bloße Verweis auf den empirischen Erfolg der angewandten KI-Methoden die ›Erzähl- und Erklärbarkeit‹ von technischen Erfolgen nicht angemessen ersetzen könne: Wie, so die rhetorische Frage des Ingenieurs Michael Kuhn in einem Aufsatz der Zeitschrift *Chemie-Ingenieur-Technik*, »will man junge Leute dazu motivieren, sich mit Wissenschaft und Technik auseinanderzusetzen, wenn das aktive Verstehen und Gestalten in immer tiefere Algorithmentschichten abgeschoben wird?« (Kuhn 2021: 368)

Diejenigen, die sich – ganz im Sinne der Rede von einem ›Ende der Theorie‹ – von den neuen Big-Data-getriebenen Ansätzen eine Befreiung vom Korsett strenger Theoriebildung und kausaler Modelle und damit eine umfassende Umwälzung der Wissenschaft erhoffen, bedienen einerseits ein weit verbreitetes Narrativ, das Wissenschaft als von der Gesellschaft zu förderndes Instrument zur Lösung konkreter Probleme legitimiert. Andererseits wird dadurch die Autonomie von Wissenschaft als erkenntnisgeleiteter Unternehmung, die zum Verstehen der Welt beiträgt, erheblich relativiert; als Bewertungsmaßstab bleiben dann in erster Linie nichtepistemische Kriterien, die sich am Beitrag der Wissenschaft zur Lösung gesellschaftlich wahrgenommener Probleme orientieren. Daran, ob ein solch instrumentalistisches Wissenschaftsbild ›nach innen‹ – also wissenschaftsintern – vermittelbar, geschweige denn wünschenswert wäre, dürfen berechtigte Zweifel gehegt werden.

4. Ausblick: Zur Komplizenschaft von »Korrelationalismus« und Solutionismus

In ihrem populären Sachbuch *Big Data: The Essential Guide to Work, Life and Learning in the Age of Insight* (2017), der erweiterten Ausgabe eines 2013 zuerst erschienenen Bestsellers, schreiben Viktor Mayer-Schönberger und Kenneth Cukier:

»Before big data, our analysis was usually limited to testing a small number of hypotheses that we defined well before we even collected the data. When we let the data speak, we can make connections that we had never thought existed.« (Mayer-Schönberger & Cukier 2017: 14)

Was in diesen Zeilen als Überwindung der engen Grenzen traditioneller hypothesenbasierter Wissenschaft gefeiert wird, muss im Lichte der vorangegangenen Diskussion als durchaus zwiespältig angesehen werden. Manch eine algorithmisch identifizierte *connection* ist womöglich nur eine Scheinkorrelation, und auch wenn es naiv wäre zu glauben, unser menschliches Gespür dafür, welche Zusammenhänge real existieren und welche nicht, sei unfehlbar, so bleibt es doch ebenso naiv zu glauben, die Daten könnten – auf gewissermaßen *uninterpretierte* Weise – »für sich sprechen«. Ebenso darf mit einigem Recht bezweifelt werden, dass das Sich-Verlassen auf Big-Data-Ansätze allein schon zu einem »Age of Insight« führt, wie es der Untertitel von Mayer-Schönbergers und Cukiers Buch suggeriert. Schließlich ist gerade *insight* – also die subjektive Einsicht bzw. das Verstehen komplexer Zusammenhänge durch ein Erkenntnissubjekt – nicht das, was Big Data und maschinelles Lernen garantieren können.

Der bewusste Verzicht auf das Primat kausaler Hypothesen hat zur Folge, dass Korrelation und Aussagen zu wechselseitigen probabilistischen Abhängigkeiten zum Strukturprinzip wissenschaftlicher Erkenntnis erhoben werden. Dieser »Korrelationalismus«, wie ihn Geoffrey Bowker nennt (2014: 1796), ist zwar auf technisch mustergültige Weise durch Methoden maschinellen Lernens auf neuronalen Netzwerken umsetzbar, erkaufte sich seinen Erfolg jedoch durch den Verzicht

auf den Anspruch, eine unabhängige theoretische Beschreibung der Welt zu entwickeln. Jedoch, wie Bowker anmerkt: »Just because we have big data does not mean that the world acts as if there are no categories.« (Bowker 2014: 1797) Auch das Generieren von Datenmodellen ist keineswegs voraussetzungslos: Zum einen ist die Existenz nackter Rohdaten eine Fiktion, die sich nur dann aufrechterhalten lässt, wenn man von der Genese der verwendeten Datenreihen absieht; zum anderen müssen Daten – insbesondere dort, wo Datenreihen aus unterschiedlichen Quellen bemüht werden – integriert und kuratiert werden.

Andersons 2008 getroffene Vorhersage, die Wissenschaft steuere dank Big Data auf ein Ende der Theoriebildung zu, hat sich in den vergangenen anderthalb Jahrzehnten nicht bewahrheitet. Zwar spielen datengetriebene Methoden in vielen Anwendungsbereichen eine wichtige unterstützende Rolle oder haben – wie in den diskutierten Fällen der Pflanzenforschung und der medizinischen Genomforschung – die Erforschung einer Reihe vorher unzugänglicher Fragen ermöglicht. Jedoch bleibt in vielen anderen Forschungszweigen das hypothesenorientierte »Prüfen, basierend auf einer Theorie- und Erkenntnisorientierung, bestehen – auch wenn diese Disziplinen mitunter die neuen informatischen Verfahren als Mittel, Methoden und Instrumente verwenden« (Schmidt 2022: 95). So argumentiert Jan Schmidt, dass die bisherige Erfolgsbilanz von Big-Data-Ansätzen nicht dazu geführt hat, Theorien und Gesetzmäßigkeiten als »Goldstandard« für die gute Prognoseleistung von Wissenschaft« vom Sockel zu stoßen; vielmehr würden sich maschinelles Lernen und Big-Data-Ansätze parallel dazu als neuer »Silberstandard« etablieren – gerade dort, wo es darum geht, »jenseits von Theorien und der Kenntnis von Gesetzen prognostische Kraft entfalten zu können« (Schmidt 2022: 91).

Ob die Koexistenz zweier Erkenntnisstandards – zumal wenn diese hierarchisch abgestuft als (traditioneller) »Goldstandard« der Theoriebildung und als (neumodischer) »Silberstandard« eines algorithmischen Turbo-Empirismus verstanden werden – auf Dauer gelingen kann, lässt sich zum jetzigen Zeitpunkt kaum sicher beantworten. Selbst wenn einige der Zuschreibungen, die von außen an die Big-Data-

Bewegung herangetragen werden, übertrieben sein mögen: Ein bloß neutrales Werkzeug im Methodenarsenal der Wissenschaften sind Big Data und maschinelles Lernen keineswegs. Vielmehr verbindet sich mit der sicheren Erwartung, die Daten ›für sich‹ sprechen lassen zu können, auch die Hoffnung, aus einer Gesamtschau aller auffindbaren Korrelationen ließen sich für jeden beliebigen Zweck unmittelbar jene Stellschrauben ablesen, die die Erreichung des definierten Ziels optimal ermöglichen. Ein weiteres Nachforschen – etwa danach, welche Tiefenstrukturen den beobachteten Phänomenen zu Grunde liegen – wäre dann nicht nur unnötig, sondern in der Kosten-Nutzen-Abwägung sogar schädlich. Damit rückt der dem Glauben an Big Data zu Grunde liegende »Daten-Fundamentalismus« (Crawford 2013) in die Nähe dessen, was der Technik-Kritiker Evgeny Morozov als technologischen »Solutionismus« bezeichnet: »An intellectual pathology that recognizes problems as problems based on just one criterion: whether they are ›solvable‹ with a nice and clean technological solution at our disposal.«⁷ (Morozov 2013)

Für die Zukunft der Wissenschaft kann dies nur bedeuten, bei aller Offenheit für die dank Big Data und maschinellem Lernen neu eröffneten Ansätze und Horizonte, die Autonomie erkenntnisgeleiteter Forschung in ihrer Methodenvielfalt auch gegenüber jenen zu verteidigen, die in datengetriebenen Methoden ein Allheilmittel sehen. Würde als Kriterium dafür, was eine legitime wissenschaftliche Fragestellung ausmacht, exklusiv die Eignung zur Anwendung datengetriebener Methoden herangezogen werden – womöglich in der Hoffnung, damit den gesellschaftlichen Nutzen von Wissenschaft zu maximieren – könnte die erhoffte Befreiung vom vermeintlichen Zwang zur Theoriebildung schnell in eine instrumentalistische Unterwerfung unter externe arbiträre Setzungen umschlagen.

7 Zur Affinität zwischen Big Data und »Techno-Solutionismus« vgl. auch (Prietl 2019).

Literatur

- AAAS (2018): *Perceptions of Science in America*, Cambridge, Mass.: American Academy of Arts & Sciences.
- Anderson, Chris (2008): »The end of theory: The data deluge makes the scientific method obsolete«, in: *Wired Magazine* 16.7. <https://www.wired.com/2008/06/pb-theory/>.
- Boon, Mieke (2020): »How scientists are brought back into science – the error of empiricism«, in: Marta Bertolaso/Fabio Sterpetti (Hg.), *A Critical Reflection on Automated Science*, Cham: Springer, S. 43-65. https://doi.org/10.1007/978-3-030-25001-0_4.
- Bowker, Geoffrey C. (2014): »The theory/data thing: commentary«, in: *International Journal of Communication* 8, S. 1795-1799. <https://ijoc.org/index.php/ijoc/article/view/2190/1156>.
- Calude, Cristian S./Longo, Giuseppe (2017): »The deluge of spurious correlations in big data«, in: *Foundations of Science* 22.3, S. 595-612. <https://doi.org/10.1007/s10699-016-9489-4>.
- Collison, Patrick/Nielsen, Michael (2018): »Science is getting less bang for its buck«, in: *The Atlantic* (16. November 2018). <https://www.theatlantic.com/science/archive/2018/11/diminishing-returns-science/575665/>.
- Cowen, Tyler (2011): *The Great Stagnation: How America Ate All the Low-hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better*, New York: Penguin.
- Crawford, Kate (2013): »The hidden biases in big data«, in: *Harvard Business Review* 1.4. <https://hbr.org/2013/04/the-hidden-biases-in-big-data>.
- Fisher, Grant/Gelfert, Axel/Steinle, Friedrich (2021): »Exploratory models and exploratory modeling in science: Introduction«, in: *Perspectives on Science* 29.4, S. 355-358. https://doi.org/10.1162/posc_e_00374.
- Gelfert, Axel (2016): *How To Do Science with Models: A Philosophical Primer*, Cham: Springer.
- Graeber, David (2016): *Bürokratie: Die Utopie der Regeln*, Stuttgart: Klett-Cotta.

- Hossenfelder, Sabine (2018): *Das hässliche Universum: Warum unsere Suche nach Schönheit die Physik in die Sackgasse führt*, Frankfurt a.M.: S. Fischer Verlag.
- Kitchin, Rob (2014): »Big Data, new epistemologies and paradigm shifts«, in: *Big Data & Society* 1.1, S. 1-12. <https://doi.org/10.1177/2053951714528481>.
- Kuhn, Michael (2021): »Big Data, AI und die Freude am Ingenieurwesen«, in: *Chemie Ingenieur Technik* 93.3, S. 364-372. <https://doi.org/10.1002/cite.202000221>.
- Leonelli, Sabina (2021): »Data science in times of pan(dem)ic«, in: *Harvard Data Science Review* 3.1. <https://doi.org/10.1162/99608f92.fbb1bdd6>.
- (2016): *Data-Centric Biology A Philosophical Study*, Chicago: Chicago University Press.
- Mayernik, Matthew S. (2019): »Metadata accounts: Achieving data and evidence in scientific research«, in: *Social Studies of Science* 49.5, S. 732-757. <https://doi.org/10.1177/0306312719863494>.
- Mayer-Schönberger, Viktor/Cukier, Kenneth (2017): *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London: Hachette.
- Morozov, Evgeny (2013): »The perils of perfection«, in: *The New York Times* (3. März 2013). <https://www.nytimes.com/2013/03/03/opinion/sunday/the-perils-of-perfection.html>.
- NSB 2018. National Science Board: *Science & Engineering Indicators*.
- Price, Derek J. de Solla (1963): *Little Science, Big Science* New York: Columbia University Press.
- Prietl, Bianca (2019): »Die Versprechen von Big Data im Spiegel feministischer Rationalitätskritik«, in: *GENDER: Zeitschrift für Geschlecht, Kultur und Gesellschaft* 11.3, S. 11-25. <https://doi.org/10.3224/gender.v11i3.02>.
- Schmidt, Jan C. (2022): »Wandel und Kontinuität von Wissenschaft durch KI. Zur aktuellen Veränderung des Wissenschafts- und Technikverständnisses«, in: In: Carl Friedrich Gethmann/Peter Buxmann/Julia Distelrath/Bernhard G. Humm/Stephan Ling-

- ner/Verena Nitsch/Jan C. Schmidt/Indra Spiecker Döhmann (Hg.), *Künstliche Intelligenz in der Forschung*. Berlin/Heidelberg: Springer, S. 79-125. https://doi.org/10.1007/978-3-662-63449-3_4.
- Shiffrin, Richard M./Börner, Katy/Stigler, Stephen M. (2018): »Scientific progress despite irreproducibility: A seeming paradox«, in: *Proceedings of the National Academy of Sciences* 115.11, S. 2632-2639. <https://doi.org/10.1073/pnas.1711786114>.
- Weingart, Peter (2011): »Die Wissenschaft der Öffentlichkeit und die Öffentlichkeit der Wissenschaft«, in: Barbara Hölscher/Justine Suchanek (Hg.), *Wissenschaft und Hochschulbildung im Kontext von Wirtschaft und Medien*, Wiesbaden: VS Verlag für Sozialwissenschaften, S. 45-61. https://doi.org/10.1007/978-3-531-92648-3_4.
- WIRED Media-Kit 2017. https://www.wired.com/wp-content/uploads/2015/03/WMG_Media_Kit_2017_v3.pdf.
- Wissenschaftsbarometer 2021. Berlin: Wissenschaft im Dialog. <https://www.wissenschaft-im-dialog.de/projekte/wissenschaftsbarometer/wissenschaftsbarometer-2021>.