



# Dimensionen von Big Data: Eine politikwissenschaftliche Systematisierung

# 3

Lena Ulbricht, Sebastian Haunss, Jeanette Hofmann, Ulrike Klinger,  
Jan-Hendrik Passoth, Christian Pentzold, Ingrid Schneider, Holger Straßheim  
und Jan-Peter Voß

## Zusammenfassung

Aus Big Data, der massenhaften Sammlung und Auswertung der vielfältigen Daten, die durch die Digitalisierung aller Lebensbereiche entstehen, erwachsen neue Phänomene, die zentrale politikwissenschaftliche Erkenntnisse und Konzepte infrage stellen und die durch moderne Gesellschaften bewertet und reguliert werden müssen. Ziel dieses Beitrags ist es, Big Data in seinen vielfältigen Bedeutungen für die politikwissenschaftliche Forschung zu erschließen und eine Systematik für künftige Forschung zu entwickeln. Fluchtpunkt ist dabei die These, dass sich durch Big Data die Bedingungen kollektiv bindenden Entscheidens verändern, indem soziale Wissensbestände, Normen und Regulierung einer radikalen Mikrofokussierung unterworfen werden. Seine Wirkung entfaltet Big Data, so die Annahme, indem es kollektiv geteilte Erwartungen weckt oder begrenzt – in kulturell-kognitiver, normativer und regulativer Hinsicht. Zugleich wird Big Data wiederum selbst durch kollektive Erwartungen geprägt. Die Tiefe und Reichweite der durch Big Data verursachten Änderungen ist allerdings je nach Dimension und Bereich ganz unterschiedlich.

## 3.1 Einleitung

Big Data bezeichnet die massenhafte Sammlung und Auswertung der vielfältigen Daten, die durch die Digitalisierung aller Lebensbereiche entstehen. Doch Big Data ist nicht nur ein technisches, sondern stets auch ein gesellschaftliches Phänomen, das sozialwissenschaftliche Erkenntnisse und Konzepte herausfordert, wie erste sozialwissenschaftliche Arbeiten aufgezeigt haben (Boyd und Crawford 2012; Barocas und Nissenbaum 2014; Lyon 2014; Pasquale 2015; Zuboff 2015; Yeung 2016).

Big Data aus politikwissenschaftlicher Perspektive zu analysieren bedeutet, ein bislang als primär technisch und einheitlich konstruiertes Phänomen vielmehr als ein ganzes Ensemble von Techniken, Strukturen, Akteuren und Praktiken zu begreifen und deren vielfältige gesellschaftliche und politische Wechselwirkungen aufzudecken. Während sich Blogs und Feuilletons der politischen Dimension von Big Data bereits ausführlich widmen, fängt eine politikwissenschaftliche Auseinandersetzung mit Big Data gerade erst an. Dabei entstehen hier neue Phänomene, die durch moderne Gesellschaften bewertet und reguliert werden müssen. Ein Beispiel etwa ist die Möglichkeit, auf der Grundlage von Big Data sehr genaue Persönlichkeitsprofile zu konstruieren. Ob diese zur Kriminalitätsbekämpfung, für die Gesundheitsprävention, in der Bildungsberatung oder zur Steigerung bürgerschaftlichen Engagements genutzt werden sollten, ist allerdings noch umstritten. Tatsächlich formulieren andere wissenschaftliche Disziplinen, die sich mit Big Data befassen, Forschungsdesiderate, die nach politikwissenschaftlichen Analysen verlangen, etwa in der Frage, wie Big Data demokratietheoretisch zu bewerten ist oder bei der Suche nach angemessenen Koordinations- und Regulierungsmechanismen. Zudem werden durch Big Data zentrale politikwissenschaftliche Erkenntnisse und Konzepte infrage gestellt. So diagnostiziert Shoshana Zuboff mit Big Data etwa die Ablösung des Markt-Kapitalismus durch einen Überwachungs-Kapitalismus mit negativen Folgen für den Wohlfahrtsstaat und individuelle Autonomie (Zuboff 2014), Karen Yeung sieht in Big Data das zentrale Element einer Regulierungsform, die Individuen stimuliert anstatt sie zu überzeugen (*nudging*) und somit potenziell manipulativ wirkt (Yeung 2016) und Barocas und Selbst zeigen anhand von Big Data auf, dass Strukturprinzipien des Rechts, wie etwa die Intentionalität bei Diskriminierung, gegenüber selbstlernenden Algorithmen zum Teil zu kurz greifen und die Reichweite staatlicher Regulierung begrenzt ist (Barocas und Selbst 2015). Eine weitere Herausforderung für politikwissenschaftliche Forschung ergibt sich daraus, dass Big Data multiple Dimensionen der Politikgestaltung betrifft. Entsprechend müssen sich die verschiedenen Strömungen und Sektionen der Disziplin mit dem Phänomen befassen, etwa die normative politische Theorie, die Policy-Analyse und die Systemlehre, und dabei Bezug aufeinander nehmen. Ziel dieses Beitrags ist es, Big Data in seinen vielfältigen Bedeutungen für die politikwissenschaftliche Forschung zu erschließen und eine Systematik für künftige Forschung zu entwickeln. *Fluchtpunkt des Beitrags ist dabei die These, dass sich durch Big Data die Bedingungen kollektiv bindenden Entscheidens verändern, indem Regulierung, Normen und soziale Wissensbestände einer radikalen Mikrofokussierung unterworfen werden. Seine Wirkung entfaltet Big Data, so die Annahme, indem es kollektiv geteilte Erwartungen weckt oder begrenzt. Zugleich wird Big Data wiederum selbst durch kollektive Erwartungen geprägt, wie im Folgenden deutlich gemacht wird.*

Big Data manifestiert sich in vielfältigen Ebenen der Politikentstehung; somit erscheint es für eine analytische Annäherung ratsam, einen theoretischen Rahmen zu wählen, der offen genug ist, um die verschiedenen Aspekte von Big Data in ihrer Bandbreite aufzuzeigen und zugleich zu systematisieren. Wir schlagen einen Ansatz vor, der das zirkuläre Verhältnis zwischen Technik und Gesellschaft in Dimensionen unterteilt, wie es etwa in Anlehnung an die Institutionentheorie möglich ist. Auch die institutionentheoretisch angelegte Bedeutung von Erwartungen als Strukturgeber menschlichen Denkens und Handelns erweist sich mit Blick auf Big Data als fruchtbar. So sehen Streeck und Thelen stabilisierte Erwartungen als das konstitutive Element von Institutionen – Erwartungen, die kollektiv unterstützt und durchgesetzt werden und das Handeln von Akteuren und die Ausführung von Aktivitäten prägen (2005, S. 9)<sup>46</sup>. Wenn die Erwartungen daran, was „angebrachte“ und „unangebrachte“, „richtige“ und „falsche“, „mögliche“ und „unmögliche“ Handlungen sind, sich nicht von Person zu Person unterscheiden, sondern geteilt werden, entsteht Raum für das Soziale. Kollektive Erwartungen schreiben Akteuren Rechte und Verpflichtungen zu und machen auf diese Weise Verhalten vorhersehbar und verlässlich. Analog dazu gehen wir somit davon aus, dass Big Data Erwartungen mit Blick auf Akteure und Aktivitäten produziert.

Für die erste Systematisierung unterscheiden wir zudem, angelehnt an eine sozialkonstruktivistische Institutionentheorie (Berger und Luckmann 1966, Giddens 1984), zwischen drei Dimensionen von Big Data: einer kulturell-kognitiven, einer normativen und einer regulativen (Scott 2008). Die *kulturell-kognitive Dimension* von Institutionen bezieht sich auf geteilte Weltansichten und Interpretationsrahmen, in denen Wissen erzeugt wird und Bedeutungen zugeschrieben werden. Institutionen bestimmen hier, welche Handlung als selbstverständlich und sinnvoll beziehungsweise sinnlos gilt und welcher Akteur als kompetent oder ahnungslos angesehen wird. Die *normative Dimension* von Institutionen umfasst geteilte Werte und Normen, aus denen Ziele und Standards abgeleitet werden, sowie Wege, diese Ziele zu erreichen. Auf diese Weise werden (un)erwünschte und (un)angemessene Handlungen definiert und Akteure als (un)moralisch oder (un)ehrenhaft charakterisiert. Die *regulative Dimension* von Institutionen findet man schließlich da, wo Regeln gesetzt werden und ihre Einhaltung überwacht und durchgesetzt wird. In regulativer Hinsicht entscheiden Institutionen darüber, welche Handlungen regelkonform sind oder Regelverstöße darstellen und welche Akteure als (un)schuldig beziehungsweise (un)bestechlich gelten.

---

46 „... institutions [...] represent socially sanctioned, that is, collectively enforced expectations with respect to the behavior of specific categories of actors or to the performance of certain activities” (Streeck und Thelen 2005, S. 9).

Ohne Big Data notwendigerweise zur Institution zu erklären, nähert dieser Beitrag sich Big Data als einem kulturell-kognitiven, normativen und regulativen Rahmen, der Akteuren Rollenerwartungen vorgibt und deren Handlungsspektrum bestimmt.<sup>47</sup> Darüber hinaus übt Big Data nicht nur Einfluss aus, sondern wird selbst durch Erwartungen beeinflusst und strukturiert – was sich ebenfalls in kognitiver, normativer und regulativer Hinsicht darstellen lässt. Aus diesem zirkulären Verhältnis zwischen Technik und Gesellschaft ergeben sich sechs analytische Zugänge zu Big Data (siehe Tabelle 2). Dabei ist zu beachten, dass die drei Dimensionen ineinander übergehen und sich sowohl gegenseitig verstärken als auch miteinander konfliktieren können (Scott 2008, S. 62). Künftige Studien können Schwerpunkte auf einzelne dieser Dimensionen von Big Data legen, den jeweiligen theoretischen Zugang stärker pointieren und empirisch unterfüttern.

**Tab. 2** Dimensionen für eine politikwissenschaftliche Analyse von Big Data

	Kulturell-kognitive Dimension	Normative Dimension	Regulative Dimension
Wirkungen von Big Data: wie Big Data Erwartungen produziert	1 <i>kulturell-kognitiv hergestellte Erwartungen durch Big Data (z. B. Zukunftserwartungen, neue Formen der Expertise)</i>	2 <i>normativ hergestellte Erwartungen durch Big Data (z. B. Klassifizierungen, Verhaltenssteuerung)</i>	3 <i>regulativ hergestellte Erwartungen durch Big Data (z. B. teil-automatisierte Regulierung)</i>
Einflüsse auf Big Data: Erwartungen, die Big Data prägen	4 <i>kulturell-kognitiv hergestellte Erwartungen an Big Data (z. B. menschliche versus maschinelle Agenten, Epistemologie)</i>	5 <i>normativ hergestellte Erwartungen an Big Data (z. B. Anforderungen an Fairness und Legitimität)</i>	6 <i>regulativ hergestellte Erwartungen an Big Data (z. B. Regulierung durch Datenschutz)</i>

Quelle: Eigene Darstellung ©

47 Die Bedeutung von Rollen und Handlungsrahmen geht auch aus der oben benannten Definition von Streeck und Thelen hervor: “expectations with respect to the behaviour of specific categories of actors or to the performance of certain activities” (2005, S. 9) (Hervorhebung durch die Autorinnen und Autoren).

Die folgenden Unterkapitel widmen sich ausführlicher den sechs vorgestellten Dimensionen von Big Data und werfen Licht auf die entsprechenden Akteure, Diskurse, Praktiken, Akteursinteressen, Machtkonstellationen usw. Leitend ist die Frage, wie sich die Wirkung von Big Data durch kollektive Erwartungen theoretisch und empirisch analysieren lässt und ob Big Data, wie angenommen, die Bedingungen kollektiv bindenden Entscheidens verändert.

Die Unterkapitel widmen sich der Frage, in welcher Form und welchem Ausmaß Big Data zu Veränderungen führt und politikwissenschaftliche Gewissheiten und Konzepte herausfordert, ergänzt oder bestätigt. Dies umfasst auch den jeweiligen Stand der Forschung, der mit Blick auf genuin politikwissenschaftliche Analysen zumeist noch sehr überschaubar ist und entsprechend mit relevanten Erkenntnissen aus anderen Disziplinen, wie etwa der Soziologie, den Kommunikations- und Rechtswissenschaften, ergänzt wird. Diese Analysen zeigen auf, wo politikwissenschaftliche Forschung ansetzen kann, um Antworten auf die Fragen zu finden, die sich für jede Dimension in der Auseinandersetzung mit Big Data ergeben. In der Zusammenschau entsteht so eine politikwissenschaftliche Forschungslinie, die unterschiedliche empirische und theoretische Vorhaben zu Big Data entlang dieser Linie verorten und miteinander in Verbindung bringen will.

---

### **3.2 Big Data als epistemische Innovation? Kulturell-kognitiv hergestellte Erwartungen durch Big Data**

Jan-Peter Voß

Sich der kulturell-kognitiven Dimension von Big Data zu widmen, legt den Fokus auf geteilte Weltansichten und Interpretationsordnungen, die durch Big Data erzeugt werden. Sie geben den Rahmen vor, in dem (menschliche und technische) Akteure Wissen erzeugen und Bedeutungen zuschreiben. Sie dienen auch der Bewertung von Akteuren als kompetent oder inkompetent und von Handlungen als sinnvoll oder sinnlos. Dieser Beitrag entwickelt die These, dass Big Data neben seinen viel-diskutierten gesellschaftlichen und politischen Wirkungen auch eine epistemische Innovation darstellt. Soziologisch gesehen kann man hier eine neue Form beobachten, in der die Gesellschaft Wissen über sich selbst generiert, und damit eine neue Art, das gesellschaftliche Selbstbewusstsein (mit) zu gestalten. Zu den klassischen Formen wie öffentliche Debatten, politische Problematisierungen, künstlerische Aufarbeitung und, besonders in modernen Gesellschaften, auch Statistik und klassische Sozialwissenschaft auf der Basis von Demographie, Umfrageforschung,

Interviews, Textanalysen und teilnehmender Beobachtung, scheint mit Big Data nun eine neue Form hinzuzutreten.

Bisher ist offen, wie die Innovation sich entwickelt und welche Folgen damit verbunden sein werden. Eine zentrale Frage ist, was sich dadurch inhaltlich verändert im Wissen einer Gesellschaft von sich selbst (Russell Neuman et al. 2014). Eine andere Frage ist, wie das Verhältnis verschiedener Formen der gesellschaftlichen Selbstreflexion beeinflusst wird und wie Big Data den Einfluss verschiedener Akteure in der Artikulation des Verständnisses von Gesellschaft von sich selbst verändert (Savage und Burrows 2007). Damit verbunden ist die Frage, wie Big Data, vermittelt über diese Verschiebungen im Wissen der Gesellschaft von sich selbst, auch die zukünftige Entwicklung von Gesellschaft beeinflusst (Law und Urry 2004; Law et al. 2011). Gegenwärtige kontroverse Debatten um Big Data können vor dem Hintergrund dieser Fragen als ein Prozess der „informellen Technikfolgenabschätzung“ für ein neues Instrument der gesellschaftlichen Selbstbeobachtung verstanden werden (Rip 1987).

Um auf bisher wenig beachtete Zusammenhänge hinzuweisen, bürsten wir die gegenwärtige Debatte um die epistemischen Wirkungen in zwei Punkten gegen den Strich. Wir problematisieren nicht die instrumentelle Macht, die Big-Data-Analysen einzelnen Akteuren an die Hand geben, sondern fragen nach kulturellen Wirkungen, die unabhängig davon existieren, ob die Analyse tatsächlich Wirklichkeit abbilden und Steuerung erlauben kann. Zuerst stellen wir die Frage, inwieweit Big Data Ordnungsmuster in der gesellschaftlichen Wirklichkeit nicht nur beschreibt, sondern aktiv dazu beiträgt, sie herzustellen. Statt der epistemisch-instrumentellen Macht stellen wir also die ontologisch-performative Macht von Big Data heraus, die daran hängt, dass mit Big-Data-Methoden bestimmte Vorstellungen des Wesens sozialer Ordnung (Ontologien) verbunden sind, die übernommen und praktisch umgesetzt werden, wenn diese Methoden und damit produzierte Ergebnisse verwendet werden, so dass diese Vorstellungen damit tatsächlich als soziale Ordnung real werden (Performativität). Hier geht es also darum, dass soziale Ordnungen, die sich nicht unmittelbar in der gesellschaftlichen Wirklichkeit finden lassen, erst durch Annahmen und Modelle erzeugt werden, die der Erhebung und Auswertung von Daten zu ihrer Beschreibung zugrunde liegen. Zweitens diskutieren wir auf dieser Basis, wie Big-Data-Analysen im Zusammenhang mit politischen Repräsentationsansprüchen dazu beitragen können, dass sich die „allgemeine Ökonomie der Repräsentation“, in der die kulturellen Ressourcen politischer Autorität hergestellt werden, ändert (Rosanvallon 2002; Disch 2008 65-67). Wenn z. B. Analysen von Twitterrends herangezogen werden, um zu behaupten, kollektive Befindlichkeiten und Interessen zu repräsentieren und daher im Namen einer größeren Gruppe oder „der Mehrheit der Bevölkerung“ zu sprechen, dann trägt Big Data zur Generierung von politischer

Macht, Legitimation und Mobilisierungsfähigkeit bei. Beide Punkte diskutieren wir im Folgenden als die epistemische und die politische Performativität von Big Data. Dabei geht es um die Umsetzung von Annahmen über die objektive Realität des Sozialen (epistemische Performativität) und über kollektive Subjektivitäten (politische Performativität) (zu epistemischer und politischer Performativität siehe Voß 2016a, 2016b). Wir illustrieren sie jeweils anhand von Beispielen.

### 3.2.1 Epistemische Performativität: „Enacting Big Data Realities“

Die epistemische Wirkung von Big-Data-Analysen leitet sich nicht direkt aus ihrem Bezug zur gesellschaftlichen Wirklichkeit ab, sondern primär daraus, dass sie zu Wissen werden: entweder explizit, indem Ergebnisse zur Handlungsorientierung dienen, oder indirekt, indem die kategoriellen Annahmen und ontologischen Modelle übernommen werden, die ihnen zugrunde liegen. Sofern Big-Data-Analysen im praktischen Umgang mit Selbst und Welt als Beschreibungen übernommen werden, prägen sie Selbstwahrnehmung und Handlungsmuster. Dann werden sie zu gesellschaftlicher Realität, unabhängig davon, wie akkurat sie vorab existierende Wirklichkeiten abbilden. Ein Beispiel ist die Repräsentation von Gesellschaft als *crowd*, als ein Aggregat von individuellen Usern, deren *agency* im Wesentlichen darin besteht, durch Software präkonfigurierte *klicks*, *likes*, *tweets* usw. abzugeben. Derartige Repräsentationen gesellschaftlichen Lebens können also *ex post* wahr werden, auch wenn sie es nicht unbedingt *ex ante* waren, bevor sie praktisch handlungsleitend wurden (Merton 1948; Austin 1975 [1962]; Foucault 1980; Butler 1988; Keller et al. 2005). Diese „performative“ Wirkung wurde bereits für andere Methoden der Sozialforschung untersucht (Osborne und Rose 1999; Callon 2007; Law 2009). Big Data bringt hier einerseits lediglich eine weitere Methode ins Spiel, was hinsichtlich der Diversifizierung wissenschaftlich angebotener Realitätsordnungen zu begrüßen ist. Andererseits treten die Befürwortenden und Anbietenden von Big-Data-Analysen mit sehr weitreichenden Behauptungen zur epistemischen Qualität und den damit verbundenen Steuerungsoptionen auf. Dabei stechen unabhängig von der Frage, wie Big Data die Diversität von Ansätzen der Sozialforschung beeinflussen wird, einige Eigenheiten der Wissensproduktion mit Big Data ins Auge, die zu einer kritischen Auseinandersetzung geführt haben.

#### Selektivität der Rohdaten

Die vermeintliche Grundgesamtheit gesellschaftlicher Interaktion, die in Big-Data-Analysen betrachtet wird, umfasst tatsächlich nur digital vermittelte Interak-

tionen und davon nur den Anteil und die Aspekte, die registriert und gespeichert werden. Was nicht durch digitale Medien und Geräte erfasst wird, bleibt außen vor. Dabei werden systematisch ganze Menschengruppen nicht berücksichtigt (Boyd und Crawford 2012, S. 669; Gillespie 2014, S. 172). Twitter-Feeds sind nicht einfach mit „the public mood“ gleichzusetzen, wie manche Analysen verkürzend behaupten (Bollen et al. 2011). Bei Analysen der Daten, die über ein bestimmtes Medium wie Twitter oder Facebook produziert werden, werden Interaktionen durch das jeweilige Format der Kommunikation bestimmt, das technisch vorgegeben ist (z. B. 140 Zeichen-Nachrichten auf Twitter, algorithmisch selektierter Newsfeed auf Facebook). Darüber hinaus fallen in der automatisierten Link-, Klick- und Textanalyse viele qualitative Nuancen heraus, an denen sich der Sinn von Verhaltensäußerungen ablesen ließe (z. B. Ironie) (Boyd und Crawford 2012, S. 669). Auch die Aufbereitung von Datensätzen kann leicht zu Verzerrungen führen, seien sie intendiert oder nicht (Hardt 2014; Portmess und Tower 2015, S. 4).

### **Mangelnde Zurechenbarkeit von Handlung**

Hinzu kommt, dass nicht jeder digitale Account beziehungsweise jede Nutzerin oder jeder Nutzer direkt einem Menschen zuzuordnen ist: Ein Mensch kann verschiedene Accounts und Computer benutzen, ein Computer und Account kann von mehreren benutzt werden. In Bezug auf politische Kommunikation sind hier besonders Trolle und ganze Trollfarmen relevant, von denen aus gezielt Kommunikation in digitale Medien eingespeist wird, um den Effekt zu erzielen, dass die dort wahrgenommene und gemessene öffentliche Meinung sich verschiebt (Dahlberg 2001; Bishop 2014). Schließlich kommen zu den menschlichen Trollen vermehrt *social* und *political bots* hinzu, also Accounts, hinter denen Computerprogramme stehen, die automatisiert Klicks erzeugen, Suchbegriffe eingeben, Nachrichten versenden und weiterverbreiten usw. (Forelle et al. 2015).<sup>48</sup>

### **Spekulative Statistik**

Die Ordnungen, die auf der Basis dieser selektiven und zum Teil manipulierten Daten „festgestellt“ werden, sind statistisch konstruiert auf Grundlage von Korrelationen in Stichproben (vgl. Kitchin 2013, S. 265; Boyd und Crawford 2012, S. 668). Zum Beispiel werden entsprechend von Korrelationskoeffizienten zwischen den Facebook-Likes bestimmter User und ihren eigenen Profil-Angaben bzw. per Umfrage erhobenen Daten zu Religion, Geschlecht, sexueller und politischer Orientierung usw. Vorhersagewahrscheinlichkeiten für diese Eigenschaften ermittelt (Kosinski

---

48 Siehe auch weitere Publikationen auf: <http://politicalbots.org/?p=711>

et al. 2013). Wenn diese dann ab einer bestimmten Trefferwahrscheinlichkeit als erwiesen gelten und damit wiederum die Treffsicherheit anderer Vorhersagealgorithmen überprüft wird, etwa bezüglich des politischen Engagements oder der Konsumorientierung, dann können aus stichprobenbasierten Korrelationen statistische Realitätsblasen erwachsen – ähnlich wie monetäre Wertblasen an Finanzmärkten, auf denen Schuldtitel zur Absicherung weiterer Kredite verwendet werden.

Ungeachtet dieser grundlegenden Begrenzungen und vieler spezieller methodischer Probleme einzelner Big-Data-Analysen, hängt ihre epistemische Autorität und performative Wirkung aber primär daran, welcher Wahrheitswert ihnen in der öffentlichen Debatte oder auch von einzelnen Akteuren zuerkannt wird. Wenn durch Big-Data-Analysen generierte Beschreibungen als Abbild der gesellschaftlichen Wirklichkeit gelten und das Wissen prägen, mit dem Akteure sich und die Welt deuten und ihren Umgang miteinander bestimmen, können sie „wirklich werden“ – direkt im individuellen Handeln, aber auch vermittelt über den journalistischen, wissenschaftlichen und politischen Diskurs.

So kann zum Beispiel, ähnlich wie für Methoden der politischen Umfrageforschung, auch für Big-Data-basierte Methoden der politischen Meinungsforschung die Frage gestellt werden, welche ontologischen Annahmen von Gesellschaft, Öffentlichkeit und Politik ihnen eingeschrieben werden, wie sie die Menschen als politische Bürgerinnen und Bürger konzipieren und auf welche Weise sie diese Annahmen und Konzepte im Forschungsprozess (bei der Datenerhebung und -analyse) wie auch bei der Verwendung der gewonnenen Ergebnisse real werden lassen (Osborne und Rose 1999; Law 2009). Ein Beispiel sind etwa Analysen von Social Media-Daten, die vom Verhalten einer selektiven Population auf die gesamte Gesellschaft schließen. Dass junge Nutzerinnen und Nutzer sozialer Netzwerke zuweilen sehr großzügig ihre persönlichen Daten teilen, hat die These der Post-Privacy-Gesellschaft genährt, also der Vorstellung, dass Gesellschaften heute weniger Wert auf den Schutz ihrer Privatsphäre legen als früher (Ganz 2014; Johnson 2010). Dass die Daten über das Verhalten von Nutzerinnen und Nutzern in Netzwerken häufig wenig Kontext liefern, wird nur selten thematisiert. So kann man bezweifeln, dass Nutzerinnen und Nutzer von Sozialen Medien repräsentativ für die Bevölkerung sind, dass sie ihre Daten im weiteren Lebensverlauf weiterhin ebenso freigiebig teilen werden, dass ihr Nutzungsverhalten Ausdruck einer unbekümmerten Haltung gegenüber ihrer Privatsphäre bedeutet etc. (Acquisti 2014). Wie bei anderen Methoden zeichnet sich das durch Big Data generierte Wissen also durch spezifische Beschränkungen aus, was sich auch auf die Realitäten auswirkt, die auf dieser Grundlage konstruiert werden.

### 3.2.2 Politische Performativität: Big-Data-gestützte Repräsentation kollektiver Interessen

In Bezug auf die epistemischen Wirkungen, die Big Data im Bereich der Politik zukommen, ist neben der Prägung gesellschaftlicher Realitäten auch der Einfluss entscheidend, den Big-Data-Analysen für die Generierung politischer Macht haben können. Big Data hat in diesem Sinn eine politische Performativität bei der Konstruktion kollektiver Subjekte mit spezifischen Formen der Identität, des Willens, der Bedürfnisse und Interessen.

#### Performative politische Repräsentation

Kollektive Subjekte, die politisch repräsentiert werden, existieren nicht bereits der Repräsentation vorgängig, so dass sie einfach wiedergegeben und originalgetreu abgebildet werden könnten (sei es „das Volk“, bestimmte Gruppen oder Klassen oder auch „die Weltgemeinschaft“). Sie können aber zunächst als der Politik zugrundeliegende „necessary fiction“ angerufen werden (Rosanvallon 2006 91, FN 40; Ezrahi 2012). In dem Maße, in dem derartige Artikulationen anerkannt werden, können sie *performativ* wirksam werden und tatsächlich das geteilte Bewusstsein einer gemeinsamen Identität, eines Willens und Interesses hervorbringen. Über die Mobilisierung kollektiver Handlungen und die Etablierung legitimer kollektiv bindender Normen können sich Kollektivrepräsentationen schließlich auch materiell-praktisch realisieren, z. B. durch körperliche Anwesenheit in Situationen, gemeinsame Bewegungen und Erfahrungen und die Ausbildung geteilter Vorlieben und Gewohnheiten (Bourdieu 2009 [1984], 2009 [1981]; Latour 2003; Butler 2015). Praktisch geht es in der Politik also um die Artikulation von „representative claims“ (Saward 2006) bezüglich der Identität, des Willens und der Interessen von Kollektivsubjekten und um die Anerkennung dementsprechender Behauptungen von denjenigen, die diese Kollektive konstituieren sollen.

#### Vielfältige und verteilte Repräsentationsformen

In westlich-modernen Gesellschaften erfolgt die performative Repräsentation kollektiver Identitäten, Werte und Interessen in vielfältiger Form verteilt über Kunst und Kulturbetrieb, Wissenschaft, Journalismus usw., nicht nur in explizit politischen Zusammenhängen oder formal regulierten Verfahren wie der demokratischen Wahl. Pierre Rosanvallon spricht von einer „allgemeinen Ökonomie der Repräsentation“ (Rosanvallon 2002, zitiert nach Disch 2008, S. 65-66; auch Rosanvallon 2006, S. 199-217). Mit der Ökonomie der Repräsentation wird das Arsenal der liberalen, direktdemokratischen und deliberativen Verfahren, aber auch der statistischen,

sozialwissenschaftlichen und ästhetischen Praktiken bezeichnet, mit deren Hilfe politische Identitäten vor dem Hintergrund sich wandelnder medial vermittelter Welterfahrungen konstruiert werden. Politikerinnen und Politiker müssen auf die in diesen verteilten Repräsentationsaktivitäten erzeugten Kollektivvorstellungen Bezug nehmen, sie müssen sich darauf abstützen und „daran andocken“, um selbst glaubhaft im Namen kollektiver Identitäten und Interessen sprechen zu können.

### **Big Data als neue politische Repräsentations-Technologie**

Der Begriff der Ökonomie der Repräsentation scheint sinnvoll, um den Einfluss neuer Technologien wie Big Data auf die Konstruktion der oder des Repräsentierten zu untersuchen. Sozialforschung und politische Meinungsumfragen spielen besonders in „verwissenschaftlichten“ Politikkulturen eine zentrale Rolle, also in solchen Kulturen, in denen wissenschaftlicher Evidenz überragende Bedeutung in der Legitimation kollektiver Ordnungsentscheidungen zukommt. In diesem Zusammenhang können die epistemischen Repräsentationen von gesellschaftlichen Gruppierungen und ihren Wünschen, Bedürfnissen und Interessen, die durch Big-Data-Analysen erzeugt werden, eine bedeutende politische Wirkung haben. Sie treten in Konkurrenz zu politischen Meinungsumfragen („polling“) und anderen Ansätzen der Sozialforschung (O’Connor et al. 2010; Law 2009).

Dort wo Big-Data-Analysen zur Beglaubigung politischer Repräsentationsansprüche und zur Generierung politischer Autorität verwendet werden, potenziert sich die performative Wirkung, die sie epistemisch besitzen (Voß und Amelung 2016). Im politischen Diskurs- und Handlungszusammenhang können Big-Data-Analysen, wenn sie als glaubhaft eingeschätzt werden, die Artikulation kollektiver Interessen und daraus abgeleitete politische Programme mitbestimmen, z. B. wenn mit Verweis auf Analysen digitaler Interaktion glaubhaft behauptet wird, im Namen der Bedürfnisse und Sorgen „der Öffentlichkeit“, „des Volkes“ oder bestimmter Gruppen wie der „99%“, „der Globalisierungsverlierer“ oder „der Jugend“ zu sprechen. Hier kann Big Data Auswirkungen darauf haben, auf welche Weise, für welche Themen und durch welche Akteure politische Autorität erzeugt werden kann.

Big-Data-Analysen können beispielsweise zum Einsatz kommen, um staatliche Entscheidungen und Handlungsprogramme zu legitimieren, wenn mit ihnen die Repräsentation der Interessen und Bedürfnisse von Bürgerinnen und Bürgern behauptet wird, z. B. in der Ausrichtung von öffentlichen Dienstleistungen. So wird daran gearbeitet, mit Methoden der „sentiment analysis“ und des „opinion mining“ die Bedürfnisse „der Bevölkerung“ oder bestimmter Anspruchsgruppen zu repräsentieren, um davon ausgehend die Abschaffung, die Ausweitung oder den Zuschnitt von staatlichen Leistungen zu legitimieren (Clarke und Margetts 2014). In der einschlägigen Literatur wird das Potenzial der Analyse von Online-Kommuni-

kation diskutiert, die klassische, umfragebasierte Meinungsforschung zu ersetzen (O'Connor et al. 2010 1; Tumasjan et al. 2010). Ein konkretes Beispiel findet sich bei der nationalen Statistikbehörde der Niederlande (CBS), die einen monatlichen Indikator für Verbraucherinnen- und Verbrauchervertrauen auf der Grundlage von *sentiment analysis* mit Daten aus Sozialen Medien erstellt. Dieses Vorgehen bietet dem CBS zufolge den Vorteil gegenüber umfragenbasierten Indizes, dass er schneller und häufiger aktualisiert werden könne. Das *Data Mining* selbst wird durch ein Unternehmen (Coosto) vorgenommen. Einem Bericht der niederländischen Zentralbank zufolge erwägen auch andere politische Entscheidungsträgerinnen und -träger, Wahrnehmungen von Bürgerinnen und Bürgern durch *social media sentiment analysis* zu ermitteln, etwa bezüglich der öffentlichen Sicherheit (Europäische Zentralbank 2014).

Ausgehend von der These, dass es sich bei Big Data um eine epistemische Innovation mit offenen Folgen handelt, sind politikwissenschaftlich besonders die Folgen von Interesse, die sich damit verbinden, dass Methoden der Big-Data-Analysen als ein Instrument der gesellschaftlichen Selbstbeobachtung Verwendung finden. Die Beachtung der Performativität dieser Methoden eröffnet dabei einen neuen Blick auf mögliche Folgen. Als Beitrag zum laufenden Prozess der informellen Technikfolgenabschätzung für Big-Data-Analysen kann es daher fruchtbar sein, den Horizont der gegenwärtigen Kontroversen, die sich auf die instrumentelle Macht konzentrieren, die aus ihrem Einsatz resultiert, um diejenigen Aspekte zu erweitern, die die Performativität der Methoden betreffen. Dann kommen neben technologisch versprochenen Folgen auch die „kollateralen Realitäten“ mit in den Blick, die durch sie erzeugt werden (Law 2012): Big-Data-Methoden sind epistemisch performativ (und in einem weiten Sinne politisch relevant), wenn die ihnen eingeschriebenen Ontologien von Gesellschaft, Politik, Öffentlichkeit und Bürgerinnen und Bürgern auf subtile Weise das Wissen der Gesellschaft von sich selbst prägen. Sie gestalten dann inhaltlich, welche Realitäten von Gesellschaft, Politik, Öffentlichkeit und Bürgerinnen und Bürgern kollektiv gelebt werden. Den Entwicklerinnen und Entwicklern und Anwenderinnen und Anwendern dieser Methoden kommt eine zentrale Rolle zu, kollektive Ordnungen mit zu gestalten, besonders dann, wenn ihnen eine gegenüber anderen Methoden erhöhte epistemische Glaubwürdigkeit zukommt. Dies gilt es anhand konkreter Anwendungsfälle zu erforschen.

Politisch performativ (und im engeren Sinne politisch relevant) werden Big-Data-Methoden, wenn die mit ihnen produzierten Repräsentationen kollektiver Subjekte – und ihrer Identität, ihres Willens, ihrer Bedürfnisse und Interessen – zur Generierung politischer Autorität und der Mobilisierung kollektiven Handelns sowie der Legitimation von Normen und von „öffentlichem Handeln“ beitragen. So steht es an, verschiedene Big-Data-Methoden auf die durch sie beförderten

Realitätsordnungen hin zu analysieren und in dieser Hinsicht sowohl miteinander wie auch mit anderen Methoden der Sozialforschung zu vergleichen (Marres und Gerlitz 2016). Dafür wäre es aber in einem ersten Schritt dringend erforderlich, dass die Methoden mit den ihnen zugrunde liegenden Annahmen, Modellierungen und Operationalisierungen öffentlich verfügbar gemacht werden. Man müsste etwa die Programmcodes der datenproduzierenden Geräte (Social Media-Plattformen etc.) sowie der Applikationen zur Analyse von Mustern in den so produzierten Datenmengen offenlegen.

Die politikwissenschaftliche Forschung zur Frage, in welcher Form Big Data das Bild prägt, das eine Gesellschaft von sich hat, und wie sich dies im politischen Prozess auswirkt, steht noch am Anfang. Im Nachgang der US-Präsidentenwahl von 2016 und mit Blick auf die ausstehende Bundestagswahl 2017 ist jedoch eine öffentliche Debatte darüber entflammt, ob und in welcher Form Big Data den öffentlichen Diskurs, Wahlkampfstrategien und Wahlergebnisse beeinflusst. Das folgende Kapitel leistet einen Beitrag zu dieser Debatte, indem es herausarbeitet, in welchem Maße Big Data Wahlkämpfe verändern kann und welche demokratietheoretischen Implikationen dies hat.

---

### **3.3 Big Data im Wahlkampf: Wählerinnen- und Wählermodellierung, Micro-Targeting und Repräsentationsansprüche**

Jeanette Hofmann

Die gegenwärtige Debatte über Big Data im Wahlkampf befasst sich damit, ob neuartige Daten und Auswertungsverfahren effektiver darin sind, Wählerinnen und Wähler zu überzeugen, und welche Kosten und Risiken damit verbunden sind. Meine These lautet jedoch, dass Big Data nicht nur (wie auch immer definierte) Effektivitäts- oder Effizienzgewinne in Wahlkämpfen mit sich bringen kann. Big Data beeinflusst vielmehr das Bild, das Parteien sich von Wählerinnen und Wählern und Bürgerinnen und Bürgern machen und dieses Bild wirkt wiederum darauf ein, wie Bürgerinnen und Bürger sich selbst im politischen Prozess verstehen.

### 3.3.1 Repräsentation als ein interaktiver Schaffensprozess

Das Verhältnis zwischen Parteien und (Wahl-)Volk wird durch die Idee der politischen Repräsentation bestimmt: Parteien und ihre Mandatsträgerinnen und Mandatsträger sollen die Bürgerinnen und Bürger im politischen System vertreten. Dabei ist das Repräsentationskonzept, auch aufgrund der Möglichkeiten der Mikrofokussierung durch Big Data, im Wandel. Repräsentation, so die Politikwissenschaftlerin Pitkin (1967, S. 8), zielt darauf ab, „etwas Abwesendes (wieder) anwesend zu machen“. Während Pitkin davon ausging, dass das zu Repräsentierende eine bekannte Größe darstellt, die sich in Form von Interessen oder Präferenzen ermitteln lässt, zeigt die neuere Repräsentationsforschung, dass schon das Repräsentationskonzept selbst widersprüchlich und problematisch ist. Die Idee eines Abbildungsverhältnisses wurde paradoxerweise in dem Moment zum Problem, in dem Individuen Trägerinnen und Träger von Freiheitsrechten werden: „The transition from a corporatist to an individualist society makes society less representable. For how to give a form – one open to description and recognition – to an agglomeration of individuals?“ (Rosanvallon 2006, S. 78).

Wenn man mit Rosanvallon davon ausgeht, dass demokratische Wahlen und die gewählten Repräsentierenden die Komplexität des Wahlvolkes faktisch nicht widerspiegeln können, dann stellt sich die Frage, wie man das Verhältnis zwischen Repräsentierenden und Repräsentierten angemessener fassen kann. Saward (2006) hat vorgeschlagen, die traditionelle Annahme eines (mehr oder minder gelingenden) Abbildungsverhältnisses durch das Konzept eines herzustellen „representative claims“ zu ersetzen. Die Interaktion zwischen Repräsentierenden und Repräsentierten erschöpft sich demzufolge nicht in einem unidirektionalen Informationsfluss, sondern erweist sich als ein schöpferisches, performatives Verhältnis. Die Rollen der beteiligten Akteure sind schöpferisch, weil Behauptungen über die aggregierbaren Anliegen, Interessen und Prioritäten der Beteiligten aktiv produziert werden müssen (Saward 2006, S. 310). Die Herstellung von *claims of representation* erfolgt weder willkürlich noch alternativlos, sie bedürfen der Anerkennung und Identifikation der Repräsentierten und müssen folglich mit diesen ausgehandelt werden (Saward 2006, S. 303-304). Ihre Wirksamkeit besteht im Erfolgsfalle darin, „to bring a potential audience to a self-conscious notion of itself as an audience“ (Saward 2006, S. 303). Repräsentationsbehauptungen sind somit wirklichkeitsschaffend in dem Maße, in dem sie den Repräsentierten, Individuen wie Kollektiven, im Rahmen eines Repräsentationsverhältnisses zur Identität verhelfen (Rosanvallon 2006).

Die Konzeptualisierung von politischer Repräsentation als interaktiver Schaffensprozess stellt das landläufige Verständnis auf den Kopf. Weder die Wählerschaft noch der Wille der Wählerinnen und Wähler können als eine unabhängige, dem

Repräsentationsvorgang vorausgehende Größe gelten; sie werden vielmehr im Akt der Repräsentation erst dialogisch konstituiert (Young zit. n. Disch 2008, S. 59). Der Wille der Wählerinnen und Wähler, so auch Rosanvallon (2006), sollte als eine bestätigende oder widersprechende Antwort auf etwas Drittes verstanden werden, das diesen Willen erst mobilisiert. Soziale oder politische Identität als Repräsentationsobjekt erscheint somit als ein Produkt des politischen Prozesses, der Kommunikation zwischen Wählerinnen und Wählern und den Gewählten. Wahlkämpfe können als ein spezieller Modus verstanden werden, um den Willen der Wählerinnen und Wähler hervorzubringen.

Die performative Perspektive auf politische Repräsentation rückt somit die Frage nach den Ressourcen und Techniken der Herstellung politischer Identitäten und ihrer Abbildung in den Mittelpunkt: Welcher Mittel bedienen sich politische Akteure, um den Willen der Wählerinnen und Wähler zu erkunden, zu deuten und zu erzählen? Welche Rolle kommt hier Big Data zu?

### 3.3.2 Wählermodelle: Objektivierung auf Widerruf

Folgt man den Berichten über Big-Data-gestützte Verfahren in den US-Präsidentenwahlkämpfen 2008 und 2016, steht im Fokus das *micro-targeting* von Individuen. Dieses zielt anders als früher nicht mehr in erster Linie auf Eigenschaften von Wählerinnen und Wählern, sondern auf ihr Verhalten. Mithilfe großer, vielfältiger Datensätze und lernender Algorithmen soll ihr künftiges Wahlverhalten besser vorhersagbar und damit kontrollierbar gemacht werden, so das Versprechen. Dies schließt auch Wählerinnen und Wähler des gegnerischen Lagers mit ein (Nickerson und Rogers 2014, S. 68). Die US-amerikanischen Wahlkampfteams rekrutieren seit einigen Jahren Verhaltenswissenschaftlerinnen und -wissenschaftler, um politische Aussagen zielgenau auf Individuen zuzuschneiden (Bimber 2014, S. 145). Die Verhaltenswissenschaftlerinnen und -wissenschaftler testen die Reaktionen auf verschiedene Versionen politischer Botschaften und setzen die erfolgreichen Varianten im Wahlkampf ein, um ihr Identifikationspotenzial dort erneut zu prüfen (Tufekci 2014; Kreiss 2012). Mithilfe solcher Analysen lässt sich die Interaktion mit den Bürgerinnen und Bürgern nun kontinuierlich aktualisieren und in Echtzeit an die hervorgerufene Gefühlslage anpassen: „Data on the fly“ so Kuchler, helfen beim „canvassing on the doorstep or [...] enable a change of tack if sentiment is not going their way“ (Kuchler 2016, S. 4; Chideya 2015). Zugleich wird das Modellieren der Wählerinnen und Wähler komplexer und berücksichtigt immer mehr „predictor variables“ (Bimber 2014), aus denen eine Objektivierung des Verhaltens der Wählerinnen und Wähler gewonnen wird.

Die Berechnung von Verhaltenswahrscheinlichkeiten der Wählerinnen und Wähler wird als *modeling* bezeichnet. Wie Mahr (2003, S. 70-73) ausführt, sind Modelle „Urteilsinhalte“, die sich bewähren müssen. Sie „treten zwischen das Subjekt und die Realität, für die sie [...] gleichsam zum Stellvertreter werden“. Nach außen hin tragen Modelle zur Objektivierung hypothetischer Sachverhalte bei, während sie „im Inneren der Inbegriff des Subjektiven und eine Quelle der Skepsis“ sind (Mahr 2003, S. 74). Eine äußere Objektivität erlangen Modelle, sofern Dritte sie in der Rolle, die ihnen zugeordnet ist, akzeptieren (Mahr 2003, S. 78). Das Modellieren der Wählerinnen und Wähler erzeugt einen objektivierten Zusammenhang zwischen den verfügbaren Datenpunkten und dem vermuteten künftigen Wahlverhalten. In der Praxis aber handelt es sich um eine Objektivierung auf Widerruf, da die teils impliziten, teils expliziten Annahmen, auf denen die Modellierung des Wahlverhaltens beruht, beständigen Prüfungen ausgesetzt sind. Die Modelle unterliegen einem kontinuierlichen Prozess der Neuberechnung von Verhaltensvorhersagen; sie können korrigiert und widerlegt werden.

### 3.3.3 Big Data für die Beeinflussung des Wählerinnen- und Wählerverhaltens

Funktional haben Modelle zwei Seiten; sie sind zugleich Modelle „von etwas“ und Modelle „für etwas“ (Mahr 2003, S. 74). Die Big-Data-gestützte Analyse der Wählerinnen und Wähler zeigt, dass der Zweck des Modellierens unmittelbar Einfluss auf das Modellierete nimmt. Die Modelle der US-amerikanischen Wählerinnen und Wähler orientieren sich an der Wettbewerbslogik des Wahlkampfes, die wiederum durch das US-Wahlsystem geprägt ist. Aus operativer Sicht lautet das Ziel, knappe Wahlkampffressourcen möglichst effektiv einzusetzen. Alle Parteien konzentrieren sich daher auf Individuen in den sogenannten *swing states*, die voraussichtlich an der Wahl teilnehmen. Insofern spielen Merkmale, die Rückschlüsse auf die Wahlbereitschaft zulassen, eine wichtige Rolle für die Eingrenzung der zu modellierenden Individuen.<sup>49</sup> Innerhalb dieser Gruppe richtet sich das Interesse vor allem auf die politische Orientierung und das wahrscheinliche Abstimmungsverhalten. Eine Rolle spielen ferner die Bereitschaft, für den Wahlkampf zu spenden sowie die Partei im Wahlkampf aktiv zu unterstützen. Das Verhalten der erfassten Personen wird anhand dieser Merkmale mit fortlaufend aktualisierten *predictive scores* modelliert, um tägliche verschiedene Szenarien des Wahlausgangs zu simulieren (Nickerson und Rogers 2014, S. 54; Scherer 2012). Im US-Wahlkampf von 2016 haben sich jedoch

---

49 Big Data war dafür bislang weniger relevant als der Zugang zu alten Wahlregistern.

auch die Grenzen der (auch Big-Data-basierten) Vorhersage gezeigt: Der Wahlsieg Donald Trumps war für viele Analytinnen und Analysten überraschend. Tausende von Datenpunkten pro Person ermöglichen keine umfassende Kontrolle über das Verhalten der Wählerinnen und Wähler. Vor diesem Hintergrund erscheint es angebracht, dem provisorischen Charakter von Wählermodellen in der öffentlichen Debatte mehr Aufmerksamkeit zu widmen: Sind Big-Data-basierte Modelle anderen Modellen überlegen? Wie gehen die Analytinnen und Analysten mit den Defiziten der Modelle um, etwa mit dem Problem, dass manche Bevölkerungsgruppen sehr viel weniger Datenspuren hinterlassen als andere? Welche Autorität erhalten (verschiedene) Modelle unter Wahlkampfstrateginnen und -strategen? Wie verhält sich die öffentliche Debatte zu Wählermodellen, was erfahren Wählerinnen und Wähler über die Modellierung ihres Verhaltens?

### 3.3.4 Big Data für die Herstellung von Repräsentation

Das „Modellieren von“ Wählerinnen und Wählern verknüpft die Verhaltensanalyse zudem mit politischen Themen und Positionen. Darauf aufbauend haben sich die Wahlkampftechniken, vor allem die gezielte Kommunikation mit den Wählerinnen und Wählern, grundlegend geändert (Tufekci 2014; Kreiss und Howard 2010). Das sogenannte „microtargeting“ (Bimber 2014; Kreiss 2012) tritt an die Stelle von zielgruppenorientierten Wahlkampfstrategien („Latinos“, „high-income voters“, „soccer moms“, etc.) und strebt danach, die Wählerinnen und Wähler als „normale Leute“ zu erfassen, „die auf jeweils unterschiedliche Weise mit dem politischen Prozess interagieren“ (Alter zit. n. Bimber 2014, S. 141). Auf der vorliegenden Datenbasis werden politische Themen und Argumente arrangiert, die Individuen persönlich adressieren, ihren Sinn fürs Politische mobilisieren, ihnen helfen, sich im politischen Raum zu verorten und sich mit einem der Kandidatinnen und Kandidaten zu identifizieren. Ob dies jedoch zu einer Abkehr der Parteien von traditionellen Positionen bedeutet, steht zu untersuchen. Generell ist bislang wenig bekannt darüber, in welcher Form und in welchem Ausmaß die Parteien maßgeschneiderte politische Botschaften und Mobilisierungsstrategien für einzelne, möglicherweise neuartig definierte Subgruppen entwerfen.

*Microtargeting*, so könnte man provokativ sagen, lässt Repräsentierte und Repräsentierende näher zusammenrücken und die Kommunikation zwischen ihnen dichter werden. Ohne dass die Akteurinnen und Akteure jemals ein Wort miteinander gesprochen haben müssen (Tufekci 2014), entwickelt sich doch ein nahezu intimes Verhältnis zwischen ihnen: Die politischen Parteien folgen ihren potenziellen Wählerinnen und Wählern auf Schritt und Tritt, sie reagieren auf An-

zeichen des Missfallens oder der drohenden Distanzierung mit einer Modifikation ihrer Botschaften. Mit Big Data und *microtargeting* scheint der von Rosanvallon hervorgehobene Widerspruch zwischen Individualisierung und Repräsentation in der Demokratie eine neue technokratische Antwort gefunden zu haben.

Doch das Verhältnis zwischen Repräsentierten und Repräsentanten ist hochgradig asymmetrisch: Die Wählerinnen und Wähler kontrollieren weder, welche Informationen die Wahlkampfapparate über sie sammeln, noch verfügen sie über vergleichbare Informationen bezüglich ihrer Kandidatinnen und Kandidaten. Diese Asymmetrie wird sukzessive institutionalisiert durch die Errichtung einer neuen technischen Infrastruktur: Während Wahlkampffressourcen früher nach einer Wahl aufgelöst wurden, sind nun Datenzentren entstanden, die dauerhaft Informationen über alle US-Wählerinnen und Wähler sammeln und unter einer persönlichen Kennung archivieren (Graff 2016). Die Datensätze werden zunehmend integriert und inzwischen auch kommerziell verwertet, etwa in Form von Themen- und Stimmungsanalysen, die durch *Data Mining* von sozialen Netzwerken gewonnen werden (Tufekci 2014: 6). Kandidatinnen und Kandidaten zahlen mutmaßlich hohe Beträge für den Zugriff auf die Datenbanken ihrer Parteien (Kreiss 2012). Diesem großen Informationsgefälle zum Trotz können *representative claims* auch in Zeiten von Big-Data-basiertem microtargeting nur in (impliziter oder expliziter) Zustimmung mit den Wählerinnen und Wählern politisch wirksam werden. Wie weit entsprechende Prozesse vorangeschritten sind, ist eine offene Frage. Eignen Repräsentierte sich die Big-Data-basierten Zuschreibungen an, wo widersetzen sie sich? Welche Erwartungen entstehen durch die Big-Data-basierten Repräsentationsansprüche wiederum an Parteien und ihre Kandidatinnen und Kandidaten? Verlieren sie möglicherweise ebenso an Autonomie wie ihre Wählerinnen und Wähler?

Anders als es die mediale Berichterstattung suggeriert, dient Big Data im Wahlkampf nicht allein der Analyse und Steuerung von individuellen Wählerinnen und Wählern. Big Data hat auch die Funktion, Repräsentationsansprüche zu formulieren und politische Identitäten und Kollektive zu schaffen. Doch hat dieser Prozess gerade erst begonnen. Ob die mithilfe von Big Data generierten *representative claims* durch die Repräsentierten angenommen werden, in welcher Art sie Identitäten und Kollektive mitprägen und welche Rolle die Repräsentierenden in diesen Deutungen erhalten, gilt es weiter zu erforschen.

### **3.4 Normativ hergestellte Erwartungen durch Big Data. Normierung, Normalisierung und Nudging**

Jan-Hendrik Passoth und Holger Straßheim

Durch Big Data können auch Werte geschaffen, Ziele gesetzt und Wege bestimmt werden, diese Ziele zu erreichen. Aus normativer Sicht stellt sich dabei die Frage: Erklärt Big Data Handlungen zu erwünschten oder unerwünschten, angemessenen oder unangemessenen Handlungen? Wie werden Akteure durch Big Data in moralischer Hinsicht bewertet?

In Bezug auf das Bündel von Technologien, Verfahren und Praktiken, die derzeit unter dem Schlagwort Big Data verhandelt werden, von einer dezidierten normativen Dimension zu sprechen, erscheint zunächst kontraintuitiv. In der klassischen, institutionentheoretischen Lesart, der sich die hier zugrundeliegende Unterscheidung verdankt, deutet die normative Dimension – im Gegensatz zu den beiden anderen – auf wenig oder kaum kodifizierte Phänomene: Normen, Werte oder Wertmaßstäbe. Dass aber auch so etwas Formales wie die Sammlung, Speicherung, Verarbeitung, Weitergabe und Auswertung von großen Datenmengen eine normative Dimension hat, macht ein Blick in die Sozial- und Kulturgeschichte der (amtlichen) Statistik deutlich, die in vielerlei Hinsicht als Ahne und als überkommenes Paradigma vieler Big-Data-Technologien und Verfahren gelten kann.

#### **3.4.1 Normierung durch Big Data**

Denn nicht nur etymologisch sind Staat und Statistik eng miteinander verbunden: Als „politische Arithmetik“ in England um 1676 und als „Abriss der neuesten Staatswissenschaft der vornehmsten Europäischen Reiche und Republiken“ in Preußen um 1749 war Statistik zunächst eine neben anderen Wissenschaften vom und im Staat. Die Hoffnung, über das „Gesetz der großen Zahl“ zu den Grundlagen einer „Göttliche(n) Ordnung in den Veränderungen des menschlichen Geschlechts“ (Süßmilch 1761) vorzustoßen, setzt sich dabei im 17. und 18. Jahrhundert nur langsam gegen einen eher symbolischen politischen Zahlengebrauch in kame-ralistischer Perspektive durch: die Sache des Staates, die Staatsräson, entsteht, so kann man sagen, mit der Umstellung von einem politischen Zahlengebrauch, der Größen, Reichtümer und Soldaten als Ausdruck fürstlicher Macht versteht, hin zur Entwicklung statistischer Mittel zum Zweck der Regierung von Bevölkerung, Territorium oder institutioneller Wirksamkeit.

Als Demographie, Gesundheitsstatistik und als Kriminalitätsstatistik wurden Fragen der politischen Regulierung eng mit epistemischer Praxis verbunden. Ihre Verbindung besteht seither in einem messenden und auswertenden Zugriff auf Populationen, deren Konturen und Eigenarten sich durch Normierung und Normalisierung politischer Subjekte (und Objekte) verschieben und verändern lassen. Dass es überhaupt so etwas wie Populationen im Werkzeugkasten der regulären Regierungspraxis gibt, ist nicht allein die Folge der Entwicklung von Messverfahren und statistischen Kenngrößen, mit denen dann etwas genauer bestimmt werden kann, was auch vorher schon Gegenstand politischer Eingriffe gewesen wäre. Vielmehr werden der Gesundheitszustand der Bevölkerung oder die Frage, wie sich Teile einer Region an deren Wohlstand beteiligen, erst in der Form von Quoten und statistischen Anteilen als mögliche Regulierungsgegenstände verfügbar. Der Erfolg von Regierungshandeln bemisst sich statistisch nicht an der Regulierung von Handeln, Devianz oder Gesundheit einzelner Regierter, auch nicht an der Regulierung des Handelns von festen Gruppen, sondern an der Veränderung statistischer Kennzahlen: Wie viele Tage sind Bürgerinnen und Bürger im Mittel krank? Sind Einkommen normalverteilt und wenn ja: lässt sich die Varianz durch Maßnahmen verändern? Welchen Effekt haben fiskalpolitische Maßnahmen auf disjunkte Klassen von Bürgerinnen und Bürgern? Die normativen Erwartungen, die durch diese „Governance by numbers“ (Heintz 2008) wirksam werden, sind dabei nicht notwendig an einer expliziten Vorstellung vom guten und gesunden Leben, vom angemessenen Lebensstandard oder eines erwünschten Verhältnisses von Stadt und Land, Bildungsgraden oder sozioökonomischer Lebenschancen orientiert, sondern immer auch an der statistischen Verteilung der in den Blick genommenen Merkmale. Der so entstehende „flexible Normalismus“ (Link 2006) orientiert sich an einem „homme moyen“, einen „Repräsentanten unserer ganzen Gattung, der außerdem im mittleren Maße alle Eigenschaften der anderen besitzt“ (Quetelet 1914, S. 165) – politische Regulierung kann sich normativ an ihm orientieren, indem sie entweder darauf setzt, dass ihm Bürgerinnen und Bürger ähnlicher sind oder darauf, seine Eigenschaften zu optimieren.

Die Allgegenwart der Sammlung großer Datenmengen, die durch die Verfügbarkeit großer Datenmengen erschließbaren statistischen Regelmäßigkeiten und die entweder neu entwickelten oder aufgrund gesteigerter Speicher- und Rechenkapazitäten endlich anwendbaren Verfahren der Verarbeitung großer Datenmengen, die gemeinhin unter dem Schlagwort Big Data zusammengefasst werden, lassen zumindest drei Dimensionen erkennen, in denen sich die normative Orientierung am „homme moyen“ in den letzten Jahren verschoben hat: Vervielfältigung, Personalisierung bzw. Granularisierung und zyklische Neuberechnung.

## Vervielfältigung

Während sich die an amtlicher Statistik orientierten Strategien der Normalisierung auf einen relativ festgelegten Satz von Eigenschaften (demographische, sozioökonomische oder ortsbezogene Faktoren etwa im Fall des Zensus) beziehen, die zudem noch unter den rechtlichen Vorgaben des Zensusgesetzes von einer systematisch eingeschränkten Zahl von Akteuren erhoben, ausgewertet und verwendet werden können, nimmt sowohl die Vielzahl als auch die Heterogenität der datengenerierenden und -verarbeitenden Akteure ebenso zu wie die Vielzahl und die Heterogenität der verfügbaren und verfügbar gemachten Daten. Ein Beispiel für die Veränderungen, die diese Vervielfältigungen und Diversifizierungen für das Gefüge normativer Erwartungen und die damit einhergehenden Normalisierungsstrategien haben können, ist die Nutzung datenbasierter Dienste im Gesundheitsbereich (vgl. Duttweiler et al. 2016): Neben den standardisierten Messungen im Rahmen der ärztlichen Versorgung ist in den letzten Jahren ein großes Angebot an Smartwatches, Fitness-Trackern, Smartphone-Apps und Plattformen im Netz getreten, die auf jeweils ganz unterschiedliche Weise Daten über den Gesundheitszustand ihrer Nutzerinnen und Nutzer erheben, mit den Daten anderer Nutzerinnen und Nutzer zusammenführen und auf ihrer Grundlage Empfehlungen und *Nudges* für die Anpassung des Verhaltens produzieren. Dabei stehen zwar zum Teil recht altbekannte Kenngrößen wie gelaufene Schritte oder Kalorienzufuhr und -verbrauch im Mittelpunkt. Aber schon bei ihnen ist die Art und Weise, mit Hilfe welcher Daten und aufgrund welcher Algorithmen aus den verbauten Accelerometern und Gyroskopen „Schritte“ werden, ebenso wenig standardisiert wie bei obskureren Berechnungen von Stimmungswerten aus Hautwiderstands- oder Pulsmessungen. Welche normierenden Effekte die Orientierung an einem derart vervielfältigten Gefüge von Daten hat, ist derzeit noch wenig erforscht. Erste Untersuchungen zu Mitgliedern der sogenannten „Self Tracking“-Szene (z. B. Duttweiler 2016; Vormbusch und Kappler 2016) zeigen jedoch, dass weder die im öffentlichen Diskurs so dominante Orientierung an permanenter Optimierung (vgl. Selke 2016) noch der erhoffte Aufbau von individueller Gegenexpertise zur eigenen Gesundheit gegenüber der schulmedizinischen Praxis (Lupton 2014) für die Einzelnen eine besonders starke Rolle spielen. Vielmehr ist eine zuweilen spielerische Orientierung an der experimentellen Modifikation der eigenen Position im jeweils verfügbaren Datenraum zu beobachten sowie eine selektive Orientierung an einzelnen Merkmalskombinationen (Crawford et al. 2015; Lupton 2016).

## Personalisierung und Granularisierung

Die Vermutung, dass sich die mit Big Data verbundenen normativen Erwartungen vor allem an den Einzelnen richten, ist in der öffentlichen Imagination der Effekte von Big Data allgegenwärtig. Nachdem schon 2006 Arvind Narayanan und Vitaly Shmatikov zeigen konnten, wie sich relativ einfach anonymisierte Filmbewertungen in einem Netflix-Datensatz de-anonymisieren lassen (Narayanan und Shmatikov 2006) und nachdem insbesondere Anfang der 2010er Jahre Meldungen über die recht einfache Identifikation individueller Nutzerinnen und Nutzer aus anonymisierten Mobilfunkdaten die Runde machten (Hardesty 2013), hat sich die Umkehrung des alten Sprichworts von der Nadel im Heuhaufen zu einer Art Maßgabe des Nachdenkens über die Konsequenzen von Big Data entwickelt: Will man eine Nadel finden, braucht man zunächst einen hinreichend großen Heuhaufen. Viele der aktuell unter Schlagworten wie *micro targeting* etwa im Bereich des Wahlkampfes geführten Debatten drehen sich um diese Figur: Big Data erlaubt die zielgenaue Einflussnahme auf das Verhalten einzelner Individuen. Am Fall des *targeting* allerdings lässt sich zeigen, dass es auf diese zielgenaue Personalisierung eigentlich in den wenigsten Fällen wirklich ankommt. Vielmehr arbeiten eine Reihe von Verfahren der Analyse großer Datenmengen weniger auf Vereinzelung als auf Granularisierung (Kucklick 2014) hin – auf die Identifikation mehr oder weniger sinnvoller „taxonomischer Kollektive“ (Wehner 2008). Denn die Merkwürdigkeit, dass z. B. gerade nach einem Einkauf bei einem großen Online-Shop das gerade gekaufte Produkt in den Werbebannern in den sozialen Netzwerken oder auf Nachrichtenseiten über Tage hinweg wieder auftaucht, ist kein Zeichen dafür, dass die Algorithmen nicht richtig arbeiten. Vielmehr hat der Kauf dazu geführt, das eigene Nutzungsprofil für einen bestimmten Zeitraum mit anderen zu einem Cluster zusammengefasst wurde, so dass das gekaufte Produkt anderen Nutzerinnen und Nutzern, die hinreichend ähnlich sind, auch angezeigt wird. Denn *targeting* setzt nicht darauf, einen einzelnen Nutzer oder eine einzelne Nutzerin zum Kauf zu bewegen, sondern darauf, in einer Menge von Nutzern die Konversionsrate<sup>50</sup> prozentual zu steigern.

## Zyklische Neuberechnung

In beiden Fällen – Vervielfältigung und Granularisierung – wird die Verarbeitung großer Datenmengen auch bei den heute verfügbaren Speicher- und Rechenkapazitäten nur in seltenen Fällen in Echtzeit durchgeführt, sondern zyklisch immer

---

50 Der Begriff Konversion bezeichnet den Statuswechsel von Kunden, etwa vom Nutzer einer Suchmaschine zum Besucher eines Online-Shops oder vom Interessenten zum Käufer.

wieder aufs Neue. Das hat zum Teil methodische Gründe – bei einigen Verfahren, etwa beim *clustering*, bekommt man mit einer Neuberechnung nach dem Hinzufügen neuer Daten eine andere Zusammensetzung der *cluster* als beim Einordnen der neuen Daten in die bestehenden *cluster* –, zum Teil aber auch schlicht Performancegründe: Auch heute dauert das Zusammenkopieren und selbst die parallelisierte Verarbeitung großer Datenmengen Zeit. In Bezug auf die normativen Erwartungen durch Big Data bedeutet das, dass auch datenbasierte Verfahren der Verhaltensmanipulation wie *nudging* oder *targeting* nicht darauf zielen müssen, das Verhalten der Einzelnen in eine bestimmte Richtung zu steuern: Vielmehr kann es passieren, dass Nutzerinnen und Nutzer nach einer Neuberechnung z. B. nicht mehr als gesundheitsbewusste Käuferinnen und Käufer von *functional food*, sondern als experimentierfreudige Kundinnen und Kunden von Pharmaprodukten angesprochen werden. Kurzum: Änderungen in der Verhaltenssteuerung sind nicht notwendigerweise intendiert.

### 3.4.2 Normierung und Verhaltenssteuerung

Vervielfältigung, Personalisierung und zyklische Neuberechnung verändern nicht nur die normative Orientierung am „*homme moyen*“. Sie fließen auch ein in staatliche Strategien der Verhaltenssteuerung. Die Analysemethoden von Big Data, in Verbindung mit verhaltenswissenschaftlich fundierten Policy-Instrumenten, gilt stellenweise gar als nächster Trend modernen Regierens, das weniger kodifizierte Regeln benötigt und letztlich die Erzeugung und Wirkung normativer Erwartungen auf die Mikroebene individuellen Verhaltens verlagert (Sunstein 2015).

#### Big Data und Nudging

Im vergangenen Jahrzehnt haben *nudges* die Praxis des Regierens bereits nachhaltig beeinflusst und beeinflussen sie noch immer. Big Data wird dabei zunehmend als natürliche Ergänzung gesehen, mit deren Hilfe sich individuelles Verhalten in dynamischen Kontexten maßgeschneidert beeinflussen lässt. Im vergangenen Jahr führte China, zunächst in einer Pilotphase, den so genannten „*Citizen Score*“ ein. Dabei werten Algorithmen Daten aus sozialen Netzwerken, Online-Shops und zahlreichen anderen Quellen aus und weisen jeder Bürgerin und jedem Bürger eine individuelle Punktzahl zu. Diese Punktzahl, gewissermaßen ein staatlicher „*Reputationsindex*“, wird nicht nur durch die eigenen Aktivitäten in sozialen Netzwerken bestimmt, sondern auch durch die Aktivitäten von Freundinnen und Freunden und Bekannten. Ihr Effekt auf den individuellen Alltag beschränkt sich nicht darauf, mit den eigenen Bekannten in einem Wettbewerb um den höchsten

„Citizen Score“ zu stehen. Vielmehr hat die Punktzahl erheblichen Einfluss auf das eigene Leben, z. B. auf die Chance, ein Visum für eine Auslandsreise zu erhalten (Hanfeld 2015). Wer bestimmte Lebenschancen anstrebt, muss sein Verhalten sowie sein soziales Umfeld also den Maßstäben angleichen, anhand derer Algorithmen ihre Bewertungen vornehmen. Auch wenn es sich beim „Citizen Score“ um einen Extremfall handelt, steht er doch emblematisch für die zunehmende Konvergenz von Techniken der Datenerhebung und -verarbeitung und Techniken der Verhaltensbeeinflussung.

*Nudging* bezeichnet den Versuch, im Rückgriff auf Erkenntnisse der Verhaltensökonomik das Verhalten von Menschen zu beeinflussen; eine wesentliche Rolle spielt dabei die Einsicht, dass menschliches Verhalten zwar systematisch vom Modell des homo oeconomicus abweicht, aber durch das planvolle Ausnutzen dieser Abweichungen beeinflusst werden kann (Thaler und Sunstein 2009). Eine solche Verhaltensbeeinflussung erzielen *nudges*, zu Deutsch etwa „Stupser/in“, indem sie scheinbar irrelevante Details von Entscheidungssituationen verändern. Zur Illustration dient häufig das „Kantinenbeispiel“, demzufolge sich der Konsum von Obst erheblich steigern lässt, wenn es auf die richtige Art und Weise präsentiert wird. Im vergangenen Jahrzehnt haben *nudges* sowie verwandte Techniken der Verhaltensregulierung und -beeinflussung die Praxis des Regierens nachhaltig verändert und verändern sie noch immer. Insbesondere im Vereinigten Königreich werden zunehmend verhaltensökonomisch informierte Politikinstrumente eingesetzt, z. B. in der Verbraucherschutz- und der Energiepolitik oder zur Steigerung der Steuermoral (Behavioural Insights Team 2014). Regierungen in den USA, Australien, Neuseeland, Frankreich, den Niederlanden oder Singapur befassen Teams aus Expertinnen und Experten mit der Entwicklung von Verhaltensinterventionen und deren experimenteller Überprüfung – zuletzt auch die Bundesregierung mit der Projektgruppe „Wirksam Regieren“ im Kanzleramt (Straßheim et al. 2015). Auf transnationaler Ebene kommt es zur Vernetzung öffentlicher, privater und zivilgesellschaftlicher Organisationen mit dem Ziel, verhaltensökonomisches Wissen zu produzieren und anzuwenden. So hat etwa die Weltbank in ihrem jüngsten Bericht „Mind, Society, and Behavior“ (2015) eine systematische Übertragung von Erkenntnissen der Verhaltensforschung auf den Bereich der Entwicklungspolitik vorgeschlagen. In der Europäischen Kommission arbeitet die „Foresights and Behavioural Insights Unit“ an der Entwicklung verhaltensregulierender Instrumente für mehrere Politikfelder, etwa für die Gesundheitspolitik, Finanzdienstleistungen oder den Datenschutz (Alemanno und Sibony 2016; Straßheim i. E.).

Laut Cass Sunstein, Ko-Autor des einflussreichen Buches „Nudge“ (Thaler und Sunstein 2009), liegt erst in der Verbindung von Techniken der Datenanalyse mit verhaltensregulierenden Maßnahmen die eigentliche Vollendung des *nudging*

(Sunstein 2015). Denn durch die datenbasierte Personalisierung von Verhaltensinterventionen gelingt, so Sunstein, die Anpassung von Verhaltensarchitekturen (*choice architectures*) an alle Lebenssituationen. Die Kombination von Big Data und *nudging* in Form von „personalized nudges“ ermöglicht es in diesem Sinne, individuelles Verhalten in dynamischen Kontexten maßgeschneidert zu beeinflussen. Statt eine „one size fits all“-Strategie zu verfolgen, wie es besonders bei frühen *nudges* erkennbar war, lassen sich im Falle personalisierter Strategien die Personengruppen, deren Verhalten beeinflusst werden soll, und das Maß der Verhaltensbeeinflussung ausdifferenzieren. Im Online-Handel beispielsweise geben Algorithmen auf der Grundlage von Daten der Kundinnen und Kunden individualisierte Produktempfehlungen, um bestimmte Kundinnen und Kunden zum Kauf spezifischer Produkte zu bewegen; in manchen Fällen ermitteln Algorithmen sogar individualisierte Preise (Reisch et al. 2016, S. 19-23).

### Staatliches „Big Nudging“

Die Nutzbarmachung verhaltenswissenschaftlicher Erkenntnisse für die Beeinflussung von Konsumierenden reicht bis in die 1920er Jahre zurück (Lemov 2005, S. 24f.). Auch der Einsatz von Big Data ist in der Privatwirtschaft bereits seit mehreren Jahren gängige Praxis. Neu und, wie das eingangs erwähnte Beispiel Chinas zeigt, von erheblicher Tragweite ist der Einsatz datengestützter Instrumente zur Verhaltensbeeinflussung durch den *Staat*. Da die Beziehung zwischen Staat und Bürgerinnen und Bürgern traditionell eine andere ist als die zwischen Käuferinnen und Käufern und Verkäuferinnen und Verkäufern, lassen sich die Erkenntnisse über privatwirtschaftliche Verhaltensbeeinflussungen nicht ohne weiteres auf den staatlichen Einsatz personalisierter Verhaltensinstrumente übertragen. In der Diskussion um den „libertären Paternalismus“ entwickeln sich nun neue Legitimations- und Politikvorstellungen. Gestützt auf diese Vorstellungen werden vergangene Krisenerscheinungen im Rückblick als Ergebnis begrenzter Rationalität erklärbar (so z. B. Thaler und Sunstein 2009, S. 255ff.); zugleich liefern sie für die Zukunft eine Neubeschreibung des Verhältnisses zwischen Bürgerinnen und Bürger und Staat. Mit Unterstützung von Verhaltensexpertinnen und -experten, so die Grundidee, kann es dem Staat gelingen, die begrenzte Rationalität im Verhalten von Bürgerinnen und Bürgern zu korrigieren und diesen so die Freiheit zur Verfolgung ihrer eigenen Ziele zurückzugeben (Sunstein 2014). Weil es sich bei *Big nudging* um ein neuartiges Phänomen handelt, liegen bisher, abgesehen von wenigen Ausnahmen (Helbing 2016; Hacker 2016), nur wenige Arbeiten zum Thema vor. Angesichts der zunehmenden Bedeutung von personalisierten (Big-Data-gestützten) Verhaltensregulierungen sowohl auf staatlicher als auch auf nichtstaatlicher Ebene ist es jedoch

nötig, die politischen und gesellschaftlichen Implikationen dieses Instruments in einem interdisziplinären Zugriff zu analysieren.

In Teilen der jüngeren Forschungsdiskussion wird die Kombination von Verhaltensinstrumenten und Big Data vor allem als prognostische Technologie diskutiert, als eine Form der algorithmischen Bewältigung komplexer Vorhersageprobleme („prediction problems“, vgl. Kleinberg et al. 2015). In der Gesundheits-, Finanz- oder Arbeitsmarktpolitik sollen mit Hilfe des *Machine Learnings* abweichende Verhaltensweisen im Voraus erkannt und dann durch den Einsatz entsprechender Verhaltenspolitiken regulierbar werden. Es geht also, so ein zentrales Argument, vor allem um Erkenntnisse über die Lebenserwartungen und Heilungschancen bestimmter Risikogruppen in der Gesundheitspolitik, die Vermittlungschancen jugendlicher Arbeitsloser oder die Erfolgsaussichten spezifischer Lehrmethoden in der Schulpolitik. Vor dem Hintergrund dieser Voraussagen können dann, gewissermaßen in Echtzeit, individuell abgestimmte Verhaltensregulierungen implementiert werden.

### Grenzen der Steuerbarkeit

Gleichwohl ist derzeit noch unklar, in welchem Ausmaß die Bewältigung von *prediction problems* durch die maschinenbasierte Erzeugung kognitiver (lernbereiter) Erwartungen nicht zugleich auch verknüpft ist mit der Bildung normativer (kontrafaktischer) Erwartungen. In seinem Buch über „Imagined Futures“ hat Jens Beckert jüngst eine Theorie fiktionaler Erwartungen vorgelegt, die als soziale und kulturelle Projektionen angesichts ungewisser Zukünfte immer auch das Ergebnis zugrundeliegender Wertvorstellungen und Deutungen sind. Dieser Einfluss von „extra-technical, non-calculative assessments expressed as imagined futures“ (Beckert 2016, S. 283) gilt in besonderem Maße für personalisierte, algorithmische Verhaltensinstrumente. Das hat vor allem drei Gründe:

1. Die maschinengestützte Analyse von Verhaltensweisen beruht nicht auf vorgegebenen Theorien oder Annahmen, sondern ist das Ergebnis eines in Echtzeit ablaufenden Prozesses der algorithmischen Reduktion von Komplexität durch Schwellenwerte und Clusterbildungen. Die sich daraus ergebenden Ordnungs- und Bewertungsmodelle beruhen also auf hochkomplexen, für die Nutzerinnen und Nutzer nicht rekonstruierbaren Rechenschritten. Jedes Ergebnis dieser algorithmischen *black box* muss daher gleichsam im Rückblick und unter Rückgriff auf bestehende Annahmen und Überzeugungen gedeutet werden. Verhaltensprognosen durch Big Data beruhen also nicht auf vorherigen Erwartungen, sondern sie müssen retroaktiv mit Sinn verknüpft und damit für politische oder andere Entscheidungen erst erzeugt werden. Sie sind eingebettet

in eine jeweils kontextspezifische und mit Anwendungszielen – z. B. dem Ziel des Abbaus devianter Verhaltensweisen – verbundene Erwartungskonstellation. Die Praktiken dieses *ex-post* Zuschreibungsprozesses sind bisher nicht systematisch untersucht worden.

2. Personalisierte Verhaltenspolitik ist insofern das Ergebnis multipler Erwartungsketten, die in der Abfolge von algorithmischer Kalkulation, sozialer Ordnungszuschreibung und Kategorisierungen, Irritation bisheriger Annahmen und politischer Projektion bzw. Trivialisierung das Ergebnis zahlreicher Teilentscheidungen und Relevanzannahmen sind. Damit wird eine doppelte Intransparenz erzeugt, weil a) die algorithmischen Analysen nicht rekonstruierbar und daher deutungsoffen sind und b) die daraus abgeleiteten Verhaltensregulierungen als Zugriffe auf unhinterfragte kognitive Heuristiken für Individuen auch nicht ohne weiteres erkennbar sind. Die doppelte Intransparenz personalisierter Verhaltenspolitiken macht diese zugleich in ihren normativen Vorannahmen unhinterfragbar.
3. Diese doppelte Intransparenz gewinnt ihre besondere Brisanz in Situationen wechselseitiger Beobachtung. Wenn – etwa in Form von Smartwatches oder Smartmeters – Ergebnisse algorithmischer Analysen zum Ausgangspunkt sozialer Selbst- und Fremdbewertung werden, dann kann es unter Umständen zu einer Mobilisierung von Wert- und Wertungsschemata kommen, deren Folgen für soziale Interaktionen einerseits, für individuelle Verhaltens- oder Wahrnehmungsweisen andererseits unklar sind. Dazu gehört auch die Frage, ob etwa durch *Health Apps* und *Wearables* eine Kultur der Mikrokonkurrenz entsteht, die letztlich Prinzipien der Solidarität in der Gesundheitspolitik erodieren läßt. Unklar ist damit dann auch, ob sich im Zuge der algorithmischen Beobachtung von Gesellschaft nicht womöglich zeitgleich die gesellschaftlichen Bedingungen für Verhaltensregulierungen selbst dynamisch verändern. Ein bekannter Nebeneffekt von Smart Meters sind beispielsweise Probleme des *moral licensing*, also einer Art Logik der moralischen Verrechnung z. B. der für das Energieverbrauchsverhalten gutgeschriebenen Einsparungen mit einem übermäßigen Wasser- oder Kraftstoffverbrauch. Diese Selbstlizenzierungen könnte sich mit der auf Big Data beruhenden Verhaltenssteuerungen noch verstärken. Es handelt sich in gewisser Weise um „*biases* zweiter Ordnung“, die überhaupt erst aus Korrekturen verhaltensökonomisch bereits beschriebener „*biases* erster Ordnung“ entstehen und dann vermutlich eine Kaskade weiterer Verhaltenssteuerungen auslösen. Die daraus resultierenden Nebenfolgen und unintendierten Effekte können erst dann abgeschätzt werden, wenn die Problematik normativer Erwartungsbildung im Zusammenhang mit personalisierten, algorithmischen

Verhaltenspolitiken systematisch zum Gegenstand interdisziplinärer und transdisziplinärer Diskussionen wird.

---

### 3.5 Wenn Big Data Regeln setzt. Regulativ hergestellte Erwartungen durch Big Data

Lena Ulbricht und Sebastian Haunss

Verändert Big Data, wie angenommen, die Bedingungen kollektiv bindenden Entscheidens, indem Regeln anders entstehen, indem neue Regeln gesetzt werden oder bestehende Regeln neuartig implementiert werden? In welcher Hinsicht wird Big Data für Regulierung eingesetzt? Wie dient Big Data der Feststellung regelkonformer Handlungen und (un)schuldiger oder (un)bestechlicher Akteure?

#### 3.5.1 Big Data und Regulierung: eine Systematisierung

Die regulative Dimension der Wirkungen von Big Data betrifft die Setzung und Implementierung kodifizierter Regeln. Es geht hier also nicht um Vorstellungen und Normen – diese Aspekte sind bereits in den vorangegangenen Abschnitten behandelt worden –, sondern um Gesetze, Verordnungen und andere Formen von üblicherweise schriftlich fixierten Regeln und Handlungsanweisungen. Aus politikwissenschaftlicher Sicht sind dabei insbesondere die Regeln interessant, die die Herstellung bzw. Bereitstellung öffentlicher Güter betreffen und die Folgen für die Bevölkerung als Ganzes oder für einzelne Bevölkerungsgruppen haben. Regeln dagegen, die vor allem für die Produktion privater Güter relevant sind und die beispielsweise innerhalb einer Organisation bestimmte (Produktions-)Abläufe festlegen, können zwar ebenfalls durch Big Data beeinflusst werden, stehen aber hier nicht im Fokus. Auch wie Big Data selbst Gegenstand von Regulierung werden kann, ist nicht Gegenstand dieses Beitrags, sondern wird in Kapitel 6 behandelt.

#### Forschungsstand über Big Data und Algorithmic Regulation

Es gibt inzwischen einen schnell anwachsenden Korpus an Forschung, der sich damit befasst, wie Big Data in ganz unterschiedlichen Bereichen die Setzung und Implementierung von Regeln beeinflusst. Eine Wendung, die dabei immer wieder auftaucht, ist die der *algorithmic regulation* (Morozov 2014) oder der *governance by algorithms* (Just und Latzer 2016). Zahlreiche Studien zeigen auf, dass Verfah-

ren der automatisierten Datenverarbeitung, die auf sehr große Datenbestände zurückgreifen, inzwischen dazu genutzt werden, um zwischen regelkonformen Handlungen und Regelverstößen zu unterscheiden und Akteure als schuldig oder unschuldig, bestechlich oder unbestechlich zu beurteilen. Beispiele finden sich in allen Politikfeldern: In der Gesundheitspolitik werden über *Data Mining* Zielgruppen identifiziert und auf dieser Grundlage ihr Zugang zu Fördermaßnahmen bestimmt, wie es etwa durch das US-Department of Veterans Affairs (VA) praktiziert wird (Fihn et al. 2014; Conn 2014). In der Sozialpolitik dient Big Data der Kontrolle von Sozialhilfeempfängerinnen und -empfängern, um Betrug zu vermeiden: Mithilfe von kombinierten Datensätzen und *Data Mining* Verfahren werden Risikogruppen bestimmt und durch Vorladungen und Hausbesuche kontrolliert (Maki 2011, S. 52-54). In der Verkehrspolitik dient Big Data der Lenkung von Verkehrsströmen, etwa zur Vermeidung von Staus im Autoverkehr (Biem et al. 2010) oder von Engpässen im öffentlichen Nahverkehr (Noursalehi und Koutsopoulos 2016), indem Verkehrsteilnehmerinnen und -teilnehmer auf entsprechende Informationen zugreifen können. In der Energiepolitik werden die von Smart Meter generierten Daten dafür eingesetzt, um Endnutzerinnen und -nutzer zum Stromsparen zu animieren (McKenna et al. 2012). Auch in der Kriminalitätsbekämpfung wird *Data Mining* in großen Datensätzen eingesetzt, um zu bestimmen, wo besonders eingehende Kontrollen einer Straftat vorbeugen oder diese aufdecken können, etwa durch die flächendeckende Prüfung von Steuererklärungen (Hoyer und Schönwitz 2015) oder bei der Analyse von Einbruchdiebstahl (Merz 2016). Big Data in Form von integrierten Datenquellen und *Data Mining* Verfahren dient zudem der Kontrolle von grenzüberschreitender Mobilität, wie etwa beim Warenimport/-export (Barbero et al. 2016, S. 40) und bei Personenkontrollen an Flughäfen (Ajana 2015). Nachrichtendienste wie die NSA nutzen und analysieren große Mengen von Kommunikationsdaten, um zwischen US-Bürgerinnen und Bürgern und anderen Personen unterscheiden zu können. Auf diese Weise entsteht eine „algorithmically defined citizenship“, welche bestimmt, ob die Daten einer Person durchsucht werden dürfen oder nicht (Cheney-Lippold 2016). Dabei kann Big Data auch von zivilgesellschaftlichen Organisationen zur Bewertung staatlicher Leistungen genutzt werden, der Rekrutierung von Unterstützerinnen und Unterstützern dienen oder zur Entwicklung neuartiger Handlungsstrategien (Goerge 2014). Weitere Beispiele für Big-Data-basierte Anwendungen im öffentlichen Sektor propagiert die EU (Europäische Kommission 2017). Allerdings fehlt es an einer systematischen Aufbereitung der Arten, in denen Big-Data-bezogene Praktiken zur Regulierung eingesetzt werden. Regulierung wird an dieser Stelle verstanden als die absichtsvolle Setzung von Regeln, ihre Kontrolle und Implementierung (Black 2008). Im Folgenden wollen wir zunächst einmal eine Reihe von Dimensionen nennen, anhand derer die regulative Wirkung von

Big Data systematisiert werden könnte. Diese Dimensionen werden dann anhand von zwei Fallbeispielen diskutiert.

### **Versuch einer Systematisierung**

Big Data kann sowohl die Setzung als auch die Implementierung von Regeln beeinflussen. Hinsichtlich beider Ebenen sind wiederum drei Aspekte zu berücksichtigen:

1. Welche Aspekte von Big Data beeinflussen Regulierung?
2. Welche Regeln beeinflusst Big Data?
3. Wie weitreichend ist der Einfluss von Big Data?

*Welche Aspekte von Big Data beeinflussen Regulierung?* Hier ist zu unterscheiden, ob der Einfluss von Big Data auf Regulierungsprozesse in erster Linie auf einen oder mehrere der drei häufig zitierten Eigenschaften von Big-Data-Datensätzen – großes Volumen, große Vielfalt und hohe Geschwindigkeit der Datengenerierung<sup>51</sup> – zurückzuführen ist, oder ob es primär neue Analysemethoden wie *Data Mining*, künstliche Intelligenz, maschinelles Lernen, prädiktive Analysen etc. sind, die eine regulative Wirkung entfalten. Natürlich hängen beide Aspekte miteinander zusammen: Die neuartigen Eigenschaften der Daten rufen nach neuen Methoden zu ihrer Analyse. Aber auch Big-Data-Datensätze können mit ganz klassischen Methoden der Häufigkeitsauszählung und der Schlagwortsuche bearbeitet werden und Mustererkennung oder maschinelles Lernen funktioniert auch mit relativ kleinen, traditionellen Datensätzen.

Wichtig ist zudem zu unterscheiden, ob zur Regulierung durch Big Data vor allem sektorspezifische oder unspezifische Daten genutzt werden. Sektorspezifische Daten werden mit Absicht erhoben und es besteht die Erwartung, dass sie mit einem bestimmten Ziel im Zusammenhang stehen. Spezifische Gesundheitsdaten sind etwa Daten wie sie in Arztpraxen, Krankenhäusern und Apotheken generiert werden. Unspezifische Daten wurden hingegen für andere Zwecke oder ohne besondere Intention erhoben. Als unspezifische Gesundheitsdaten könnte man zum Beispiel Smartphonedaten verstehen, die u. a. auch zur Analyse von Gesundheitszuständen und Lebensstil verwendet werden können. Mit Blick auf die Informationsbasis für staatliche Entscheidungen ist bedeutsam, dass Regierungen heute zunehmend damit experimentieren, neben den sektorspezifischen auch unspezifische Daten zu verwenden. Unspezifische Daten bedeuten aber größere Risiken mit Blick auf die Qualität, Legalität und gesellschaftliche Akzeptanz im Vergleich zu traditionellen sektorspezifischen Daten, sie wecken aber auch größere Erwartungen mit

---

51 <http://www.gartner.com/it-glossary/big-data/>.

Blick auf ihr Innovationspotenzial (Schintler und Kulkarni 2014, Washington 2014). Die Frage nach Datenquellen und Analyseverfahren ist somit stets auch eine nach dem angemessenen Gleichgewicht zwischen Risiken und Potenzialen datenbasierter Innovation.

*Welche Regeln beeinflusst Big Data?* In der zweiten Dimension wäre zu fragen, auf welcher Ebene eine Big-Data-basierte Regulierung wirkt: Beeinflusst Big Data die (Meta-)Regeln der Regelsetzung und Implementierung oder ist der Einfluss auf einzelne konkrete Regeln und Implementierungen beschränkt? Ändern sich durch Big Data also die Rahmenbedingungen, unter denen Regeln entstehen und modifiziert werden, also beispielsweise parlamentarische Entscheidungsprozesse oder Verwaltungshandeln? Denkbar wäre etwa, dass Datensätze, die stetig wachsen und sich verändern, dazu führen, dass Regeln häufiger überprüft und überarbeitet werden. Möglich wäre auch, dass Regeln weniger präzise kodifiziert werden (in Gesetzen, in Handlungsanweisungen) und verstärkt durch Interpretationsarbeit geschaffen werden, etwa durch Auslegung von Berechnungen. Möglicherweise beeinflusst Big Data aber gar nicht die Regeln der Regelsetzung sondern nur einzelne Regeln, die zum Beispiel im Rahmen von *Predictive Policing* den Einsatz von Funkstreifen im Stadtgebiet regeln? Aus dieser Perspektive ist wiederum zu fragen, ob durch Big Data bestehende Regeln nur verändert bzw. anders implementiert werden oder ob durch Big Data andersartige Regeln gesetzt werden, zum Beispiel solche, die stärker nach feiner granulierten Subgruppen unterscheiden. Hier wäre insbesondere interessant, ob es bestimmte Politikfelder gibt, die besonders anfällig oder empfänglich für Big-Data-basierte Regulierung sind und welche Gründe dies gegebenenfalls hat. Beeinflusst Big Data vor allem sicherheitspolitische Regulierung, während bildungspolitische Entscheidungen weitaus weniger davon betroffen sind? Oder werden sicherheitspolitische Regeln weiterhin durch Expertise, Werte und Ziele menschlicher Akteure bestimmt? Möglicherweise gibt es Lebensbereiche und Politikfelder, die der Effizienzorientierung von Big Data zuwiderlaufen, etwa Kunst und Kultur?

*Wie weitgehend ist der Einfluss von Big Data?* Die dritte Dimension betrifft schließlich den Grad der Beeinflussung und der verbleibenden Entscheidungsspielräume: Ersetzt Big Data menschliche Regelsetzung und Implementierung durch automatische Verfahren, wie dies bereits im *high velocity trading* praktiziert wird, oder liefern Big-Data-Analysen nur einen Baustein, der letztlich von menschlichen Entscheidungsträgerinnen und Entscheidungsträgern berücksichtigt werden kann oder auch nicht? Steht der Mitarbeiterin oder dem Mitarbeiter einer Versicherung frei, trotz negativem *scoring* einen günstigen Versicherungstarif anzubieten, oder wird die Entscheidung darüber letztlich in die Analysealgorithmen verlagert? Man muss also zwischen Big-Data-basierter Regulierung mit großen und mit kleinen

Umsetzungsspielräumen unterscheiden. Große Entscheidungsspielräume bietet ein Verfahren, bei dem durch Big Data ermittelte Muster eine von vielen Informationen sind, die bei der Regelumsetzung als Entscheidungsgrundlage dienen. Geringe Spielräume finden sich dann, wenn das durch Big Data ermittelte Ergebnis quasi-automatisch umgesetzt wird. Mit dieser Frage verbunden ist auch jene der Folgen für die menschliche Autonomie und für das Verhältnis zwischen Bürgerinnen und Bürgern und Staat. Stimmt die These, dass durch Big-Data-Anwendungen die staatliche Kontrolle zunimmt und bürgerliche Freiheiten systematisch eingeschränkt werden?<sup>52</sup>

Für alle drei Dimensionen gilt: Die Unterscheidung ist tatsächlich nicht dichotom. Die beiden Optionen beschreiben vielmehr die Pole eines Kontinuums, auf dem die möglichen regulatorischen Wirkungen von Big Data angesiedelt sind.

Anhand ausgewählter Beispiele wird im Folgenden aufgezeigt, in welcher Form Big Data zur Regelsetzung und Implementierung eingesetzt werden kann – mit besonderem Fokus auf die Daten, die hierzu herangezogen werden und der Verbindlichkeit der so generierten Regeln. Das erste Beispiel widmet sich den PNR-Datenbanken, die unspezifische Daten für Terrorismus und Kriminalitätsbekämpfung nutzen und die sehr geringe Spielräume für Entscheidungen lassen. Der zweite Fall widmet sich Big Data in der Gesundheitspolitik, wo überwiegend spezifische Daten verwendet werden und in vielen Fällen größere Handlungsspielräume verbleiben. Die Auswahl der Beispiele als möglichst unterschiedliche Fälle soll zeigen, dass es kein einheitliches Modell und keinen eindeutigen Trend für Big-Data-basierte Regulierung gibt, sondern eine große Spannweite an Möglichkeiten.

### 3.5.2 Fallbeispiele zu Regulierung durch Big Data

#### Fluggastdaten (PNR)

Fluggastdaten (passenger name records, PNR) gehören zu den sogenannten *computerized reservation systems* (CRS) oder auch *global distribution systems* (GDS): Diese Reservierungssysteme werden von kommerziellen Anbietern betrieben (Has-

---

52 Die Forschung über „governing by numbers“ und Quantifizierung deutet auf eine große Autorität quantifizierter Regulierungsverfahren hin (Miller 2001, Espeland und Stevens 2008). Allerdings gibt es nur wenige empirische Studien, die dies auch mit Blick auf neuere algorithmische Verfahren belegen können (zur Unterstreichung der Forschungslücke etwa Christin et al. 2016, S. 7).

brouck 2016; Moenchel 2010)<sup>53</sup> und in großem Ausmaß zu kommerziellen Zwecken genutzt: Reiseveranstalter benötigen sie, um die verschiedenen Elemente einer Reise integriert zu organisieren, etwa Flüge unterschiedlicher Fluggesellschaften, verschiedene Verkehrsmittel, Hotels, Mietwagen etc. Diese Reservierungssysteme umfassen Datenbanken, die global Daten über Reisende erfassen, die PNR. Diese werden seit den Terroranschlägen des 11. September zunehmend von Regierungen für Sicherheitszwecke genutzt, etwa den Regierungen der USA, Kanadas, Australiens. Auch in der EU werden die PNR für Flüge zwischen der EU und dem EU-Ausland sowie ausgewählten innereuropäischen Flügen von staatlichen Sicherheitsbehörden ausgewertet, ab 2017 in allen EU-Staaten (European Parliament 2016).

#### *Welche Aspekte von Big Data beeinflussen Regulierung?*

Die PNR können als Big Data verstanden werden in der Hinsicht, dass hier große Datensätze angelegt werden, die sich aus vielfältigen Quellen speisen: Reiseveranstalter, Buchungsportale, Fluggesellschaften, Hotels, Autoverleihe etc. Reisende sind mit bis zu 60 Datenpunkten abgebildet (House of Lords 2007, S. 9): Daten zur Reise, Kontaktdaten, Zahlungsdaten, Daten zu den Umständen der Buchung, Mitreisende etc. Diese Daten sind für Sicherheitsanalysen als unspezifisch zu betrachten, da sie keine direkten Aussagen über legales oder illegales Verhalten beinhalten. Vielmehr bilden sie Mobilitätsverhalten ab, aber auch Konsum, Freizeit, Kommunikation, Privatleben und berufliche Aktivitäten. Ob diese Daten mit weiteren Daten angereichert werden, etwa aus sozialen Netzwerken, ist derzeit nicht erwiesen.<sup>54</sup> Die PNR sind mit Blick auf Datenschutz nur teilweise reguliert: Es gibt Begrenzungen des staatlichen Zugriffs<sup>55</sup>, aber kaum Kontrolle der kommerziellen Datenbanken. Entsprechend muss man davon ausgehen, dass die Daten hier nicht systematisch nach einer bestimmten Frist gelöscht werden und somit wachsen und sich stetig verändern (Hasbrouck 2016). Die PNR werden durch Sicherheitsbehörden mittels *Data Mining* Verfahren analysiert: Algorithmen suchen nach Mustern und Faktoren

---

53 Die drei größten Unternehmen sind Amadeus, Sabre und Travelport. Letzteres pflegt zwei CRS: Worldspan und Galileo. Es gibt aber weitere Unternehmen, die CRS betreiben, unter anderem Alphabet (Google) (Hasbrouck 2016).

54 Die Praxis bei Grenzkontrollen, Passagiere aufzufordern, Sicherheitsbeamten Zugriff auf ihre Facebook-Accounts zu gewähren, deutet jedoch darauf hin, dass Sicherheitsbehörden ein Potenzial darin sehen, Social Media-Daten für Sicherheitszwecke auszuwerten (Nixon 2016). <http://edition.cnn.com/2017/01/29/politics/donald-trump-immigrant-policy-social-media-contacts/index.html>

55 Gesetzliche Regelungen setzen staatlichen Sicherheitsbehörden Grenzen für die Zugriffsdauer: US-Behörden haben etwa 15 Jahre lang Zugriff auf die Daten einer Person, in der EU sind es 5 Jahre.

(*classifiers*), mit denen man mutmaßlich unerwünschte Personen klassifizieren kann: Drogenkuriere, Terroristinnen und Terroristen, illegale Einwanderinnen und Einwander oder Personen, die zu solchen werden könnten (Amoore und de Goede 2012). Anders als zuvor ist es also nicht mehr nötig, Informationen über konkret geplante terroristische Taten einer Person zu haben, um diese am Flughafen zu verhaften, so die Argumentation. Vielmehr erlaube es das *Data Mining*, Menschen anhand ihres Reiseverhaltens<sup>56</sup> als (zukünftige) Terroristinnen und Terroristen oder Kriminelle zu identifizieren (Deutscher Bundestag 2017, Europäische Kommission 2016c).

### *Welche Regeln beeinflusst Big Data?*

Die sicherheitspolitischen Analysen auf Grundlage der PNR bilden die Grundlage dafür, dass Individuen in Risikogruppen eingeteilt werden: Passagiere, für die ein hohes Risiko festgelegt wird, werden intensiveren Kontrollen an den Flughäfen unterzogen, in Extremfällen werden sie auf sogenannte *no fly lists* gesetzt und am Reisen gehindert. Passagiere hingegen, die als wenig riskant eingestuft werden, müssen weniger strikte Kontrollen auf sich nehmen. Big Data verändert also die Art und Weise, wie Regeln implementiert werden: etwa durch die Zuteilung von Personen in (immer feiner granuliert) Gruppen. Die Regeln selbst („Terroristen dürfen kein Flugzeug besteigen.“ „Hochrisikopersonen werden am Flughafen besonders penibel kontrolliert.“) ändern sich nicht durch Big Data. Die (Meta-)Regeln der Regelsetzung für die Regulierung von Flugreisenden scheinen sich nicht grundlegend durch die PNR zu ändern. Nach wie vor werden die entsprechenden Regeln durch Parlamente und Verwaltungen nach deliberativen und bürokratischen Verfahren beschlossen. Was sich allerdings ändert, ist eine Zunahme an Interpretationsleistungen und Klassifikationsregeln, die für die Auslegung der Regeln notwendig ist. Zudem kann man davon ausgehen, dass die PNR-basierte Evidenz die Formulierung von Regeln beeinflusst: Wo man Regeln auf Subgruppen zuschneiden kann, werden manche Regeln vielleicht restriktiver und andere liberaler gestaltet und es häufen sich die Regeln (Amoore 2011, S. 31f.). So lautet ein Argument für die Nutzung der PNR für Sicherheitszwecke etwa, dass die Mehrheit der Reisenden schneller und leichter reisen kann: Intensive Kontrollen an den Flughäfen entfallen, da ihre Daten im Vorfeld der Reise bereits überprüft wurden (Sales 2015). Zugleich ist eine kleinere, aber dennoch substantielle Gruppe an Passagieren – die Risikogruppe – anderen Regeln unterworfen, die ihnen das Reisen erschweren oder unmöglich machen. Bei

---

56 Zum Reiseverhalten, so wie es durch die PNR abgebildet wird, zählen nicht nur die tatsächlich getätigten Reisen, sondern auch die reservierten und nicht angetretenen Reisen, die Reisevorbereitungen, die Kommunikation, das Zahlungsverhalten und nicht zuletzt (geplante sowie zufällige) Beziehungen zu anderen Personen im Datensatz.

immer stärker ausdifferenzierten Subgruppen läuft die Praxis entsprechend auf eine Big-Data-gestützte Individualisierung von Mobilitätsregeln hinaus.

#### *Wie weitgehend ist der Einfluss von Big Data?*

Diese Regeln für die Erteilung von Reiseerlaubnissen und Kontrollintensität an Flughäfen werden nicht maschinell umgesetzt. Die letzte Entscheidung darüber, wer ein Flugzeug besteigen darf oder wer eine besondere Kontrolle durchlaufen muss, liegt bei den Sicherheitsbeamtinnen und -beamten. Allerdings ist davon auszugehen, dass die algorithmisch ermittelten Risiken über eine große Autorität verfügen (Amoore 2011, S. 31). Sollte es zutreffen, dass Sicherheitsbeamtinnen und -beamte und Mitarbeiterinnen und Mitarbeiter von Fluggesellschaften die Anweisung erhalten, ihre Entscheidungen eng an die PNR-basierten Risiken zu orientieren, könnte man von einer quasi-automatischen Regulierung der Mobilität von Flugreisenden sprechen. Ob und in welcher Art Sicherheitsbeamtinnen und -beamten tatsächliche Entscheidungsspielräume beim Umgang mit den PNR-basierten Analysen verbleiben, bleibt zu erforschen. Allerdings sind kaum Informationen über die konkreten Schritte der Nutzung der PNR verfügbar.<sup>57</sup> Dies wird mit Sicherheitsinteressen und den Geschäftsgeheimnissen der Unternehmen begründet, ist aber aus demokratietheoretischer Perspektive hoch problematisch.

Ein weiteres Problem der regulativen Wirkung der PNR ist, dass die meisten „Risikopassagiere“ kein Sicherheitsrisiko darstellen<sup>58</sup>. Nicht aufgrund ihrer Identität, sondern aufgrund eines datenbasierten Risikowerts sehen sie ihr Grundrecht auf Mobilität substantziell eingeschränkt, was in manchen Berufen und Lebenssituationen eine große Belastung bedeuten kann (Lyon 2007). Ein weiteres Problem ist, dass Reisende kaum Einfluss auf die Daten haben, anhand derer ihre Risiken ermittelt werden (Korff und Goerge 2015). Dies verletzt ihre informationelle Selbstbestimmung. Auch das rechtsstaatliche Prinzip der Gleichheit vor dem Recht ist betroffen, denn nicht einmal Reisende mit gleichem Verhalten sind den gleichen Regeln unterworfen: Die Reiseveranstalter und sonstigen Intermediäre, die Daten über Reisende übermitteln, kooperieren mit unterschiedlichen Reservierungssystemen. Für eine Reise entstehen somit Daten bei unterschiedlichen Reservierungssystemen und das Ergebnis sind lückenhafte und unterschiedliche Dossiers für jeden Fluggast je nach Reservierungssystem (Hasbrouck 2016).

---

57 So geben Europäische Fluggesellschaften Passagieren etwa über ihre Daten im PNR Auskunft, aber nicht darüber, wer auf diese Daten Zugriff erhalten hat (Hasbrouck 2016).

58 Regelmäßig wird darüber berichtet, dass Personen irrtümlich auf No Fly Listen gesetzt werden (Krieg 2015).

## Routinedaten für Gesundheitsforschung

Der Gesundheitsbereich spielt in Studien zu Möglichkeiten und Risiken von Big Data häufig eine prominente Rolle (z. B. OECD 2015; Barbero et al. 2016). Betrachten wir die durch Big Data beeinflussten Regulierungen im Gesundheitsbereich, dann lässt sich zuerst einmal feststellen, dass viele der in der Literatur diskutierten Anwendungsszenarien von Big Data bisher häufig noch nicht den Status von Forschungs- und Pilotprojekten verlassen haben und daher die tatsächlichen regulatorischen Konsequenzen von Big Data im Gesundheitsbereich oft noch kaum zu beurteilen sind.

### *Welche Aspekte von Big Data beeinflussen Regulierung?*

Im Gesundheitsbereich existieren sehr große, oft weit zurückreichende personalisierte Datensätze, die besonders sensible Daten enthalten. Daher wird auch in praktisch allen Beiträgen zu Big Data im Gesundheitsbereich auf die Problematik des Datenschutzes hingewiesen. Neben den Daten aus klinischen Studien geht es hier vor allem um die im Gesundheitswesen anfallenden „Routinedaten“, also Diagnosen, Arzneimittelgaben, Therapien und Abrechnungsdaten. Je nach Struktur des nationalen Gesundheitssystems sind diese Daten in staatlich oder korporatistisch organisierten Gesundheitssystemen an einer oder an wenigen zentralen Stellen verfügbar oder – in Staaten mit überwiegend privater Gesundheitsvorsorge – bei einzelnen privaten Akteuren in großem Umfang vorhanden (Swart et al. 2014). In Deutschland verfügt beispielsweise der Verband der Gesetzlichen Krankenkassen (GKV) über die Routinedaten von gut 70 Millionen Versicherten. In Großbritannien laufen alle Patientendaten beim National Health System zusammen, in den USA verfügt der größte private Gesundheitsanbieter, Kaiser Permanente, über mehr als 9 Millionen Patientendatensätze (Gothe 2014).

Bei diesen Daten handelt es sich zwar um sehr große, aber um bereichsspezifische Datensätze, die im Gesundheitssystem erhoben werden. Sie werden in der Regel mit klassischen statistischen Methoden aggregiert und ausgewertet. Neuere Entwicklungen wie z. B. die Übermittlung von Daten aus Gesundheitsapps an Krankenkassen führen allerdings dazu, dass die Grenze zwischen im Gesundheitssystem und außerhalb des Systems erhobener Daten schwieriger zu ziehen ist.

Zudem gibt es eine Reihe von Pilotprojekten, in denen andere Daten und andere Analysemethoden zum Einsatz kommen. So haben beispielsweise die kanadische Public Health Agency in Zusammenarbeit mit der Weltgesundheitsorganisation (WHO) seit Ende der 1990er Jahre ein globales Gesundheitsinformations-Netzwerk (Global Public Health Intelligence Network, GPHIN) aufgebaut, das globale Nachrichtenquellen zur Vorhersage der Ausbreitung von Infektionskrankheiten

nutzt und das damit darauf abzielt, gesundheitspolitische Entscheidungen auf der Basis von nicht im Gesundheitswesen erhobenen Daten zu ermöglichen.

Eine immer wieder genannte Limitation der Verknüpfung verschiedener Datenquellen in Big-Data-Anwendungen im Gesundheitswesen ist der besondere Schutz, den personenbezogene Gesundheitsdaten genießen. Dies gilt sowohl für Routinedaten als auch für die in Forschungsprojekten entstehenden Daten (White House 2014, S. 29; March et al. 2014).

#### *Welche Regeln beeinflusst Big Data?*

Auswertungen von Routinedaten werden genutzt, um sowohl Entscheidungsprozesse über generelle Regeln, Gesetze und Verordnungen zu beeinflussen als auch um spezifische Regeln zu setzen, die z. B. die Abläufe in einzelnen Kliniken steuern. Konkret werden Routinedaten in Deutschland insbesondere dafür genutzt, um Entscheidungen über die Wirksamkeit und Kosteneffizienz von Therapien und Arzneimitteln zu treffen (OECD 2015, S. 342). Big Data wird also bisher vor allem als zusätzlicher Input in bestehenden Entscheidungsprozessen genutzt.

Allerdings werden große zukünftige Möglichkeiten oft im Bereich der prädiktiven Medizin gesehen. Durch die Kombination von klassischen Gesundheitsdaten, aus denen individuelle Krankheits- und Therapieverläufe hervorgehen, mit Verhaltensdaten und genetischen Informationen sollen Dispositionen für Krankheiten individuell vorhergesagt werden, bereits bevor sie ausbrechen. Diese Zusammenführung verschiedener Datenströme findet gegenwärtig noch nicht flächendeckend, sondern erst im Rahmen einzelner Forschungsprojekte statt. In der Regel versuchen Studien für einzelne, sehr spezielle Erkrankungen Faktoren zu bestimmen, die für das Auftreten der Krankheit verantwortlich sind. Oder es wird nach Faktoren gesucht, die einen erfolgreichen Therapieverlauf sicherstellen (Raghupathi und Raghupathi 2014). Hier deutet sich ein umfassenderer Einfluss von Big Data auf Entscheidungsprozesse im Gesundheitswesen an, indem Therapieentscheidungen auch auf andere Faktoren als die ärztliche Diagnose bereits bestehender Erkrankungen gestützt werden.

#### *Wie weitgehend ist der Einfluss von Big Data?*

Aktuell bleiben im Gesundheitsbereich große Entscheidungsspielräume für menschliche Entscheiderinnen und Entscheider. Bisher werden Entscheidungen für oder gegen bestimmte Therapien oder Entscheidungen über strukturelle Veränderungen im Gesundheitssystem nicht durch Big Data determiniert. Die größten Spielräume für Big-Data-getriebene Regulierung bieten private Gesundheitssysteme, da nur dort – anders als in Solidarsystemen – in nennenswertem Umfang eine Individualisierung von Gesundheitsleistungen möglich ist.

### 3.5.3 Fazit und Ausblick

Zur Frage, welche Aspekte von Big Data regulierend wirken, lässt sich bilanzieren, dass Big Data zwar primär zu kommerziellen Zwecken eingesetzt wird, jedoch zunehmend auch Anwendung durch staatliche Akteurinnen und Akteure oder in deren Auftrag findet. Ein Blick auf die entsprechenden Studien macht dabei deutlich, dass das Gros der Anwendungen bereichsspezifische Daten einsetzt. Das liegt unter anderem am z. T. strengen regulatorischen Rahmen. Allerdings gibt es in Forschungs- und Pilotprojekten auch Experimente mit unspezifischen Daten. Es fehlt jedoch an vergleichenden Analysen, die darüber Auskunft geben, welcher Mehrwert durch die Hinzunahme unspezifischer Datenquellen im Vergleich zu den bisher verwendeten Quellen entsteht. Ob Regulierung durch Big Data also effizienter und effektiver wird, wie das Versprechen lautet, kann man derzeit nicht beantworten (Christin et al. 2016, S. 2). Hierzu bedarf es empirischer Forschung, die die Versprechen der Big-Data-Industrie und ihre positiven (Selbst-) Evaluationen überprüft.

Mit Blick darauf, in welcher Art Big Data gesellschaftlichen Erwartungen durch Regeln Ausdruck verleiht, scheint es, dass Big Data weitgehend dafür eingesetzt wird, konkrete Regeln zu setzen und zu implementieren. Der gemeinsame Nenner Big-Data-basierter Regelsetzung ist es, Regeln stärker zu differenzieren: nach geografischen Gebieten oder nach Subpopulationen. Big Data wird so für die Kontrolle normgerechten Verhaltens von Bürgerinnen und Bürgern eingesetzt. Die entsprechende Praxis von Überwachung, Disziplinierung und *social sorting* wird bereits kritisch durch sozialwissenschaftliche Forschung begleitet. Wenig ist hingegen darüber bekannt, in welchem Ausmaß Big Data die Regeln der Regelsetzung beeinflusst: Welche Rolle spielen die Möglichkeiten und Grenzen von Big Data dafür, wie Entscheidungen in Politik und Verwaltung zustande kommen? Deutlich ist, dass neue Akteurinnen und Akteure und Professionen an Regulierung beteiligt sind: jene, die die Daten sammeln, mit ihnen handeln, sie aufbereiten, auswerten und interpretieren. Welche Macht sie innehaben und in welchem Verhältnis sie zu anderen Regulierungsakteurinnen und -akteuren stehen, bleibt zu spezifizieren: Stimmt die These, dass den Programmierinnen und Programmierern und den Anwenderinnen und Anwendern, die die errechneten Befunde interpretieren, große Kompetenzen übertragen werden? Oder setzen sie eng gesetzte Vorgaben von Verwaltungsjuristen um?

Mit Blick auf die Verbindlichkeit von Regeln, die durch Big-Data-basierte Verfahren gesetzt und implementiert werden, zeigt sich, dass die Erwartung, dass Big Data im staatlichen Bereich zu einer Renaissance kybernetischer Steuerung führt, überzogen ist. Beispiele für eine vollautomatisierte algorithmische Steuerung konnten wir nicht finden, stets gab es einen „menschlichen Filter“. Vielmehr dienen Big-Data-basierte Analysen staatlichen Akteurinnen und Akteuren oder

Bürgerinnen und Bürgern als Information oder Empfehlung und es verbleiben Handlungsspielräume bei der Umsetzung. Wie groß diese formalen Entscheidungsspielräume aber tatsächlich sind, ist offen. Hier tut sich eine große Forschungslücke auf: Welche Autorität üben die Ergebnisse von Big-Data-basierten Verfahren in der Praxis aus? Wie verhalten sich (staatliche) Regulierung und (soziale) Selbstregulierung zueinander? In welchen Bereichen sollten Spielräume verbleiben und wo nicht, etwa bei richterlichen Urteilen?<sup>59</sup>

Die Beispiele aus der Sicherheits- und Gesundheitspolitik machen deutlich, dass beim Einsatz von Big Data für öffentliche Güter stets Abwägungen zwischen Gemeinwohl und individuellen Rechten getroffen werden müssen: öffentliche Sicherheit versus Freiheitsrechte, Gesundheit versus Schutz vor Diskriminierung etc. Dabei formulieren Sozialwissenschaftlerinnen und -wissenschaftler besondere Anforderungen, wenn Big Data im öffentlichen Sektor eingesetzt wird: Transparenz, Qualität des Verfahrens, Datensicherheit, Datenschutz, soziale Verträglichkeit und viele weitere. Dies gilt insbesondere dort, wo staatlichen Akteurinnen und Akteuren sowie Bürgerinnen und Bürgern geringe Handlungsspielräume bei der Regelumsetzung verbleiben und wo private Akteurinnen und Akteuren die Regelsetzung auf der Grundlage von Big Data (mit) verantworten. Unternehmen scheinen in vielen Fällen, in denen Big Data für Regulierung eingesetzt wird, in zentraler Funktion beteiligt zu sein. Entsprechende Anforderungen an verschiedene Akteurinnen und Akteuren in den verschiedenen Anwendungsbereichen zu formulieren ist unter anderem auch Aufgabe sozialwissenschaftlicher Forschung.

---

## 3.6 Kulturell-kognitiv hergestellte Erwartungen an Big Data

Ulrike Klinger und Christian Pentzold

### 3.6.1 Was erwarten wir, wenn wir von Big Data reden?

Der Gebrauch des Begriffs Big Data geht einher mit einer Reihe an Erwartungen und Befürchtungen, so etwa hinsichtlich der technologischen Machbarkeit, der politischen Signifikanz, dem kommerziellen Wert oder der kulturellen Implikationen des Sammelns und Auswertens digitaler Datenmengen. Hierbei dient Big

---

59 Zu den Vor- und Nachteilen des Einsatzes von algorithmisch basierten Regeln und Entscheidungen im Gerichtswesen siehe etwa Christin et al. (2015) und Angwin et al. (2016).

Data nicht so sehr als exakter wissenschaftlicher Terminus, sondern eher als mehrdeutige Metapher und argumentative Ressource, die in verschiedenen und nicht notwendigerweise komplementären Begründungszusammenhängen mobilisiert werden kann. Daraus ergeben sich eine Reihe grundsätzlicher Fragen, so etwa was Big Data ist und was es sein kann, wer Big Data verstehen oder erzeugen kann, wer Big Data kontrollieren kann oder wer bestimmt, was Big Data ist bzw. sein soll.

Die Implikationen von Big Data für politische Akteurinnen und Akteure, politische Prozesse und politische Strukturen müssen nicht zwangsläufig in die Dichotomie utopischer versus dystopischer Szenarien zerfallen, entlang derer der Diskurs über Big Data strukturiert ist. Der Einsatz algorithmisch kontrollierter Analysen großer Datenmengen wird nicht nur positive oder negative Konsequenzen nach sich ziehen, wohl aber werden entsprechend einseitige Vorstellungen mobilisiert und instrumentalisiert. Die Etablierung des Begriffs Big Data und der damit benannten Problemlage in öffentlichen Diskursen zeigt an, dass verschiedene, auch gegensätzliche Bestrebungen bestehen, Big Data gesellschaftlich relevant zu machen. Entsprechend ist zu fragen und empirisch zu untersuchen, wie Big Data in öffentlichen Debatten gerahmt wird und welche Akteurinnen und Akteure mit welchen Interessenlagen hierbei eine Rolle spielen (vgl. Kitchin 2014a; Beer 2016).

Der Blick auf die kognitiv-kulturellen Erwartungen an Big Data verweist auf die instrumentellen Funktionen dieser Polarisierungen. Erwartungen oder Befürchtungen werden genau so formuliert, um den Status Quo bestehender Akteurskonstellationen, Prozesse oder Strukturen zu rechtfertigen oder um deren notwendige Transformation einzufordern. Verweise auf die nutzbringende oder unheilvolle Wirksamkeit von Big Data werden somit im Rahmen normativer Ordnungsvorstellungen angeführt, die wiederum rekursiv auf die Gestaltung und Relevanzsetzung von Big-Data-Technologien und daraus abgeleiteter Einsichten und Erfordernisse zurückwirken. In diesem Zusammenhang erscheint dann Marcuses Hinweis auf die der Technik innewohnende Ideologie immer noch zeitgemäß: „Bestimmte Zwecke und Interessen der Herrschaft sind nicht erst „nachträglich“ und von außen der Technik oktroyiert – sie gehen schon in die Konstruktion des technischen Apparats selbst ein; die Technik ist jeweils ein geschichtlich-gesellschaftliches Projekt; in ihr ist projektiert, was eine Gesellschaft und die sie beherrschenden Interessen mit den Menschen und mit den Dingen zu machen gedenken“ (Marcuse 1965, S. 179). Die Ambivalenz des Begriffs Big Data wird dadurch bedingt, dass er mit variablen Definitionen erfasst, in Bezug zu verschiedenen historischen, literarischen oder räumlichen Analogien konkretisiert und mit unterschiedlichen Bedingungen, Konsequenzen und Bewertungen in Verbindung gebracht wird (Beer 2016).

In der Konsequenz finden sich nicht zu wenige, sondern eher zu viele unterschiedliche Ansätze, um eine mehr oder minder umfassende oder bündige Definition von

Big Data vorzulegen. Sie tendieren letztlich bei einem allzu losen Wortgebrauch dazu, aus Big Data die rhetorische Phrase eines *elevator word* zu machen, das ohne weitere inhaltliche Begründung eingeworfen wird, um den *level of discourse* vermeintlich anzuheben, wie Hacking (1999, S. 21) schreibt. Dabei wird Big Data als Mittel und Ziel zum Selbstzweck – zum Phänomen, das nicht auf vorgelagerten Fragen, Hypothesen oder methodischer Innovation basiert, sondern auf technologischem Fortschritt. „Big data is thus akin to the Everest of contemporary social research. ‚Because it’s there’, so massively there, is sufficient rationale to tackle it“, konzediert Smith (2014, S. 184) für die sozialwissenschaftliche Nutzung – und verweist damit zugleich auf ein *mindset*, das auch in anderen Anwendungsszenarien virulent ist.

Grundsätzlich greift der Blick auf die kulturell-kognitiv hergestellten Erwartungen an die Unmengen von Big Data zurück auf Ansätze, welche die objektive Faktizität von Daten hinterfragen (Bowker 2013). Sie verweisen auf das stets vorhandene interpretative Moment ihres Herstellens und Darstellens. „Data need to be imagined as data to exist and to function as such“, erklären demgemäß Gitelman und Jackson (2013, S. 3), „and the imagination of data entails an interpretative base“. In dieser Hinsicht hat gerade die unbestimmte Rede von Big Data auch eine ideologische Konnotation. Nach Boyd und Crawford (2012, S. 663) impliziere sie eine gleichsam mythologische Idee: „that large data sets offer a higher form of intelligence and knowledge ... with the aura of truth, objectivity, and accuracy“.

Diese Erwartung muss indessen nicht die einzige oder auch nur eine allgemein geteilte sein, sondern vielmehr ist Big Data semantisch offen und geht einher mit verschiedenen, auch in Konflikt stehenden Vorstellungswelten, die sich etwa in Worten wie *dataverse*, *data deluge* oder *data explosion* bzw. in Slogans wie Big Data als *the new oil* oder in den Wendungen des *mining*, *dredging* oder *harvesting* von Daten manifestieren. „The implicit meanings of these words“, so Portmess und Tower (2015, S. 3), „carry suggestive implications for exploring different ways of envisioning our relationship to emerging information technologies and embody forms of thought and practice“. Folglich verweisen Puschmann und Burgess (2014) auf die fortwährenden Deutungskämpfe um die korrekte Bedeutung des Konzeptes, wobei sie bei ihrer Analyse US-amerikanischer Wirtschafts- und Computer/IT-Zeitschriften zwei vorherrschende, zueinander gegensätzliche Stimmungen abbilden können. Entweder wird hier Big Data mit einer (kaum) zu bändigenden Naturgewalt verglichen oder Big Data wird mit einem zu gebrauchenden Rohstoff bzw. Treibstoff gleichgesetzt. Allgemein gesehen bilden diese beiden Pole die grundsätzliche Dynamik von utopischen versus dystopischen Visionen ab, welche generell die Diffusion neuer Technologien zu begleiten scheinen (Feenberg 2002).

Abseits der Frage, ob der unpräzise Ausdruck Big Data durch vermeintlich exaktere Begriffe wie „Trace Data“ (Jungherr 2015) oder „Meta-Data“ (van Dijck 2014)

zu ersetzen ist, bleibt grundlegend zu klären, welche polarisierenden Perspektiven auf Big Data im öffentlichen Diskurs artikuliert werden. Erwartungen an Big Data als Technologie beziehen sich etwa auf Datensätze, zu deren Bearbeitung leistungsstarke Rechenzentren nötig sind, sowie Techniken, um Gegebenheiten in einzelne Daten zu überführen und diese maschinell zu aggregieren, zu analysieren und zu kombinieren. Mit Erwartungen an Big Data als analytische Ambition verbindet sich die Hoffnung auf präzisere Einsichten und verlässlichere Prognosen (Ruppert et al. 2013). Hingegen heben beispielsweise Erwartungen an Big Data als regulatorische Herausforderung auf die Überwachung und Kontrolle durch staatliche Institutionen und privatwirtschaftliche Unternehmen ab (Lane et al. 2014; Tufekci 2014).

### 3.6.2 (Wie) verändert Big Data Politik?

Unbeschadet der Unschärfe des Begriffes, verbinden sich mit Big Data sehr konkrete Erwartungen eines Wandels politischer Akteurskonstellationen, Strukturen und Prozesse. Die Perspektiven auf Big Data als Treiber eines Wandels, als *game changer*, verlaufen entlang der bereits beschriebenen utopischen versus dystopischen Visionen. „Big“ Data bezieht sich aber nicht nur auf die schiere Größe der Datenmengen, sondern auch auf die Vielfalt der Daten, die Geschwindigkeit, mit der sie ausgewertet werden können, und ihren ökonomischen Wert – den Schatz, der mit ihnen gehoben wird (Chang et al. 2013, S.2). Integral damit verbunden sind Algorithmen, „sets of defined steps structured to computationally process instructions/data to produce an output“ (Kitchin 2016, S.1), ohne die Big Data wertlose Datenhaufen wären, weil keine Muster erkennbar würden.

#### Big Data als Akteurin bzw. Akteur?

Die Idee, dass Technologien handeln können, also ebenso wie Individuen oder Organisationen als Akteurinnen und Akteure angesehen werden können, ist in der Sozialwissenschaft nicht neu. Ein Beispiel dafür ist die Actor-Network-Theory, die generell von *actants* ausgeht – unabhängig davon, ob es sich dabei um Menschen oder Maschinen handelt (vgl. Latour 2005). Dass indessen Daten autonom agieren, ist bislang nicht explizit argumentiert worden – implizit schwingt diese Vorstellung von und Erwartung an Big Data aber in vielen Bereichen mit. Die Kernfrage hierbei lautet, ob Daten und, damit verbunden, Algorithmen eine Form von *agency* zukommt, die mit menschlicher Handlungsbefähigung und Handlungsmacht vergleichbar bzw. auf Augenhöhe ist – und ob diese Erwartung eher Anlass zu Besorgnis oder Zuversicht ist.

Wenn wir mit Emirbayer und Mische (1998) argumentieren, dass *agency* immer aus drei Elementen besteht, die Handlungen durch (1) Iterationen mit der Vergangenheit, durch (2) Projektionen mit der Zukunft und durch (3) praktische Evaluationen mit den Ambivalenzen und Problemen der Gegenwart verbinden, wird deutlich, wieso Daten und Algorithmen kein Handlungsvermögen haben, wie es Menschen zugesprochen wird. Auf der Basis von großen Datensätzen filtern, selektieren, aggregieren, bewerten, empfehlen, überwachen, verteilen und prognostizieren Algorithmen politisches Handeln und Entscheiden (vgl. Just und Latzer 2016). Ihr Agieren besteht dabei zu einem großen Teil aus Iterationen, teilweise auch Projektionen – während praktische Evaluationen und der Umgang mit ambivalenten Situationen (noch) eine signifikante Hürde darstellen. Insofern können Daten und Algorithmen zwar als Akteurin bzw. Akteur verstanden werden und ihnen kann Handlungsmacht zugeschrieben werden, jedoch nur im Sinne einer direkt übertragenen (*imposed*) und delegierten *agency*, (noch) nicht aber als eigenständige und reflexive (Um-)Gestaltung von Wirklichkeit (vgl. Mitcham 2014; Just und Latzer 2016). *Imposed agency* bedeutet, dass Technologien nur in einem eng gesteckten Rahmen ausführen, wozu ihre Entwicklerinnen und Entwickler sie anweisen. Delegierte *agency* ist weiter gefasst und betrifft zum Beispiel selbstlernende Algorithmen, die sich in einem gewissen, aber auch vordefinierten Rahmen autonom weiter entwickeln können. Ein Beispiel dafür wäre Googles Go-Algorithmus, der sich selbstlernend dazu befähigte, den amtierenden Weltmeister des Strategiespiels Go zu schlagen – auf der Basis umfangreicher, weitgehend autonom ausgeführter strategischer Berechnungen. Dennoch kann auch dieser Algorithmus nicht über seine vorprogrammierte Grundstruktur hinaus, indem er etwa lieber Schach lernt als Go. Emmanuel Mogenet, Forschungsleiter bei Google, beschreibt es so: „Ich glaube nicht, dass ich in meiner Lebenszeit Computer sehen werde, die mehr können als Menschen. Der Computer, der Go spielen kann, kann nur Go spielen. Er kann keine Katzen erkennen, er kann kein industrielles Rechenzentrum optimieren. Das sind sehr kleine, abgesteckte Bereiche, in denen die Computer etwas leisten können, was der Mensch nicht kann“ (Theile 2016).

In den Debatten über die Handlungsanweisung bzw. -delegation an Big Data können positiv wie negativ konnotierte Beschwörungen ihrer vermeintlichen Autonomie letztlich den Blick verstellen auf die eigentlichen Akteurinnen und Akteure hinter den Benutzeroberflächen und damit auch auf die der Technologie innewohnenden Machtstrukturen. Kulturelle und kognitive Erwartungen an Big Data richten sich letztlich nicht nur an die technischen Tools selbst, sondern in erster Linie eben auch an diejenigen, die sie bereitstellen, entwickeln und denen Daten und Technologien gehören. In diesem Sinne agierten etwa im US-Wahlkampf 2016 die Entscheidungsträgerinnen und -träger und Programmiererinnen und

Programmierer von Plattformen wie Facebook und Twitter selbst als „political employees“ (Agho 2015), die beratend für politische Kandidatinnen und Kandidaten und Wahlkampfteams auftraten, wobei ihre Dienstleistungen zum einen begrüßt wurden, um Wählerinnen und Wähler gezielt anzusprechen und Botschaften zielgruppenorientiert zu fabrizieren und zu adressieren. Zum anderen zielt die Kritik an diesen neuen Agenten der Intermediation zwischen Politikerinnen und Politikern, politischen Botschaften und Bürgerinnen und Bürgern insbesondere auf vermutete Manipulationen und die Besorgnis, dass die Plattformen ein verzerrtes Meinungsklima beförderten.

Folglich besteht in diesem Bereich dringender Forschungsbedarf. So ist zu fragen, welche Geschäftsmodelle und Weltbilder das Handeln jener leiten, denen Datensätze gehören und die sie auswerten und interpretieren (vgl. Mager 2012)? Algorithmen, die Big Data überhaupt erst zu einer Informationsressource machen, sind in dieser Hinsicht Manifestationen sozialer Werte und Normen, keine neutralen „schwarzen Kisten“: „Algorithmic systems are not standalone little boxes,“ so erklärt entsprechend Seaver (2013, S. 10), „but massive, networked ones with hundreds of hands reaching into them, tweaking and tuning, swapping out parts and experiencing with new arrangements ( ) we need to examine the logic that guides the hands“.

### **Big Data und politische Prozesse**

Die Verfügbarkeit und Auswertung großer Datensätze, das Entdecken neuer Muster und Verfolgen sozialer Prozesse in Echtzeit kann auch politische Prozesse beeinflussen. Wie hoch die Erwartungen an Big Data in diesem Bereich sind, wird besonders dann deutlich, wenn Akteurinnen und Akteure auf ihren Einsatz verzichten. Gilt *Data Mining* schon als elementarer Bestandteil moderner Wahlkampfkommunikation, so wurde die Beobachtung mit Erstaunen quittiert, Donald Trump habe die US-Präsidentschaft scheinbar ohne den Einsatz von Big Data gewonnen. David Karpf (2016) etwa bemerkt, dass Trump im Wahlkampf mehr Geld für bedruckte Basecaps („Make America Great Again“) als für Umfragen ausgegeben habe, und die extensive Liste von Email-Adressen potenzieller Wählerinnen und Wähler tatsächlich für nichts Innovativeres als das Versenden von Emails verwendet habe. Der Sieg Trumps ohne jegliche datengestützte Kampagnenarbeit wie sie in den letzten zehn Jahren üblich wurde, stelle somit ein ganzes System infrage: „A Trump win means, in effect, that decades of research designed to organize and influence voters can be overpowered by a chaotic, from-the-gut performance“ (Karpf 2016). Indes scheint der Schluss zu voreilig, die Kampagne von Trump habe nicht auch auf die algorithmisch programmierte Auswertung von sehr großen Mengen Daten aus sozialen Netzwerken gesetzt – und damit letztlich doch einen wahlentscheidenden Vorsprung errungen (Krogerus und Grassegger 2016).

Big Data stellt in dieser Lesart ein wirkungsvolles Instrument dar, Wahlkämpfe als kampagnenförmige politische Prozesse effizienter, berechenbarer und rationaler zu führen. Nicht an alle Haustüren müssten Politikerinnen und Politiker klopfen, sondern nur bei jenen Wählerinnen und Wählern, die datenbasiert für eine persönliche Ansprache auserkoren wurden – weil sie zu einer kritischen Kohorte gehören, als Multiplikatoren identifiziert wurden oder anderen wahlkampfrelevanten Mustern entsprechen (vgl. Nickerson und Rogers 2013; Hersh 2015). Durch Daten würden Wahlkämpfe planbarer und kontrollierbarer – so zumindest die durch die Siege der datenintensiv geführten Kampagnen von Barack Obama bestätigte Erwartung.

Auch im Bereich der Politikgestaltung, des *policy making*, sind mit Big Data sowohl positive als auch negative Erwartungen verbunden. Schon immer verfügten Regierungen über große Datensätze (z. B. Zensusdaten), die aber weder in Echtzeit erhoben und ausgewertet wurden, noch zu einer ergebnisoffenen Suche nach Problemmustern genutzt wurden (vgl. Scott 1999). Dies ist aber gerade das Novum der mit Big Data einhergehenden analytischen Ambitionen: über die algorithmische Auswertung von Datenmengen lassen sich Muster identifizieren, nach denen man zunächst nicht explizit gesucht hatte. Daraus ergeben sich Konsequenzen für die Erwartungen an Big Data. Denken wir an die Bekämpfung von Kriminalität, die Fahndung nach Terroristinnen und Terroristen oder auch Klimapolitik, zeigt sich das Potenzial datengestützter Problemdefinition und Problemlösung. Nicht durch aggregierte Forderungen diverser Anspruchsgruppen, sondern durch Muster oder Irregularitäten in Datensätzen werden Themen in den *policy cycle* eingespeist. Problematisch daran ist die Frage, wer eine entsprechende *voice* hat und für wen Politik gemacht wird: Kommt der Input in den Gesetzgebungsprozess von betroffenen gesellschaftlichen Gruppierungen oder von Datenexpertinnen und -experten? Da Big Data in erster Linie Korrelationen offenlegt, nicht aber Kausalitäten, stellt sich zudem die Frage, wer legitimerweise statistische Zusammenhänge in Ursache-Wirkung-Relationen übersetzen und ihnen gesellschaftliche Relevanz zuweisen kann und soll. Wird, so ist in künftiger empirischer Forschung zu fragen, Gesetzgebung tatsächlich responsiver, wenngleich Probleme bearbeitet werden, die von keiner gesellschaftlichen Gruppe artikuliert, sondern in Datensätzen identifiziert wurden?

Big Data legt zudem nicht nur bislang versteckte Muster im Handeln von Individuen offen, sondern ermöglicht auch Voraussagen auf zukünftiges Verhalten. Während Unternehmen darin große Potenziale für fokussierte Werbung sehen (wenn z. B. Schwangerschaften der Kundinnen vom Einzelhandel frühzeitig erkannt werden), ist für staatliche Akteurinnen und Akteure die Vorhersagbarkeit von Rechtsverletzungen interessant – sowohl in der Polizeiarbeit (*Predictive Policing*) als auch in Fragen der Sicherheit. Dadurch verschiebt sich möglicherweise der Gestaltungsbereich politischer Prozesse, indem nicht nur individuelles und kollektives

Handeln reguliert, sondern auch zukünftige, noch ungeschehene Entscheidungen sanktionierbar werden. Besorgniserregend scheinen diese Entwicklungen nicht zuletzt aus den Perspektiven von staatlicher Überwachung (*surveillance*), Datenschutz und damit verbundenen rechtsstaatlichen Aspekten.

Wir sehen also sowohl optimistische Erwartungen an Big Data, die auf eine gesteigerte Effizienz und vermeintlich neutrale Informiertheit – und somit höhere Rationalität – politischer Prozesse abzielen, als auch pessimistische Erwartungen, für die Big Data vor allem eine unentrinnbare Überwachung und Kontrolle bedeutet.

### **Big Data und politische Strukturen**

Auch wenn Big Data und Algorithmen ein relativ neues Forschungsfeld konstituieren, lassen sie sich doch mit etablierten sozialwissenschaftlichen Theorien und Konzepten analysieren. So ist überzeugend argumentiert worden, dass diese Technologien als Institutionen verstanden werden können (vgl. Katzenbach 2012; Napoli 2014), da sie individuelles und kollektives Handeln sowohl begrenzen als auch ermöglichen. Dabei ist es ganz unerheblich, ob Individuen die Funktionsweise von institutionell aufgestellten und institutionell wirksamen Algorithmen und Daten nachvollziehen können. „Wenn der Mensch den Sinn oder die objektive Wirkung nicht begreift, wird ihre objektive Wirklichkeit nicht geringer“, führten bereits Berger und Luckmann (1980 [1966], S. 64) für Institutionen allgemein aus.

Die Frage mit Blick auf politische Strukturen aber ist, wie diese Technologien als Institutionen selbst Struktur sein und bestehende Strukturen ersetzen oder verändern können. Damit verbunden wird die sowohl zustimmend als auch ablehnend aufgenommene Erwartung einer reflexiven datenbasierten „Verwissenschaftlichung“ politischer Systeme in einer sozialen Wirklichkeit, die sich über Daten praktisch selbst abbildet und in der soziale Fakten bzw. Entwicklungen objektiv modelliert und auf die Gestaltung von politischen Strukturen zurückgespiegelt werden können. Ein intermediäres System von Akteuren, die zwischen Herrschaftsträgerinnen und -trägern und Bürgerinnen und Bürgern vermitteln, wäre damit nicht mehr nötig. Journalistische Medien, Verbände, Vereine, Parteien müssten nicht mehr vermitteln, informieren, erklären oder fordern, weil Daten selbst Themen und Probleme identifizieren sowie die Effizienz und Effektivität politischen Outputs messbar machen. Diese Vorstellung basiert auf einer sehr vereinfachten Vorstellung gesellschaftlicher Zusammenhänge, die soziales Leben als Produkt, nicht als Prozess versteht und die Aggregate über soziale Praktiken stellt.

Ein Beispiel für die Materialisierung solcher Effizienzerwartungen sind die sogenannten *smart cities*. Unter diesem Begriff wird die Implementierung von datenbasierten Technologien in die regulativen politischen Strukturen von Städten diskutiert, in denen umfassende Datensätze in Echtzeit generiert werden, etwa

durch automatisiertes Monitoring, beispielsweise über lesbare *Smart Cards* im öffentlichen Nahverkehr statt Papiertickets, Überwachungskameras, das Auslesen von Kfz-Kennzeichen durch Mautkameras oder RFID-Chips in Mülltonnen, die die Entleerungen überwachen (vgl. ausführlich Kitchin 2014b). Hinzukommen „smarte“ Gegenstände wie Straßenlaternen, die über Wifi verbunden sind, über Bewegungsmelder gesteuert werden und Videoaufnahmen machen oder Lüftungs- und Heizsysteme, welche die Anwesenheit im Gebäude registrieren und darauf reagieren. Solche Daten werden zum einen genutzt, um Prozesse innerhalb der Stadt zu steuern, etwa zur Steuerung von Verkehrsströmen in Echtzeit. Das bedeutet zum anderen aber auch, dass eine klassische Steuerung über politische Strukturen an Bedeutung verliert und technokratische Weisungs- und Entscheidungsketten zunehmen, die mit einem Machtgewinn durch *private-public partnerships* für diejenigen Akteurinnen und Akteure einhergehen, welche Daten erheben, speichern und auswerten können. Dies führt aber, wie Kitchin (2014b) bemerkt, nicht zwangsläufig zu besseren politischen Ergebnissen: „Technological solutions on their own are not going to solve the deep rooted structural problems in cities as they do not address their root causes. Rather they only enable the more efficient management of the manifestations of those problems.“

Neben den positiven Erwartungen bezüglich einer datenbasierten Politik in *smart cities* finden sich auch hier dystopische Vorstellungen einer „gehackten“ Stadt, die von Cyberattacken aus dem In- und Ausland heimgesucht wird – die politische Kontrolle über die automatisierten Bereiche der Stadt also nicht mehr ausschließlich in den Händen der politisch legitimierten Akteurinnen und Akteure liegt (Kitchin et al. 2015).

### 3.6.3 Ausblick

Aus diesen Überlegungen folgt, dass ein zentrales Desiderat politikwissenschaftlicher Beschäftigung mit Big Data kritische Reflektion sein muss. Big Data *ist* nicht, sondern wird hergestellt: zum einen durch die Arbeitsprozesse und Technologien von Datenexpertinnen und -experten, zum anderen durch Bedeutungszuschreibungen in fachlichen und öffentlichen Diskursen. Die Technologien, die Big Data ermöglichen, sind potenziell disruptiv, das heißt, sie verändern sehr wahrscheinlich relevante gesellschaftliche Strukturen und Prozesse – nicht nur im Bereich der Politik. Es ist die Aufgabe der Wissenschaft, nicht nur die konkreten Bereiche zu identifizieren, die dadurch transformiert werden, sondern über die dichotomen Erwartungen im öffentlichen Diskurs hinaus kritisch zu fragen, welche konkreten Nutzen und Kosten von Big Data für demokratisch verfasste Gemeinwesen zu erwarten sind.

Dies betrifft nicht nur, aber auch ganz zentral Fragen der Machtstrukturen, der Funktionslogik intermediärer Akteurinnen und Akteure, sowie der Herstellung und Legitimation von Politik.

---

### **3.7 Ist Big Data fair? Normativ hergestellte Erwartungen an Big Data**

Ingrid Schneider und Lena Ulbricht

In welcher Art sich der Umgang mit Big Data entwickelt, wird auch durch Normen, Werte und Ziele beeinflusst. Aus normativer Hinsicht stellen sich der Sozialwissenschaft somit folgende Fragen: Für welche Ziele und in welcher Ausprägung gilt die Anwendung von Big Data als legitim? Welche Akteurinnen und Akteure dürfen oder sollen Big Data nutzen? An welchen Werten und Zielen muss Big Data sich messen lassen? Welche gesellschaftlichen Erwartungen bilden sich heraus und wie können sozialwissenschaftliche Antworten auf diese Fragen lauten?

Es gibt Werte und Ziele, die die Ausweitung von Big Data auf immer mehr Lebensbereiche vorantreiben, etwa ein Streben nach Effizienz und Effektivität bei Entscheidungen. Ebenso gibt es Werte, die die Anwendung von Big Data auf immer weitere Lebensbereiche begrenzen (wollen), wie etwa ethische Ansprüche an die Wahrung menschlicher Autonomie (Zuboff 2015) und gesellschaftlicher Solidarität (Rouvroy 2016). Ein Wert, der die öffentliche Debatte über Big Data in den letzten Jahren zunehmend mitprägt, ist jener der Fairness: Die Befürchtung, dass Big Data zu neuen und/oder massiven Diskriminierungen führen könnte, scheint ein zentraler Faktor für die Entwicklung von Big Data zu werden.

Profilbildung und Erkenntnisse über sensible personenbezogene Informationen werden in Deutschland bisher vor allem unter der Thematik von Privatsphäre und Datenschutz oder Überwachung diskutiert. In den USA richtet sich die Debatte stärker auf die Frage, welche gesellschaftlichen Ein- und Ausschlusspotenziale mit Big-Data-Analytik verbunden sind. Diese Debatte fand etwa Ausdruck in zwei Berichten des Weißen Hauses 2014 (White House 2014) und 2016 (White House 2016) und der Federal Trade Commission (FTC), die für Verbraucherschutz zuständig ist (2016). Diese Studien beruhen auf Sekundärliteratur, Befragungen, *stakeholder meetings* und wissenschaftlichen Tagungen und haben somit erste Belege über mögliche diskriminierende Effekte von Big Data und Empfehlungen für die Prävention geliefert. Während einige analytische Konzepte durchaus übertragbar erscheinen, sind die Rechtslage sowie andere Faktoren und Kontexte nicht unmittelbar auf Eu-

ropa und Deutschland anwendbar.<sup>60</sup> Dennoch erlauben sie eine erste Annäherung, die in politikwissenschaftlichen Analysen weiter zu verfolgen sein wird.

### 3.7.1 Forschungsstand und Bedarfe

Obwohl der öffentliche Diskurs Big Data häufig als Risiko für Diskriminierungen darstellt, sind wissenschaftliche Studien, die dem auf den Grund gehen, noch spärlich gesät – selbst in den USA (dies betonen etwa Desmarais und Singh 2013 sowie Christin et al. 2016). Die Berichte des Weißen Hauses und der FTC nennen nur eine Handvoll empirischer Analysen, die sich alle auf die USA beziehen und noch kein eindeutiges Bild zeichnen. Zwar finden sich Hinweise darauf, dass bestimmte Gruppen durch Big-Data-bezogene Praktiken benachteiligt werden, wie etwa *blacks*<sup>61</sup> (Angwin et al. 2016; Lowry und MacPherson 1988; Sweeney 2013; Brevoort et al. 2015), *hispanics* (Brevoort et al. 2015), Frauen (Datta et al. 2015, Lowry und MacPherson 1988), Personen, die mutmaßlich unter Alkoholproblemen leiden (Datta et al. 2015), Personen mit geringem Einkommen (Brevoort et al. 2015) sowie Mac-Nutzerinnen und -Nutzer (White 2016). Doch werden nur wenige Anwendungsbereiche von Big Data untersucht: Online-Werbung (Datta et al. 2015; Sweeney 2013), Kreditscores (Brevoort et al. 2015), das Strafmaß bei Verurteilungen (Angwin et al. 2016; Christin et al. 2015), Preisdiskriminierung (White 2016) und Hochschulzulassungsverfahren (Lowry und MacPherson 1988). Zudem sind nicht alle Ungleichbehandlungen als Diskriminierung zu verstehen, selbst wenn *race* das Unterscheidungsmerkmal ist. Bei Skeem und Lowenkamp zeigt sich etwa, dass ein Großteil der unterschiedlichen Risikoscores zwischen weißen und schwarzen Angeklagten auf deren unterschiedliche strafrechtliche Vorgeschichte zurückgeht (2016).

---

60 Diese Studien stützen sich auf die US-Antidiskriminierungsgesetzgebung, die im Zuge von Bürgerrechtsbewegungen und der *affirmative action* erstritten wurde. Die Situation in den USA ist allenfalls bedingt auf Europa übertragbar. Denn etwa die Kategorisierung von *race* bzw. ethnischer Herkunft (*Caucasian*, *African-American*, *Asian-American* etc.) findet hierzulande – nicht zuletzt aus historischen Gründen – weder Eingang in Kategorien der Staatsbürgerschaft noch bilden sie die Basis von sozialen Anspruchsrechten (*protected groups*, Quotensysteme etc.). Gleichwohl lassen sich durchaus einige der Befunde ggf. übertragen und sie bieten Ansatzpunkte, die für analytische und empirische Studien fruchtbar gemacht werden können, wie weiter unten deutlich wird.

61 In diesem Artikel werden die Bezeichnungen für ethnische oder rassenbezogene Zuordnung verwendet, die die jeweiligen Autoren der Studien, die hier zitiert werden, wählen.

Die hier untersuchten Studien erheben keinen Anspruch auf Vollständigkeit, bekräftigen aber, dass die Antwort auf die Frage, ob Big Data zu Diskriminierung führt, nicht leicht zu geben ist. Es gibt zahlreiche methodische Hürden: Ein zentrales Problem ist die Identifizierung von Gruppen, die als schützenswert gelten (sollten). Um festzustellen, ob beispielsweise Personen mit geringem Einkommen, Alleinerziehende oder bestimmte ethnische Gruppen benachteiligt werden, muss man diese überhaupt als solche erkennen. Doch fehlt es an geeigneten Indikatoren. In der Studie über Kreditscores wurden einkommensschwache Personen etwa anhand des Wohnortes geschätzt (Brevoort et al. 2015). Um rassistische Diskriminierung bei Werbeanzeigen nachzuweisen, muss Sweeney Vornamen als „weiß“ oder „schwarz“ charakterisieren. Sie stützt sich hierbei auf Informationen aus Geburtsregistern (Sweeney 2013). Lediglich die Arbeit von Angwin und Larson et al. kann mit Daten arbeiten, die über die ethnische Zugehörigkeit der Personen direkt Auskunft geben (Larson et al. 2016). Die meisten Studien stehen allerdings vor der Herausforderung, mit mehr oder weniger präzisen Indikatoren zu arbeiten. In Deutschland werden sensible Daten wie etwa über ethnische Herkunft in der amtlichen Statistik entweder gar nicht erhoben oder sind öffentlich nicht verfügbar, sondern liegen ausschließlich den zuständigen Behörden vor, etwa Sozial- oder Gesundheitsämtern. In den USA ist dies zum Teil anders: Die Studie von Angwin, Larson und ihren Co-Autoren beruht auf ausführlichen Dossiers der Strafverfolgungsbehörden, die über die strafrechtliche Vorgeschichte von Individuen Auskunft geben, eine Kategorie *race* umfassen und öffentlich über das Internet zugänglich sind. Ein weiteres Problem ist mangelnde Transparenz: Viele Big-Data-Anwendungen werden von privaten oder staatlichen Akteurinnen und Akteuren kontrolliert, die mit Verweis auf Geschäftsgeheimnisse oder die innere Sicherheit nicht genügend Daten und Informationen bereitstellen, die belastbare wissenschaftliche Erkenntnisse erlauben würden. Wenn Sweeney also feststellt, dass die Eingabe in Suchmaschinen einer afroamerikanischen Person eher mit stigmatisierenden Werbeanzeigen einhergeht als die Eingabe des Namens einer weißen Person, weiß sie doch nicht, ob die Gründe bei den Anbieterinnen und Anbietern von Werbeanzeigen, bei den Webseiten, die Werbeanzeigen schalten, oder in Googles System der Vergabe von Anzeigepätzen liegen (2013). Dies herauszufinden ist nur diesen drei Akteurinnen bzw. Akteuren möglich. Die Öffentlichkeit ist in dieser Hinsicht auf die Bereitschaft der Unternehmen (oder staatlichen Behörden) angewiesen, die Big Data anwenden, um mögliche diskriminierende Effekte zu untersuchen und ihre Forschungsergebnisse der Öffentlichkeit zur Verfügung zu stellen. Forschung, die durch öffentliche Gelder oder durch gemeinnützige Organisationen finanziert wird, kann bestenfalls diskriminierende Effekte nachweisen – die darunter liegenden Mechanismen können sie selten untersuchen. Auch wenn

Erkenntnisse über konkrete Diskriminierungsfälle durch Big Data rar sind, gibt es doch Thesen darüber, wie Diskriminierung durch Big Data entsteht.

### 3.7.2 Differenzierung oder Diskriminierung?

Dabei muss man zuerst festhalten, dass es sich bei den Ergebnissen von Big-Data-Analysen ganz allgemein um Unterscheidungen handelt: Segmentierungen, Stratifizierungen, Rankings etc., denen eine Bedeutung zugeschrieben wird. Sofern es auf Grundlage dieser Unterscheidungen zu Bewertungen in Form von Rangfolgen, Präferenzen, Ein- und Ausschlüssen kommt, stellt sich die Frage, ob es sich dabei um *legitime Differenzierungen* handelt oder ob Differenzierung in Abwertung, Stigmatisierung oder (*illegitime*) *Diskriminierung*<sup>62</sup> umschlägt. Dies gilt etwa dann, wenn Individuen aufgrund der Zugehörigkeit zu einer sozialen Gruppe bestimmte Zugangsmöglichkeiten verwehrt werden, sie in Rangordnungen nach unten fallen oder sie Preisdifferenzierungen erfahren. Die Grenze zwischen legitimer Differenzierung und illegitimer Diskriminierung ist kontextspezifisch und Resultat von gesellschaftlichen Aushandlungsprozessen (siehe etwa White House 2016<sup>63</sup> und Kimmich und Schahadat 2016).

Eine Erwartung an Big Data liegt darin, dass algorithmische Bewertungen anhand großer Datenmengen objektivere Urteile zulassen als beispielsweise subjektive Bewertungen (White House 2016, S. 10; Boyd und Crawford 2012, S. 663, Christin et al. 2016). Allerdings mehrten sich die Hinweise darauf, dass sich sowohl aus der Datenbasis (A) als auch aus den Analysemethoden (B) diskriminierende Effekte ergeben können (White House 2016, S. 6-7), wie im Folgenden dargelegt wird.

---

62 Diskriminierung wird hier im sozialwissenschaftlichen Sinne verstanden als illegitime Ungleichbehandlung von Menschen auf der Grundlage eines oder mehrerer Kriterien. Dieses Verständnis setzt sich ab von einer (breiteren) statistischen Definition, die unter Diskriminierung jeden Unterschied zwischen Merkmalsträgerinnen und -trägern versteht, und von einer (engeren) juristischen Definition, die eng an Gesetze gebunden ist. So wird in Deutschland etwa nur im juristischen Sinne als Diskriminierung bezeichnet, was den Bedingungen entspricht, die im Allgemeinen Gleichbehandlungsgesetz (AGG) festgelegt sind. So können nur Benachteiligungen auf der Grundlage begrenzter Kriterien geahndet werden: Rasse, ethnische Herkunft, Geschlecht, Religion, Behinderung, Alter oder sexuelle Identität. Sozioökonomischer Status, politische Aktivität oder Familienstand fallen etwa nicht hinein.

63 „This report uses the term “discrimination” in a very broad sense to refer to outsized harmful impacts—whether intended or otherwise—that the design, implementation, and utilization of algorithmic systems can have on discrete communities and other groups that share certain characteristics.“ (White House 2016, S. 25).

### 3.7.3 Diskriminierung aufgrund der Dateneingabe und -aufbereitung

Eine erste Quelle für diskriminierende Effekte sind die Entscheidungen darüber, welche Arten von Daten überhaupt in die Auswertung eingehen.

Ein zentrales Problem sind Verzerrungen im Datensatz, die durch Selektion entstehen (*selection bias*) und/oder wenn die Dateneingaben zu einem Modell nicht repräsentativ für eine bestimmte Population bzw. Kohorte sind (*sample bias*). Dies führt zu Schlüssen, die bestimmte Gruppen gegenüber anderen begünstigen können (vgl. Barocas und Selbst 2016, S. 684f.; Hardt 2014). Modelle werden so entwickelt, dass sie für die Mehrheit der gewählten Population im Datensatz zutreffend sind. Dies bedeutet zugleich, dass für kleine Sub-Populationen weitaus seltener zutreffende Modelle erstellt werden. Wenn Modelle für bestimmte Gruppen (die Mehrheit) zutreffen und für andere nicht (Minderheiten), bedeutet dies eine unterschiedliche Behandlung (Barocas und Selbst 2016, S. 689, Hardt 2014). Es kann auch leicht zu sehr ungenauen Prognosen führen. So erwies sich ein Vorhersagesystem für die erneute Straffälligkeit von Verurteilten als akkurater für Männer als für Frauen – letztere wiesen zu Unrecht zu hohe *scores* auf (Larson et al. 2016). Ungenügende Daten werden besonders dann zum Problem, wenn sie exkludierend wirken. Über Personen mit niedrigerem Einkommen stehen etwa häufiger zu wenige Informationen zur Verfügung, um einen Kreditscore zu errechnen. Sie werden somit häufiger bei Finanzdienstleistungen benachteiligt, obwohl sie keinen schlechten Kreditscore aufweisen (sondern keine) (Brevoort et al. 2015). Ein anderes Beispiel sind Aufteilungen in abgegrenzte Populationen (*redlining*) – etwa nach der Postleitzahl – die Variationen in den Subpopulationen nicht beachten. Zum Beispiel wird eine Bewohnerin bzw. ein Bewohner im stigmatisierten Stadtteil Hamburg-Wilhelmsburg trotz ansonsten völlig gleicher Charakteristika mit einer Bewohnerin bzw. einem Bewohner im gutsituierten Blankenese unter Umständen längere Zeiten in der Warteschleife von Callcentern oder einen schlechteren Zinssatz bei einem Kredit in Kauf nehmen müssen (Kurz und Rieger 2012).

Eine Verzerrung entsteht auch durch unzureichende Daten, die nur bestimmte Datentypen berücksichtigen, die veraltet sind oder die Fehler aufweisen (vgl. Barocas und Selbst 2016, S. 684). Gerade für Daten, die durch Zwischenhändlerinnen und -händler, die sogenannten *data brokers* vertrieben werden, können Analytistinnen und Analysten nur noch schwer nachvollziehen, unter welchen Umständen Daten erhoben wurden und welche Aussagekraft sie besitzen. Ein weiterer Grund für Verzerrungen sind ungenaue oder fehlerhafte (d. h. falsch positive oder falsch negative) Klassifikationen (Hofstetter 2016, S. 381; Schneier 2015; Christl 2014, S. 71; Barocas und Selbst 2016, S. 680). Bei nahezu jeder Einordnung ist die Präzision der

Abgrenzung fraglich – und sie könnte theoretisch auch anders ausfallen (Barocas und Selbst 2016, S. 681). Der gleiche Datensatz kann von unterschiedlichen Personen kodiert zu unterschiedlichen Ergebnissen führen. An dieser Stelle werden die individuellen Vorstellungen und eben auch Vorurteile von Programmierinnen und Programmierern, Datenanalytistinnen und -analysten sowie Auftraggeberinnen und -gebern mitunter Teil des Big-Data-basierten Entscheidungssystems.

Diese möglichen Verzerrungen im Datensatz sind von großer Bedeutung, da Ereignisse und gesellschaftliche Gruppen unterschiedlich gut durch Daten abgebildet zu sein scheinen – und zwar in systematischer Art und Weise. Doch gerade für den deutschen Kontext fehlt es hier noch an Erkenntnissen. Die Gründe für unterschiedliche Datenverfügbarkeit sind vielfältig: Zugang zu Technologie und digitale Beteiligung werden etwa durch wirtschaftliche, sprachliche, kulturelle und sozioökonomische Faktoren bestimmt. In Datensätzen können sich auch unbeabsichtigt historische Vorurteile oder Ungleichgewichte fortschreiben: Eingaben oder Ergebnisse aus der Vergangenheit reproduzieren sich in den Outputs eines algorithmischen Systems. So können beispielsweise bei Beschäftigungsverhältnissen frühere Einstellungsmuster, bei welchen Mütter in der Auswahl nicht berücksichtigt wurden oder nur Teilzeitjobs erhielten, fortgeschrieben werden, wenn der Algorithmus Kinderzahl und Geschlecht mit Daten der bisherigen Einstellungspolitik korreliert (Barocas und Selbst 2016, S. 689).

Selbst wenn Indikatoren, die Diskriminierung ermöglichen, wie etwa Geschlecht, ethnische Zugehörigkeit, Alter etc., aus Datensätzen gelöscht werden, können sogenannte redundante Kodierungen oder *proxies* dazu führen, dass die Zugehörigkeit zu einer bestimmten sozialen Gruppe in anderen Daten enthalten ist. Dies trifft gerade auch dann zu, wenn viele Daten mit hoher Granularität in die Berechnungen eingehen. So legte eine Studie offen, dass Facebook-“Likes“ mit demografischen Profilen und einigen psychometrischen Tests<sup>64</sup> es erlauben, Schlüsse über sensible Daten zu ziehen: In 88 % der Fälle ließ sich die sexuelle Orientierung der männlichen Nutzer richtig zuordnen sowie die Religionszugehörigkeit (christlich oder muslimisch, zu 82 %), die Hautfarbe (weiß oder schwarz bzw. *Caucasian* oder *African-American*, zu 95 %), die politische Ausrichtung (demokratisch oder republikanisch, zu 85 %), und der Alkohol und Zigarettengebrauch (zu 65 % bzw. 75 %) (Kosinski et al. 2013). Scheinbar nicht sensible Informationen können so leicht „enttarnend“ wirken.

---

64 Die Art der psychometrischen Tests wird im Artikel nicht erläutert.

### 3.7.4 Diskriminierung durch algorithmisch basierte Entscheidungssysteme

Fehlerquellen in Datensätzen führen erst dann zu gesellschaftlich wahrgenommener Diskriminierung, wenn sie in Entscheidungssysteme einfließen. Dazu gehören etwa *matching systems*, etwa in Suchmaschinen oder Social-Media-Plattformen, die Nutzerinnen und Nutzern dazu dienen, Informationen zu finden: über Produkte, Dienstleistungen, Personen, Ereignisse etc. Ein Beispiel ist etwa, dass Frauen bei Google seltener Werbeanzeigen für den beruflichen Aufstieg erhalten als Männer (Datta et al. 2015) oder dass Suchtreffer zu „schwarz“ klingenden Vornamen häufiger mit Werbeanzeigen einhergehen, die stigmatisierend auf die gesuchte Person wirken, da beispielsweise eine kriminelle Vergangenheit angedeutet wird (Sweeney 2013). Algorithmen können darüber hinaus (unbeabsichtigt) den Informationsfluss auf bestimmte Gruppen beschränken und anderen damit Zugangschancen verbauen. So stellen manche Arbeitgeberinnen und Arbeitgeber ihre Stellenanzeigen nur noch in sozialen Netzwerken ein. Es ist durchaus denkbar, dass die Empfehlungsalgorithmen bewirken, dass manche Stellenangebote nur bestimmten Gruppen angezeigt werden, etwa weißen ledigen Männern zwischen 25 und 40. Andere Gruppen wären dann faktisch von der Bewerbung ausgeschlossen.

Problematisch sind zudem Entscheidungsfindungssysteme, die sich so verhalten als seien Korrelation und Kausalität gleichzusetzen (siehe auch Unterkapitel zu Dimension 4). Korrelationen sind die Grundlage für Entscheidungsmodelle und somit nicht prinzipiell problematisch, sondern notwendig. Das Problem entsteht, wenn in Modellen bestimmte Faktoren überhöht werden, weil eine Kausalitätsannahme dahinter steckt, und andere Faktoren dagegen verblassen. Konkret haben Datenanalysen ergeben, dass die Länge der Pendelstrecke zur Arbeit ein starker Prädiktor für die Verbleibsdauer einer Arbeitnehmerin oder eines Arbeitnehmers bei einer Arbeitgeberin oder einem Arbeitgeber ist. Sofern Algorithmen diesen Faktor stark gewichten, können im Ergebnis Jobbewerberinnen und -bewerber, die gleiche oder sogar bessere Qualifikationen haben, benachteiligt werden, nur weil sie weiter entfernt wohnen (Barocas und Selbst 2016, S. 680; White House 2016, S. 15). Ein weiteres vieldiskutiertes Beispiel für die illegitime Gleichsetzung von Korrelation und Kausalität ist die weiter oben erwähnte wohnortbasierte Benachteiligung bei Finanzdienstleistungen (Brevoort et al. 2015).

### 3.7.5 Ethische Prinzipien und Regulierung

Ob Big Data bestimmte Gruppen benachteiligt und welche Gründe dies hat, entscheidet darüber, ob Big Data als sozial verträglich gilt und wie Gesellschaften mit Big Data umgehen. Denn diese Informationen sind die Grundlage für die Beschäftigung damit, welche Ungleichbehandlung als legitim erachtet wird und welche nicht. Sie helfen auch bei der Entscheidung, welche Kontrollen, Strafen und Entschädigungen angemessen sind. Denn gerade in den USA, aber zunehmend auch in Deutschland, gibt es eine öffentliche Debatte darüber, dass Diskriminierungen durch Big Data nicht nur zulasten der benachteiligten Individuen gehen, sondern auch gesamtgesellschaftliche Implikationen haben. So ist ein Diskurs über die Zukunft des Solidaritätsprinzips in Zeiten der Mikrofokussierung von Entscheidungen entstanden. Wissenschaftlerinnen und Wissenschaftler warnen vor einer weiteren „Benachteiligung der bereits Benachteiligten“ (Barocas und Selbst 2016, S. 675).

Ein erster Schritt sind somit gesellschaftliche Auseinandersetzungen darüber, welche Informationen für Big-Data-basierte Entscheidungen zugrunde gelegt werden sollten: Können alle Indikatoren, die durch Mustererkennung identifiziert und als gewichtig angesehen werden, die Grundlage für Entscheidungen sein oder müssen „sachbezogene“ Informationen den Ausschlag geben? Sollten etwa Kompetenzen der zentrale Faktor für die Bewertung von Bewerberinnen und Bewerbern für einen Arbeitsplatz sein oder dürfen auch Wohnortnähe und die Nutzung sozialer Medien zurate gezogen werden? Welche *proxies* sind in Datensätzen enthalten, die zur Benachteiligung bestimmter Gruppen führen können?

Eine weitere normative Debatte muss klären, in welchen Bereichen Big Data für Entscheidungen über Ein- und Ausschluss überhaupt eingesetzt werden darf: Für die Vergabe von Kreditdienstleistungen ist dies bereits üblich – auch in Deutschland. Findet das eine gesellschaftliche Akzeptanz? Wie sieht es mit der Allokation von Schul- und Studienplätzen aus? Mit der Vergabe von Wohnungen und Programmen zur gesundheitlichen Prävention? Gibt es auch Anwendungen, in denen Big Data für eine Verbesserung der Fairness eingesetzt werden kann? Es ist durchaus denkbar, dass die deutsche Antwort auf diese normativen Entscheidungen anders lautet als jene der US-amerikanischen Gesellschaft.

Eine weitere normative Auseinandersetzung verbindet sich schließlich mit der Frage, welche technischen Verfahren Einsatz finden. Doch gerade diese Debatte kann aufgrund einer mangelnden Informationsgrundlage kaum stattfinden: Die Datensätze, Algorithmen und Entscheidungssysteme, die in Big-Data-basierte Entscheidungsprozesse eingehen, sind den Betroffenen, etwa Bewerberinnen und Bewerbern um einen Kredit oder einen Arbeitsplatz, in der Regel nicht bekannt. Häufig werden sie als Betriebs-, Geschäftsgeheimnisse oder durch geistiges Eigentum

geschützt. Wenn sich Fehler oder rechtswidrige Praktiken einschleichen, können diese nur schwer erkannt und korrigiert werden. Entsprechend nennt der White House Report Transparenz, Rechenschaftspflicht und Verfahrensgerechtigkeit als zentrale Aspekte, um sicherzustellen, dass die Eingaben in ein algorithmisches System akkurat, geeignet und angemessen sind (White House 2016, S. 6-7). Wie mehr Transparenz und Rechenschaftspflicht über Algorithmen erlangt werden kann, bleibt eine offene Frage, die derzeit unter dem Stichwort „Algorithmenethik“ und *algorithmic accountability* diskutiert wird.

Zuletzt öffnet sich das Feld der möglichen Strategien, um Big-Data-basierten Diskriminierungen entgegenzuwirken: Welche werden bevorzugt eingesetzt und welche Erfolge zeitigen sie? All dies sind derzeit noch offene Forschungsfragen: Brauchen wir Verbote, bestimmte Daten zu erheben und zu verarbeiten, etwa für *proxies* für ethnische Zugehörigkeit (Roßnagel et al. 2016, S. 150f.)? Was bringen verpflichtende gruppenspezifische *impact assessments* (Sweeney 2013)? Welche Mechanismen erlauben es, ungleiche Behandlung wieder ausgleichen? Sweeney schlägt etwa vor, dass Google sicherstellt, dass Suchergebnisse zu geschützten Bevölkerungsgruppen nicht überproportional mit stigmatisierenden Werbeanzeigen gekoppelt werden (2013). Angwin et al. schlagen vor, dass Risikoscores für einzelne Subgruppen entwickelt werden, etwa um Frauen und bestimmten ethnischen Gruppen gerecht zu werden (Larson 2016). In welcher Form können *equal opportunity by design* und *bias mitigation* eingesetzt werden? Analog zum Konzept des *Privacy by Design* impliziert *equal opportunity* oder *Fairness by Design*, dass alle beteiligten Akteurinnen und Akteure, die eine Big-Data-basierte Anwendung entwickeln, in allen Arbeitsschritten mögliche Verzerrungen und weitere Probleme bedenken, die zu einer Diskriminierung führen können, um diesen entgegenzuwirken. Dazu gehört nicht zuletzt eine eingehende Reflektion der individuellen Vorurteile (*bias mitigation*) (White House 2016, S. 10).

All diese Themen sind keineswegs neu, sondern waren immer schon mit Statistik, Klassifizierung, Ungleichbehandlung und daraus resultierender potenzieller Diskriminierung verbunden. So ist beispielsweise bekannt, dass bei der Personalauswahl für einen Arbeitsplatz die Zugehörigkeit zu ähnlichen sozialen Milieus eine Rolle spielt (der *people like me bias*). Big Data stellt allerdings in vielen Bereichen die Gewissheiten darüber infrage, auf welcher Grundlage Entscheidungen getroffen werden dürften: Sollte es Arbeitgeberinnen und Arbeitgebern gestattet sein, Bewerber aufgrund von Kriterien zu bewerten, die wenig über ihre Kompetenzen aussagen, aber über ihre statistische Verbleibsdauer Auskunft geben? Sollte es Unternehmen weiterhin erlaubt sein, in intransparenter Weise die Kreditwürdigkeit von Personen zu berechnen und zu verkaufen? Ist es legitim, wenn Versicherte von staatlichen Krankenkassen aufgrund eines „gesunden“ Lebensstils Boni erhalten? Dies ist

nur ein Ausschnitt der Fragen, die nicht in erster Linie durch Gerichte, sondern politisch zu beantworten sind. Moderne Gesellschaften müssen sich im Angesicht von Big Data damit auseinandersetzen, welche Formen von Ungleichbehandlung legitim sein sollen und welche nicht. Durch Big Data generierte Kategorien können Gruppen benachteiligen, diesen aber auch Vorteile bringen. Immer wieder entspringen anhand von gesellschaftlichen Kategorisierungen Forderungen nach Bürgerrechten (Ruppert 2012, S. 214). Der Blick in die USA kann hierbei hilfreich sein, ersetzt jedoch nicht die Wertdebatte, die in Deutschland zu führen ist.

---

### **3.8 Regulativ hergestellte Erwartungen an Big Data: Regulierung von Big Data als Deutungskonflikt?**

Lena Ulbricht

Big Data wird nicht zuletzt auch geprägt durch Erwartungen, die regulativ hergestellt werden. Die Anwendung und Ausbreitung von Big Data wird durch kodifizierte Regeln bestimmt, etwa in Gesetzen, Verordnungen und Standards. Zu den regulativ hergestellten Erwartungen an Big Data gehören darüber hinaus auch Instrumente, mittels derer die Einhaltung der Regeln überwacht und sichergestellt wird. Welche Regeln die Entwicklung von Big Data bestimmen sollten, wird in der Öffentlichkeit kontrovers diskutiert. Ein besonderer Konflikt ist etwa um eine Flexibilisierung des Prinzips der Datensparsamkeit entstanden, wie es die Bundeskanzlerin es auf dem IT-Gipfel 2016 gefordert hat (Merkel 2016). Auch die Frage danach, welche Akteurinnen und Akteure sich regelkonform verhalten und welche nicht, wird öffentlich und vor Gerichten diskutiert, etwa bei der Frage, ob eine Datenweitergabe von Kommunikationsunternehmen an Nachrichtendienste legitim und rechtskonform ist. Dies gilt auch bei der Auseinandersetzung darüber, ob Unternehmen wie Google und Facebook die persönlichen Daten ihrer Nutzerinnen und Nutzer rechtmäßig erheben und verwenden. Die Literatur über Regulierungsdefizite und neue Instrumente für die Regulierung von Big Data wächst stetig. Diese Auseinandersetzung ist aber bislang stark rechtswissenschaftlich geprägt. Eine politikwissenschaftliche Perspektive erweitert die Debatte, indem sie die besondere Bedeutung von Diskursen, Institutionen und Akteurinnen und Akteuren unterstreicht, wie dies im akteurzentrierten Institutionalismus (Scharpf 2006), dem diskursiven Institutionalismus (Schmidt 2008) oder auch dem *advocacy coalition framework* (Jenkins-Smith et al. 2014) und weiteren politikwissenschaftlichen Ansätzen praktiziert wird.

Die Debatte über die Regulierung von Big Data wirft dabei eine zentrale Frage auf, welche die Regulierungsforschung derzeit grundlegend beschäftigt. So stellt sich die Frage, ob der *regulatory state* (Majone 1997; Hood et al. 1999) weiterhin Bestand hat und sich der Bereich staatlicher Kontrolle tatsächlich ausweitet. Oder ob sich das Rad zurückdreht und vermehrt regulierungsfreie Räume entstehen (Lodge und Wegrich 2016). Auch mit Blick auf Big Data lässt sich fragen: *Markiert die Ausweitung von Big Data kontrollfreie Räume und hinkt die staatliche Regulierung der technologischen Entwicklung hinterher? Oder bieten die bestehenden Regulierungsansätze im Prinzip ein engmaschiges Netz, das es lediglich an wenigen Stellen zu flicken gilt? Erstickt der Regulierungsstaat gar viele Big-Data-zentrierte Projekte und verhindert durch Überregulierung gesellschaftlich nützliche Innovationen?*

Umfassend können diese Fragen in diesem Beitrag nicht beantwortet werden, sie werden jedoch anhand von Beispielen diskutiert. Dabei wird dafür plädiert, die Debatte um die Regulierung von Big Data nicht allein als Suche nach den besten Lösungen für Big-Data-induzierte Risiken zu verstehen. Die Debatte über die Regulierung von Big Data ist vielmehr als gesellschaftlicher Deutungskonflikt zu verstehen. Die Risiken und Potenziale, die Big Data zugeschrieben werden, sind in dieser Lesart nicht objektiv vorhanden und messbar, sondern haben zwei Funktionen: Sie sind erstens Ausdruck verschiedener Deutungen der Digitalisierung. Sie sind zweitens auch Element von Durchsetzungsstrategien verschiedener Akteurinnen und Akteure, die in der Debatte um die Regulierung von Big Data ihr Selbstverständnis formen und ihre Handlungsspielräume zu erweitern suchen. Denn diskursive Auseinandersetzungen über Regulierung haben viele Funktionen: Sie erschaffen Bedeutung, sie ermöglichen Auseinandersetzungen, Koordination und Handlung und sie tragen zur Ausbildung des Selbstverständnisses von Akteurinnen und Akteuren bei (Black 2002). Dies soll in diesem Beitrag anhand der Diskussion über die Verbindung von Datenschutz- und Wettbewerbsregulierung verdeutlicht werden.

### **3.8.1 Big Data stellt Regulierung infrage: ein Deutungskampf**

Die öffentliche Debatte zur Regulierung von Big Data konzentriert sich weitgehend darauf, wie gesellschaftlich unerwünschte Folgen von Big Data verhindert werden können. Wenn man Regulierung aber als die absichtsvolle Setzung von Regeln versteht (Black 2008), gehören zur Regulierung von Big Data auch jene Regeln, die die Ausbreitung von Big Data fördern und die Art des Einsatzes spezifizieren sollen. Dazu gehören beispielsweise die Regelungen für die öffentliche Bereitstellung staatlich generierter Daten (*Open Government Data*). Wenn diese Daten in

großem Umfang in maschinenlesbarem Format und standardisiert bereitgestellt werden, können sie von Dritten leichter integriert und ausgewertet werden und so neue Geschäftsmodelle entstehen lassen, der Forschung nutzen und politische Teilhabe und demokratische Kontrolle ermöglichen. In Deutschland besteht eine Nationale E-Government Strategie, die mit dem Zielbereich „Transparenz und gesellschaftliche Teilhabe“ das Ziel hat, Open Data und Informationsfreiheit zu fördern (IT-Planungsrat 2015, S. 15; kritisch dazu: Wewer 2014). 2016 ist die Bundesregierung zudem dem Open Government Partnership beigetreten, einem internationalen Bündnis, das die Länder dazu verpflichtet, nationale Handlungspläne zu entwickeln (Open Government Partnership 2011; Stiftung Neue Verantwortung 2016). Die zivilgesellschaftliche Open Knowledge Foundation kritisiert allerdings, die Bundesregierung behandle die öffentliche Bereitstellung staatlicher Daten nicht als Priorität (Semsrott et al. 2016). Weitere Strategien, mit denen die Bundesregierung, Wirtschaft und Zivilgesellschaft günstige Rahmenbedingungen für die Entwicklung von Big Data schaffen wollen, sind die Verbesserung der IT-Sicherheit, der Schutz von Privatheit und Änderungen im Datenschutzrecht. All diese Ansätze haben zwei Seiten: Big Data zugleich zu ermöglichen und einzuhegen.

Big Data bestimmte Grenzen zu setzen, innerhalb derer sich die entsprechenden Anwendungen gemeinwohlfördernd entwickeln können, ist den zahlreichen Befürchtungen über unerwünschte Folgen für Individuen, Gruppen und die Gesellschaft geschuldet. Risiken von Big Data sind in den vorangehenden Unterkapiteln bereits ausführlich dargelegt worden: Verletzung der Privatsphäre, unfairer Wettbewerb, Manipulation, Diskriminierung, Einschüchterung durch Überwachung, Autonomieverluste sowie weitere Risiken. Kritikerinnen und Kritiker monieren, staatliche Regulierung begegne diesen Risiken bisher nicht angemessen und zahlreiche Studien diagnostizieren ein allgemeines Regulierungsdefizit (etwa Roßnagel 2013; 2016; Mantelero 2017; Weichert 2013; Europäischer Datenschutzbeauftragter 2015; White House 2014; Rouvroy 2016). Ein häufig zu vernehmendes Narrativ besagt sogar, Big Data stelle die Regulierung, die wir kennen, tiefgreifend infrage. Die Datenschutzregulierung, in der deutschen wie internationalen Regulierungsdebatte der zentrale Ansatz für die Regulierung von Big Data, wird etwa in doppelter Hinsicht als defizitär angesehen: Zum einen würden die rechtlichen Normen den Big-Data-basierten Praktiken zum Teil zuwiderlaufen und seien somit sowohl zu streng als auch unzureichend. Zu streng, da die Datensparsamkeit und Zweckbindung viele Big-Data-basierte Geschäftsmodelle verhindere; unzureichend, da sie auf einer Unterscheidung zwischen (besonders schützenswerten) personenbezogenen Daten und nicht-personenbezogenen Daten beruhe, die angesichts von Big Data nicht mehr aufrechtzuerhalten sei. Denn maschinenbezogene Daten, die etwa durch ein Smartphone erhoben werden, erlaubten derart persönliche Rückschlüsse über

ihre Nutzerinnen und Nutzer, dass sie als personenbezogene Daten angesehen werden müssten (Schwartz und Solove 2011).<sup>65</sup> Zudem werde das Instrument der informierten Einwilligung ausgehöhlt, da Nutzerinnen und Nutzer keine echten Alternativen hätten und somit keine wirkliche Opt-Out-Option bestünde (Rouvroy 2016). Ein weiterer Kritikpunkt ist zudem, dass die staatlichen Datenschutzbehörden unzureichend ausgestattet seien und das bestehende Recht deshalb nicht implementiert würde, wenn etwa keine Kontrollen durchgeführt werden (Schulzki-Haddouti 2016). Ein weiteres Argument für die Defizithese ist zudem, dass die schiere Masse an Daten, die Auskunft über menschliches Verhalten gibt, Rückschlüsse über alle erlaube, selbst über jene Personen, die ihre persönlichen Daten penibel schützten (Barocas und Nissenbaum 2014).

Obwohl die deutsche wie die internationale Regulierungsdebatte einen deutlichen Fokus auf Datenschutz legt, werden weitere Regulierungsdefizite benannt, die nicht primär durch Datenschutz reguliert werden können. Dazu gehört das Risiko, Big Data dazu zu nutzen, Gruppen zu diskriminieren (Sweeney 2013; Datta et al. 2015; Angwin et al. 2016) oder Individuen zu manipulieren (Zuboff 2015; Yeung 2016). Weitere gesellschaftliche Folgen von Big Data, die kritisiert werden, sind etwa abnehmende Diversitätstoleranz, etwa durch die sogenannten *filter bubbles* (Zuiderveen Borgesius et al. 2016) und Einschüchterung durch Überwachung durch sogenannte *chilling effects* (Lyon 2014; Yeung 2016). Auch hier sei der bestehende Regulierungsrahmen unzureichend, so die Defizithese.

Es finden sich auch viele Hypothesen über die Ursachen dafür, warum die gegenwärtige Regulierung gegenüber Big Data zu kurz greife: Ein Grund sei die Übermacht der datenverarbeitenden Unternehmen gegenüber seinen Nutzerinnen und Nutzern und den (meist staatlichen) Kontrolleurinnen und Kontrolleuren (Zuboff 2015; Weichert 2013). Eine weitere Interpretation besagt, Regulierungs- und Kontrollinstanzen hätten nicht genügend Regulierungswillen – wegen des erhofften Nutzens von Big Data für Sicherheitspolitik und Wirtschaftswachstum und wegen der massiven Lobbyarbeit der Internetwirtschaft (Albrecht 2014; Pasquale 2015). Eine weitere Erklärung konstatiert, es mangle an gesellschaftlichem Druck für eine bessere Regulierung. Datenschutz sei kein Thema, das große Proteste mobilisiere und Wahlen entscheide. Manche Beobachterinnen und Beobachter diagnostizieren gar, wir lebten in einer „Post-Privacy-Gesellschaft“, in der die Privatsphäre immer weniger Bedeutung habe (Guardian 2010).

---

65 Wenn mehr Daten verschiedener Quellen in eine Analyse einfließen, verbessern sich die Möglichkeiten, zuvor anonymisierte Daten wieder Personen zuzuordnen (Acquisti und Gross 2009).

### 3.8.2 Große Vielfalt der Regulierungsansätze

All diese Defizitdiagnosen implizieren, dass es bislang an Vorschlägen für effektive Regulierungsansätze mangle. Allerdings legt eine erste Analyse von Regulierungsvorschlägen eher das Gegenteil nahe; es besteht eine große Vielfalt an Regulierungsvorschlägen.

Im Bereich der *primär staatlichen Regulierung* zielen viele Regulierungsvorschläge darauf, die Datenschutzregulierung anzupassen und auszuweiten (z. B. durch die Etablierung des Marktortprinzips und das Recht auf Datenportabilität), andere Vorschläge zielen auf eine bessere Implementierung des vorhandenen Datenschutzrechts (etwa durch *privacy impact assessments* und eine bessere Ausstattung der Kontrollbehörden). Eine weitere Strategie sieht eine Verbindung von Datenschutz mit Wettbewerbsregulierung vor; wieder andere Vorschläge plädieren dafür, Big-Data-induzierte Risiken stärker durch Regulierungsansätze zum Schutz vor Diskriminierung, zum Schutz vor Manipulation und zum Schutz von Freiheitsrechten zu begegnen. Hinzu kommen Ansätze, Unternehmen oder Nutzerinnen und Nutzern haftungsrechtlich zu begegnen.<sup>66</sup>

Mit Blick auf die *Selbst-Regulierung durch Unternehmen* werden Standards, Normen und Selbstverpflichtungen wie etwa CSR-Standards (*corporate social responsibility*), die Entwicklung von *privacy-* und *fairness-*freundlichen Produkten und Dienstleistungen (*Privacy by Default, Privacy by Design, Fairness by Design, equal opportunity by design, bias mitigation*) sowie Verfahren zur Zertifizierung von Produkten, Praktiken und Organisationen durch Siegel und Audits etc. diskutiert.

Im Bereich der *professionellen Selbst-Regulierung* werden Ideen diskutiert, wie relevante Fachdisziplinen und Berufe, wie etwa Informatikerinnen und Informatiker, Data Scientists und Datenschützerinnen und Datenschützer sich Leitlinien für ethisches Verhalten und Qualifizierung geben können.

Vorschläge zur *Regulierung durch die Zivilgesellschaft* betreffen die Beratung und Unterstützung von Nutzerinnen und Nutzern, Bürgerinnen und Bürgern, damit sie ihre Rechte kennen, wahrnehmen und einfordern können. Dazu gehören auch Möglichkeiten von Daten- und Verbraucherschutzorganisationen, Verbandsklagen anzustrengen.

Die *Regulierung durch Selbst-Schutz* schließlich umfasst alle Initiativen, die Nutzerinnen und Nutzer sowie Bürgerinnen und Bürger dazu befähigen, zu ihrem

---

66 Auch wenn dieser Artikel die Rolle von Nachrichtendiensten weitgehend ausgeklammert hat, ist hier zu erwähnen, dass derzeit zahlreiche Ansätze einer besseren parlamentarischen Kontrolle von Nachrichtendiensten diskutiert werden. Die Regulierungsvorschläge betreffen auch die Praktiken der Datensammlung und -verwertung der Dienste.

eigenen Schutz beizutragen, etwa durch Selbstdatenschutz oder das Erkennen von Diskriminierung.

Die Beschäftigung mit diesen Ansätzen verfolgt derzeit noch in erster Linie die Frage, wie gut sie Big-Data-induzierte Risiken verringern können. Was fehlt, sind Analysen, die herausarbeiten, wie Krisendiagnosen, Defizitbeschreibungen und Regulierungsvorschläge zusammenhängen, und welche Diskurskoalitionen sich in der Debatte zeigen. Aus politikwissenschaftlicher Perspektive ist dabei besonders interessant, welche Implikationen verschiedene Vorschläge im Hinblick auf Machtverschiebungen haben: Welche Deutungen setzen sich durch, welche Akteurinnen und Akteure können ihre Handlungsspielräume erweitern oder müssen diese verringern und wie wirkt sich dies auf ihre (individuellen und kollektiven) Selbstverständnisse aus? Anhand der Debatte über die Nutzung von Wettbewerbsregulierung für die Regulierung von Big Data lässt sich dies veranschaulichen.

### **3.8.3 Möglichkeiten der Wettbewerbsregulierung**

In den letzten Jahren ist ein Diskurs über Möglichkeiten und Grenzen entstanden, mittels Wettbewerbsregulierung einige der Big-Data-bezogenen Probleme zu bekämpfen. Der Gedanke ist erst einmal überraschend: Wettbewerbsregulierung nimmt Märkte in den Blick und nicht den gesellschaftlichen Zusammenhalt. Dass der Ansatz dennoch in den vergangenen Jahren intensiv debattiert wurde, liegt an seiner Sensibilität für Machtverhältnisse. Wettbewerbsregulierung stellt sich als die Antwort auf ein (angenommenes) Machtgefälle zwischen datenverarbeitenden Unternehmen und Verbraucherinnen und Verbrauchern sowie zwischen Unternehmen und den (meist) staatlichen Regulierungs- und Kontrollinstanzen dar. Dieses Machtgefälle ist vielfältig: Den größten Vorteil genießen die führenden Konzerne der Internetwirtschaft dadurch, dass weder Nutzerinnen und Nutzer noch staatliche Kontrollbehörden erfahren können, ob die nutzerbezogenen Daten tatsächlich so erhoben, verwaltet und genutzt werden, dass dies der Rechtsprechung und den selbstgesetzten Regeln der Unternehmen entspricht. Einen weiteren Vorteil genießen die Unternehmen darin, dass sie große Ressourcen einsetzen können, um ihre Position im Rechtsstreit verteidigen zu können. Auch die Marktmacht mancher Plattform-Industrien ist so groß, dass Nutzerinnen und Nutzer sowie Kundinnen und Kunden wenig Alternativen haben, auf die sie ausweichen können. Dies gilt etwa für die Suchmaschine von Google, das Soziale Netzwerk von Facebook und das Anzeigengeschäft beider Konzerne. Ein Fall, in dem dieses Machtgefälle öffent-

lich thematisiert wurde, waren die Verfahren von Max Schrems gegen Facebook (Schrems 2017)<sup>67</sup>.

Während Datenschutzregulierung angesichts dieser Machtasymmetrien zum Teil ins Leere geht, sehen manche Akteurinnen und Akteure in der Wettbewerbsregulierung einen vielversprechenden Ansatz. Denn wenn Unternehmen ihre marktbeherrschende Stellung missbrauchen, um Kundinnen und Kunden zu schaden, etwa in Form geringer Datenschutzstandards, kann die Wettbewerbsregulierung einschreiten (Kuner et al. 2014). Eine erweiterte Wettbewerbsregulierung würde sich entsprechend nicht allein daran orientieren, ob monopolartige Strukturen Nutzerinnen und Nutzern mit Blick auf Preis und Qualität von Produkten und Dienstleistungen Nachteile bringen, sondern auch im Blick auf den Umgang mit ihren persönlichen Daten (Pozzato 2014). Auf dieser Grundlage hat etwa das Bundeskartellamt 2016 ein Verfahren gegen Facebook eingeleitet unter dem Verdacht, dass das Unternehmen seine dominante Marktposition dazu nutzt, um Datenschutzrecht zu unterlaufen. Der mutmaßliche Verstoß besteht darin, dass Facebook seine Nutzerinnen und Nutzer nicht angemessen über die Erhebung und Verwendung ihrer persönlichen Daten informiert (Bundeskartellamt 2016a). Zudem können die Wettbewerbshüter tätig werden, wenn der Wettbewerb zwischen Unternehmen gefährdet ist, sofern manche sich an die datenschutzrechtlichen Regeln halten und andere nicht. So kritisiert der Wirtschaftswissenschaftler Eric Clemons etwa, dass die Vormachtstellung von Google in Europa nicht auf technologischer Überlegenheit, sondern auf dem systematischen Verstoß gegen europäisches Datenschutzrecht beruhe (Clemons 2015). Würden Google und andere Monopolisten verpflichtet, sich an die Regeln zu halten, hätten auch andere Unternehmen eine Chance, ihre Marktanteile zu erhöhen, etwa jene, die besonders auf Datenschutz und Datensicherheit achten (Europäischer Datenschutzbeauftragter 2014). Wenn man anerkennt, dass Marktmacht in bestimmten Bereichen zentral davon abhängt, in welcher Form und in welchem Ausmaß personenbezogene Daten Unternehmen zur Verfügung stehen, liegt die Verantwortung dafür, dass Unternehmen nicht gegen Datenschutz verstoßen, also nicht allein bei den Kontrollbehörden für Datenschutz, sondern auch bei den Wettbewerbshüterinnen und -hütern (Bundeskartellamt 2016a; Bundeskartellamt 2016c).

Aus Perspektive derer, die eine Verbindung von Wettbewerbs- und Datenschutzregulierung befürworten, stellt die Datenökonomie einen Markt mit Besonderheiten dar, der entsprechend nach neuen regulatorischen Antworten verlangt. Dabei stehe

---

67 In verschiedenen Gerichtsverfahren hat Max Schrems darauf aufmerksam gemacht, dass Nutzerinnen und Nutzer nur geringe Möglichkeiten haben, ihr Recht auf informationelle Selbstbestimmung gegenüber Facebook wahrzunehmen.

nicht allein das Marktgleichgewicht zwischen Unternehmen im Fokus, sondern auch das Verhältnis zwischen Unternehmen und Kundinnen und Kunden bzw. Nutzerinnen und Nutzern. Letztere sollen nicht nur vor unfairen Vertragsbedingungen, sondern – durch den Fokus auf personenbezogene Daten – auch in ihren Grundfreiheiten geschützt werden.

Diese Perspektive nehmen das Bundeskartellamt, der Europäische Gerichtshof und der Europäische Datenschutzbeauftragte ein. Sie sind in unterschiedlicher Weise von den Konsequenzen einer solchen Strategie betroffen. Das Bundeskartellamt lotet Möglichkeiten aus, seinen Kontrollbereich auszuweiten – durch Verfahren wie jenes gegen Facebook, aber auch durch neue Kompetenzen. So soll das Amt etwa nach einem Gesetzentwurf aus dem Bundesministerium für Wirtschaft und Energie vom 01.07.2016<sup>68</sup> in Zukunft Musterverfahren anstrengen können (ecommerce Magazin 2017) und 2015 hat das Bundeskartellamt einen „Think Tank Internet“ eingerichtet. Der Europäische Datenschutzbeauftragte fordert wiederum mehr Unterstützung durch Wettbewerbshüterinnen und -hüter: indem beide Regulierungsbehörden in bestimmten Fällen Einsicht in Datensätze und Datenverarbeitung von Unternehmen erhalten sollten, um wettbewerbsschädigendes Verhalten feststellen zu können, und ganz allgemein in Form einer engeren Zusammenarbeit (Europäischer Datenschutzbeauftragter 2014). Wo sich der Europäische Gerichtshof in dieser Kontroverse zukünftig verortet, bleibt zu erforschen. Bislang zeigte er eher eine zögerliche Haltung, Wettbewerbs- und Datenschutzregulierung zu verknüpfen (Graef 2016, S. 19), hat aber prinzipiell eine Definition von Wettbewerbsregulierung, die sich durchaus offen zeigen könnte für Fragen des Datenschutzes (Kuner et al. 2014, S. 248).

Kritikerinnen und Kritiker wenden hingegen ein, die Wettbewerbsregulierung werde überstrapaziert, wenn prinzipiell jede Rechtsverletzung eines marktdominierenden Unternehmens auch als Verstoß gegen Wettbewerbsrecht gesehen würde (Graef und Alsenoy 2016). Sie fordern, die Einhaltung des Datenschutzrechts solle durch die Akteurinnen und Akteure und Instrumente der Datenschutzregulierung überwacht werden und Konflikte nicht auf dem Nebenschauplatz der Wettbewerbspolitik austragen (Kuner et al. 2014). Sie befürchten, dass die Definitionen und Instrumente der Wettbewerbsregulierung so stark ausgeweitet würden, um Datenschutz und Verbraucherschutz auf internetbasierten Märkten zu integrieren, dass sie dadurch grenzen- und substanzlos gerieten (Kuner et al. 2014; Pozzato 2014). Beispielsweise orientiert sich Wettbewerbsregulierung bislang daran, ob Monopole Konsumentinnen und Konsumenten mit Blick auf den Preis und die

---

68 Diese und weitere Kompetenzerweiterungen für das Bundeskartellamt werden derzeit im Rahmen der Reform des Gesetzes für Wettbewerbsbeschränkungen (GWB) diskutiert.

Qualität von Gütern und Dienstleistungen schaden. In welcher Form der Umgang von Monopolen mit den persönlichen Daten der Kundinnen und Kunden bewertet werden soll, ist bislang eine offene Frage (Pozzato 2014). Eine weitere Frage, die Kartellrechtlerinnen und -rechtler beschäftigt, ist, wie die Marktmacht von datenbasierten Plattformen und Netzwerken beurteilt werden kann (Bundeskartellamt 2016b). Zusammenfassend kann man die Deutung der Digitalisierung aus dieser Perspektive als einen Markt wie andere Märkte ansehen, der entsprechend keine besondere Regulierung notwendig macht. Vertreten wird sie unter anderem von der Generaldirektion Wettbewerb der Europäischen Kommission, dem zentralen Akteur der Wettbewerbsregulierung in der Europäischen Union.<sup>69</sup> Durch eine engere Zusammenarbeit mit dem Europäischen Datenschutzbeauftragten würde sie möglicherweise Spielräume dazugewinnen, aber auch neuen Verantwortungsbereichen, Erwartungen und nicht zuletzt Koordinationszwängen unterliegen. Ob und in welchem Maß sie dies in ihr Selbstverständnis aufzunehmen bereit ist, bleibt zu untersuchen.

### 3.8.4 Fazit und Fragen für politikwissenschaftliche Forschung

Angesichts der Debatte über die Regulierung von Big Data steht die Politikwissenschaft vor zahlreichen Fragen: Zum einen gilt es zu hinterfragen, ob und in welchem Ausmaß die These zutrifft, Big Data verlange nach neuen Regulierungsansätzen. Die Defizitbeschreibungen der bestehenden Regulierung und die Regulierungsvorschläge sind Teil eines größeren Konflikts über gesellschaftliche Deutungen der Digitalisierung und ihre Gestaltung. Welche Weltbilder stehen also hinter den widerstreitenden Regulierungsvorschlägen, welche Ziele und Akteurinnen und Akteure? Ein Blick auf die Kontroverse über die Verbindung von Wettbewerbs- und Datenschutzregulierung hat offengelegt, dass die Datenökonomie Mal als Markt wie jeder andere und Mal als Markt mit Besonderheiten verstanden wird. Interessanterweise nutzt das Bundeskartellamt die Debatte, um seine Handlungsspielräume zu erweitern, während sein Pendant auf EU-Ebene, die Generaldirektion Wettbewerb der Kommission, dies (bislang jedenfalls) abzulehnen scheint. Wie dies zu erklären ist, bleibt zu erforschen. Auch im Blick auf andere Vorschläge für eine Regulierung

---

69 Allerdings hat die Generaldirektion Wettbewerb zahlreiche Beschwerdeverfahren gegen Google wegen wettbewerbsschädigender Geschäftspraktiken eingeleitet. Datenschutzverstöße stehen hier jedoch nicht im Vordergrund (Europäische Kommission 2013, Europäische Kommission 2014, Europäische Kommission 2015, Europäische Kommission 2016a Europäische Kommission 2016b).

von Big Data sollten Analysen offenlegen, welche Weltansichten, Selbstverständnisse und Interessen mit ihnen verbunden sind. So kann man etwa fragen, weshalb das Instrument der informierten Einwilligung trotz aller Defizite so populär ist.<sup>70</sup> Und weshalb ist der Schutz vor Diskriminierung in Deutschland kaum ein Thema – anders als in Frankreich oder den USA (Ulbricht 2017)? Welche Regulierungsvorschläge haben sich in der Europäischen Datenschutzgrundverordnung durchgesetzt und welche nicht? Warum konnte sich etwa dort das Instrument der *privacy impact assessments* etablieren, obwohl es bislang wenig Erfahrung gibt und Evaluationen es als wenig verbindlich darstellen (Wright und Hert 2012)? Warum wurde im Gegenzug die Registrierungspflicht von Datenverarbeitungen nicht aufgenommen obwohl es diesbezüglich langjährige Erfahrungen in Großbritannien gibt und sie eine Grundlage für die Anlegung eines öffentlichen Registers hätte werden können (Schallaböck 2014)?

Zum anderen können politikwissenschaftliche Analysen auch einen Beitrag zur Frage nach den Möglichkeiten, Grenzen und Implikationen verschiedener Regulierungsvorschläge leisten – sowohl mit Blick darauf, ob sie gesellschaftlich unerwünschte Effekte von Big Data verhindern können, als auch mit Blick darauf, welche Folgen sie für die Gesellschaft insgesamt sowie für verschiedene Akteurinnen und Akteure wie Individuen, gesellschaftliche Subgruppen, verschiedene Unternehmen, Kontrollinstanzen etc. haben. Ganz allgemein stellt sich die Frage, wie das politische System die Regulierung von Big Data beeinflusst und wie diese Regulierung wiederum auf das politische System und seine Akteure zurückwirkt. Dies ist besonders mit Blick auf das institutionelle Gefüge zwischen Legislative und Judikative interessant – in einer Zeit, in der viele wichtige Fragen zur Gestaltung der Digitalisierung nicht durch Parlamente, sondern durch Gerichte entschieden werden (Ritzi i. E., Rehder und Schneider 2016).

Einiges deutet darauf hin, dass die vielfältigen Auswirkungen von Big Data auf Individuen, Gesellschaft, Wirtschaft und Politik und die verschiedenen Bereiche, in denen Big-Data-basierte Verfahren Anwendung finden, nach neuen Antworten verlangen. Nicht im Sinne radikal neuer Regulierungsformen, sondern vielmehr durch die Kombination von Regulierungsansätzen, Instrumenten und Akteurinnen und Akteuren, die bislang wenig miteinander zu tun hatten.

---

70 Siehe kritisch dazu etwa Hull (2015), sowie Hofmann und Bergemann (2016).

### 3.9 Fazit und Ausblick

Dass technologische Entwicklungen, anders als es die häufig verwendeten Metaphern von Big Data als „Flut“ oder „Öl“ suggerieren (Puschmann und Burgess 2014), keine Naturgewalten sind, sondern sich entlang gesellschaftlich gestalteter Grenzen entwickeln, muss heute nicht mehr betont werden. Dieser Beitrag hat herausgearbeitet, dass Big-Data-basierte Anwendungen in vielerlei Hinsicht durch kollektive Erwartungen geprägt werden und selbst wiederum der Herstellung von Erwartungen dienen, in kulturell-kognitiver, normativer und regulativer Hinsicht. In allen sechs Analysekatégorien wurde jedoch auch deutlich, dass über die tatsächlichen Wechselwirkungen zwischen Big-Data-basierten Phänomenen und kollektiven Erwartungen wenig bekannt ist. Viele der vernehmbaren Aussagen über die gesellschaftlichen Implikationen von Big Data müssen angesichts des noch spärlichen sozialwissenschaftlichen Forschungsstands somit erst einmal als Thesen gelten, die es zu überprüfen gilt.

In kultureller und kognitiver Hinsicht hat Big Data das Potenzial zur epistemischen Innovation, wenn die entsprechenden Methoden und Praktiken das Bild prägen, das eine Gesellschaft von sich hat. Diese Vorstellungen können auch politische Prozesse verändern, etwa wenn im Rahmen von Big-Data-basierten Wahlkämpfen neue Repräsentationsbeziehungen und kollektive Identitäten geschaffen werden. In welchem Ausmaß sich Big Data aber als Repräsentations-Technologie etabliert, muss allerdings eingehend untersucht werden.

Auch in normativer Hinsicht kann Big Data dafür genutzt werden, kollektive Erwartungen zu produzieren. So finden auf der Grundlage von Big Data Normalisierungsprozesse statt, die neuartig sind mit Blick auf Vielfalt, Personalisierung bzw. Granularität und zyklische Neuberechnungen. Big-Data-basierte Normierungen werden auch für Verhaltenssteuerung eingesetzt. Dabei zeigen sich als besondere Eigenschaften dieser Techniken ex-post Zuschreibungsprozesse, eine doppelte Intransparenz der algorithmischen Analysen und der daraus abgeleiteten Verhaltensregulierung sowie eine Situation von Selbst- und gegenseitiger Beobachtung. Ob dies zu neuen Wert- und Wertungsschemata führt und, intendiert oder nicht, die gesellschaftlichen Bedingungen für Verhaltensregulierungen ändert, ist allerdings noch unklar.

Big Data wird auch regulativ eingesetzt und kann Regelsetzung und -implementierung in verschiedener Hinsicht beeinflussen: mit Blick auf Datengrundlage und Rechenmethoden, wo in den meisten staatlichen Anwendungsbereichen eher traditionelle Daten und Verfahren eingesetzt werden. Big Data schafft auch neue Arten der Regelsetzung, die zwar eine Differenzierung von Regeln nach Zielgruppen oder Gebieten mit sich bringen, allerdings bislang wenig Einfluss auf die Regeln

der Regelsetzung selbst haben. Über die Autorität der Big-Data-basierten Regeln ist schließlich wenig bekannt. Die These, dass Big Data im staatlichen Bereich zu einer Renaissance kybernetischer Steuerung führt, scheint allerdings überzogen.

Mit Blick auf die Frage, welche kulturellen und kognitiven Erwartungen Big Data prägen, zeigt sich, dass Algorithmen zwar als Akteure angesehen werden, aber (noch) nicht als solche, die eigenständig und reflexiv die Wirklichkeit gestalten. Dennoch entstehen mit Blick auf Big Data auch neue Ambitionen analytischer Natur, nämlich, dass Big Data die soziale Wirklichkeit direkt abbildet. Diese Erwartungen an Big Data werden auch für die Gestaltung von Politik bedeutsam, haben aber bislang noch nicht dazu geführt, dass Datenanalyse als Ersatz für Intermediäre und soziale Praktiken herhält.

Die normativen Erwartungen an Big Data gehen zum Teil davon aus, dass Big Data menschliche Vorurteile und Defizite umgehen kann und eine neutrale und faire Grundlage für Entscheidungssysteme ist. Allerdings deuten erste Forschungsbefunde darauf hin, dass Big-Data-basierte Verfahren ebenfalls Verzerrungen beinhalten und diskriminierend wirken können. Die Auseinandersetzung über normative Erwartungen an Big Data befeuert somit eine Debatte über Grenzen der legitimen Differenzierung in modernen Gesellschaften.

Mit Blick auf die regulativ hergestellten Erwartungen an Big Data wurde deutlich, dass die These, dass Big Data die bestehende Regulierung grundlegend herausfordert, zum Teil zutrifft, wenn diese etwa noch immer auf einer Unterscheidung zwischen personenbezogenen und sonstigen Daten beruht. Es trifft aber nicht zu, dass technologischen Herausforderungen in erster Linie technisch begegnet werden muss. Zwar sollte Regulierung auch in die Gestaltung von Technik einfließen, etwa bei der Entwicklung von datensparsamen Technologien. Doch vielen Risiken von Big Data kann man mit bekannten Regulierungsansätzen und -instrumenten begegnen, wie etwa Datenschutzkontrollen oder Diskriminierungsverboten. Diese müssen jedoch zum Teil mit Blick auf Big Data weiterentwickelt werden.

Mit Blick auf die eingangs gestellte These, dass sich durch Big Data die Bedingungen kollektiv bindenden Entscheidens verändern, zeigten sich in diesem Beitrag also gemischte Befunde: Zwar zeigt sich in vielen Bereichen tatsächlich eine Mikrofokussierung von Regulierung, Normen und sozialen Wissensbeständen. Dies bringt durchaus weitreichende Änderungen mit sich. Dies ist im Übrigen auch dann der Fall, wenn die konkreten Mechanismen und Implikationen von Big Data den Menschen und Organisationen, die von ihnen betroffen sind, nicht ersichtlich sind. Denn auch eine vage oder fälschliche Annahme über Big Data kann zu kollektiven Erwartungen führen. Die Tiefe und Reichweite der durch Big Data verursachten Änderungen ist allerdings je nach Dimension und Bereich ganz unterschiedlich.

Diese differenzierte Bilanz gilt es weiter auszuarbeiten und entsprechende Thesen empirisch zu überprüfen.

## Literatur

- Acquisti, A., & Gross, R. (2009). Predicting Social Security numbers from public data. *Proceedings of the National Academy of Sciences of the United States of America* 106(27), 10975-10980.
- Acquisti, A. (2014). The Economics and Behavioral Economics of Privacy. In J. Lane, V. Stodden, S. Bender & H. Nissenbaum (Hrsg.), *Privacy, Big Data, and the Public Good* (S. 76-95). New York: Cambridge University Press.
- Agho, O. (2015). *New Political Actors in the Age of Big Data* (Master's Thesis, Georgetown University, Washington, D.C., United States). <https://repository.library.georgetown.edu/handle/10822/760824>. Zugegriffen: 05. Dezember 2016.
- Ajana, B. (2015). Augmented borders: Big Data and the ethics of immigration control. *Journal of Information, Communication and Ethics in Society* 13(1), 58-78. doi:10.1108/JICES-01-2014-0005
- Albrecht, J.P. (2014). *Finger weg von unseren Daten! Wie wir entmündigt und ausgenommen werden*. München: Knauer.
- Alemanno, A., & Sibony, A.L. (Hrsg.). (2015). *Nudge and the Law: a European Perspective*. Oxford: Hart Publishing.
- Amoore, L. (2011). Data Derivatives. *Theory, Culture & Society* 28(6), 24-43. doi:10.1177/0263276411417430
- Amoore, L., & Goede, M. (2012). Introduction. *Journal of Cultural Economy* 5(1), 3-8. doi: 10.1080/17530350.2012.640548
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Zugegriffen: 13. Januar 2017.
- Austin, J.L. (1975 [1962]). *How to do things with words*. Cambridge, MA: Harvard University Press.
- Barbero, M., Coutuer, J., Jackers, R., Mouddene, K., Renders, E., Stevens, W., Toninato, Y., Peijl, S., & Verstele, D. (2016). Big data analytics for policy making: A study prepared for the European Commission DG INFORMATICS (DG DIGIT). [https://joinup.ec.europa.eu/sites/default/files/dg\\_digit\\_study\\_big\\_data\\_analytics\\_for\\_policy\\_making.pdf](https://joinup.ec.europa.eu/sites/default/files/dg_digit_study_big_data_analytics_for_policy_making.pdf). Zugegriffen: 13. Januar 2017.
- Barocas, S., & Nissenbaum, H. (2014). Big Data's End Run around Anonymity and Consent. In J. Lane, V. Stodden, S. Bender & H. Nissenbaum (Hrsg.), *Privacy, Big Data, and the Public Good* (S. 44-75). New York: Cambridge University Press. doi:10.1017/CBO9781107590205.004
- Barocas, S., & Selbst, A.D. (2016). Big Data's Disparate Impact. *California Law Review* 104(3), 671-732.

- Beckert, J. (2016). *Imagined Futures. Fictional Expectations and Capitalist Dynamics*. Cambridge: Harvard University Press.
- Beer, D. (2016). How Should We Do the History of Big Data? *Big Data & Society* 1(3), 1-10.
- Behavioural Insights Team (2014). *EAST. Four simple ways to apply behavioural insights*. [http://38r8om2xjhhl25mw24492dir.wpengine.netdna-cdn.com/wp-content/uploads/2015/07/BIT-Publication-EAST\\_FA\\_WEB.pdf](http://38r8om2xjhhl25mw24492dir.wpengine.netdna-cdn.com/wp-content/uploads/2015/07/BIT-Publication-EAST_FA_WEB.pdf). Zugegriffen: 09. Februar 2017.
- Berger, P. L., & Luckmann, T. (1980 [1966]). *Die gesellschaftliche Konstruktion der Wirklichkeit*. Frankfurt a. M.: Fischer.
- Berger, P. L., & Luckmann, T. (1966). *The social construction of reality: A treatise on the sociology of knowledge*. New York: Anchor Books.
- Biem, A., Bouillet, E., Feng, H., Ranganathan, A., Riabov, A., Verscheure, O., Koutsopoulos, H., & Moran, C. (2010). IBM infosphere streams for scalable, real-time, intelligent transportation services. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data* (S. 1093-1104).
- Bimber, B. (2014). Digital Media in the Obama Campaigns of 2008 and 2012: Adaptation to the Personalized Political Communication Environment. *Journal of Information Technology & Politics* 11(2), 130-150. doi:10.1080/19331681.2014.895691
- Bishop, J. (2014). Representations of 'trolls' in mass media communication: a review of media-texts and moral panics relating to 'internet trolling'. *International Journal of Web Based Communities* 10, 7-24.
- Black, J. (2002). Regulatory Conversations. *Journal of Law and Society* 29(1), 163-196.
- Black, J. (2008). Constructing and contesting legitimacy and accountability in polycentric regulatory regimes. *Regulation & Governance* 2(2), 137-164.
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science* 2(1), 1-8.
- Bourdieu, P. (2009 [1981]). Beschreiben und Vorschreiben. Die Bedingungen der Möglichkeit politischer Wirkung und ihre Grenzen. In H. Beister, E. Kessler, J. Ohnacker, R. Schmid & B. Schwibs (Hrsg.), *Politik: Schriften zur Politischen Ökonomie* 2 (S. 11-22). Konstanz: UVK.
- Bourdieu, P. (2009 [1984]). Delegation und politischer Fetischismus. In H. Beister, E. Kessler, J. Ohnacker, R. Schmid, & B. Schwibs (Hrsg.), *Politik: Schriften zur Politischen Ökonomie* 2 (S. 23-41). Konstanz: UVK.
- Bowker, G. C. (2013). Data Flakes: An Afterword to „Raw Data“ Is an Oxymoron. In L. Gitelman (Hrsg.), *„Raw Data“ Is an Oxymoron* (S. 167-171). Cambridge, MA: MIT Press.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society* 15(5), 662-667.
- Brevoort, K. P., Grimm, P., & Kambara, M. (2015). *Data Point: Credit Invisibles*. [http://files.consumerfinance.gov/f/201505\\_cfpb\\_data-point-credit-invisibles.pdf](http://files.consumerfinance.gov/f/201505_cfpb_data-point-credit-invisibles.pdf). Zugegriffen: 13. Januar 2017.
- Bundeskartellamt (2016a). Bundeskartellamt eröffnet Verfahren gegen Facebook wegen Verdachts auf Marktmachtmissbrauch durch Datenschutzverstöße. [http://www.bundeskartellamt.de/SharedDocs/Meldung/DE/Pressemitteilungen/2016/02\\_03\\_2016\\_Facebook.html](http://www.bundeskartellamt.de/SharedDocs/Meldung/DE/Pressemitteilungen/2016/02_03_2016_Facebook.html). Zugegriffen: 27. Januar 2017.
- Bundeskartellamt (2016b). Französische und deutsche Wettbewerbsbehörde veröffentlichen gemeinsames Papier zu Daten und ihren Auswirkungen auf das Wettbewerbsrecht. [http://www.bundeskartellamt.de/SharedDocs/Meldung/DE/Pressemitteilungen/2016/10\\_05\\_2016\\_Big%20Data.html](http://www.bundeskartellamt.de/SharedDocs/Meldung/DE/Pressemitteilungen/2016/10_05_2016_Big%20Data.html). Zugegriffen: 06. Februar 2017.

- Bundeskartellamt (2016c). Bundeskartellamt veröffentlicht Arbeitspapier zum Thema „Marktmacht von Plattformen und Netzwerken“. [http://www.bundeskartellamt.de/SharedDocs/Meldung/DE/Pressemitteilungen/2016/09\\_06\\_2016\\_Think%20Tank.html](http://www.bundeskartellamt.de/SharedDocs/Meldung/DE/Pressemitteilungen/2016/09_06_2016_Think%20Tank.html). Zugegriffen: 14. Februar 2017.
- Butler, J. (2015). *Notes Toward a Performative Theory of Assembly*. Cambridge, MA: Harvard University Press.
- Butler, J. (1988). Performative acts and gender constitution: An essay in phenomenology and feminist theory. *Theatre journal*, 519-531.
- Callon, M. (2007). What Does It Mean to Say That Economics Is Performative? In D. MacKenzie, F. Muniesa & L. Siu (Hrsg.), *Do Economists Make Markets? On the Performativity of Economics* (S. 311-357). Princeton: Princeton University Press.
- Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems* 63, 67-80.
- Cheney-Lippold, J. (2016). Jus Algoritmi: How the National Security Agency Remade Citizenship. *International Journal of Communication* 10, 1721-1742.
- Chideya, F. (2015). Political Data is Everywhere — But What Does It All Mean? *The Intercept*. <https://theintercept.com/2015/07/17/political-data-mining-2016-election>. Zugegriffen: 03. Februar 2017.
- Christin, A., Rosenblat, A., & Boyd, D. (2015). Courts and Predictive Algorithms. [http://www.datacivilrights.org/pubs/2015-1027/Courts\\_and\\_Predictive\\_Algorithms.pdf](http://www.datacivilrights.org/pubs/2015-1027/Courts_and_Predictive_Algorithms.pdf). Zugegriffen: 13. Januar 2017.
- Christl, W. (2014). Kommerzielle Digitale Überwachung im Alltag. Studie im Auftrag der österreichischen Bundesarbeitskammer. [http://crackedlabs.org/dl/Studie\\_Digitale\\_Ueberwachung.pdf](http://crackedlabs.org/dl/Studie_Digitale_Ueberwachung.pdf). Zugegriffen: 13. Januar 2017.
- Clarke, A., & Margetts, H. (2014). Governments and citizens getting to know each other? Open, closed, and big data in public management reform. *Policy & Internet* 6, 393-417.
- Clemons, E. K. (2015). The EU Files Complaints Against Google, and It's About Time! *Huffington Post*. [http://www.huffingtonpost.com/eric-k-clemons/the-eu-files-complaints-against-google\\_b\\_7069780.html](http://www.huffingtonpost.com/eric-k-clemons/the-eu-files-complaints-against-google_b_7069780.html). Zugegriffen: 14. August 2016.
- Conn, J. (2014). VA puts the 'big' in big data predictive analytics. *VitalSigns*. <http://www.modernhealthcare.com/article/20140805/BLOG/308059999>. Zugegriffen: 13. Januar 2017.
- Crawford, K., Lingel, J., & Karppi, T. (2015). Our metrics, ourselves: A hundred years of self-tracking from the weight scale to the wrist wearable device. *European Journal of Cultural Studies* 18 (4-5), 479-496. doi: 10.1177/1367549415584857
- Dahlberg, L. (2001). Computer-mediated communication and the public sphere: A critical analysis. *Journal of Computer-Mediated Communication* 7, 0-0.
- Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated Experiments on Ad Privacy Settings. *Proceedings on Privacy Enhancing Technologies* 2015(1). doi:10.1515/popets-2015-0007
- Desmarais, S. L. & Singh, J. P. (2013). Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States. <https://csgjusticecenter.org/wp-content/uploads/2014/07/Risk-Assessment-Instruments-Validated-and-Implemented-in-Correctional-Settings-in-the-United-States.pdf>. Zugegriffen: 17. Januar 2017.
- Deutscher Bundestag (2012). *Datenschutz, Persönlichkeitsrechte: Fünfter Zwischenbericht der Enquete-Kommission „Internet und digitale Gesellschaft“, Bundestags-Drucksache 17/8999*. <http://dipbt.bundestag.de/dip21/btd/17/089/1708999.pdf>. Zugegriffen: 13. Januar 2017.
- Deutscher Bundestag (2017). *Entwurf eines Gesetzes über die Verarbeitung von Fluggastdaten zur Umsetzung der Richtlinie (EU) 2016/681: (Fluggastdatengesetz – FlugDaG)*

- Bundestags-Drucksache* 18/11501. <http://dipbt.bundestag.de/doc/btd/18/115/1811501.pdf>. Zugegriffen: 22. Mai 2017.
- Dion, M., Abdel Malik, P., & Mawudeku, A. (2015). Big Data and the Global Public Health Intelligence Network (GPHIN). *Canada Communicable Disease Report* 41, 209-214.
- Disch, L. (2008). The People as „Presupposition“ of Representative Democracy – An Essay on the Political Theory of Pierre Rosanvallon. *Redescriptions: Political Thought, Conceptual History and Feminist Theory* 12, 47-71.
- Duttweiler, S. (2016). Daten statt Worte? Bedeutungspraktiken in digitalen Selbstvermessungspraktiken. In T. Mämecke, J.H. Passoth & J. Wehner (Hrsg.), *Bedeutende Daten. Verfahren und Praxis der Vermessung und Verdatung im Netz*. Wiesbaden: Springer VS.
- Duttweiler, S., Gugutzer, R., Passoth, J.H., & Strübing, J. (2016). *Leben nach Zahlen. Self-Tracking als Optimierungsprojekt?* Bielefeld: transcript Verlag.
- Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology* 103(4), 962-1023.
- Espeland, W., & Stevens, M. (2008). A Sociology of Quantification. *Archives Europeennes de Sociologie* 2008 (3), 401-436.
- Europäischer Datenschutzbeauftragter (2014). Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy. [https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2014/14-03-26\\_competition\\_law\\_big\\_data\\_EN.pdf](https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2014/14-03-26_competition_law_big_data_EN.pdf). Zugegriffen: 16. Dezember 2016.
- Europäischer Datenschutzbeauftragter (2015). Meeting the challenges of big data: A call for transparency, user control, data protection by design and accountability. [https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-11-19\\_Big\\_Data\\_EN.pdf](https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-11-19_Big_Data_EN.pdf). Zugegriffen: 24. November 2015.
- Europäische Kommission (2013). The Google antitrust case: what is at stake? [http://europa.eu/rapid/press-release\\_SPEECH-13-768\\_de.htm](http://europa.eu/rapid/press-release_SPEECH-13-768_de.htm). Zugegriffen: 27. April 2017.
- Europäische Kommission (2014). Kartellrecht: Kommission erzielt von Google vergleichbare Anzeige konkurrierender spezialisierter Suchdienste. [http://europa.eu/rapid/press-release\\_IP-14-116\\_de.htm](http://europa.eu/rapid/press-release_IP-14-116_de.htm). Zugegriffen: 27. April 2017.
- Europäische Kommission (2015). Kartellrecht: Kommission übermittelt Google Mitteilung der Beschwerdepunkte zu seinem Preisvergleichsdienst. [http://europa.eu/rapid/press-release\\_MEMO-15-4781\\_de.htm](http://europa.eu/rapid/press-release_MEMO-15-4781_de.htm). Zugegriffen: 27. April 2017.
- Europäische Kommission (2017). What can big data do for you? <https://ec.europa.eu/digital-single-market/what-big-data-can-do-you>. Zugegriffen: 27. April 2017.
- Europäische Kommission (2016a). Kartellrecht: Kommission sendet Google Mitteilung der Beschwerdepunkte zu Android-Betriebssystem und -Anwendungen. [http://europa.eu/rapid/press-release\\_IP-16-1492\\_de.htm](http://europa.eu/rapid/press-release_IP-16-1492_de.htm). Zugegriffen: 27. April 2017.
- Europäische Kommission (2016b). Kartellrecht: Weitere Schritte der Kommission in Untersuchungen zum Preisvergleichsdienst und zu den Werbepraktiken von Google wegen mutmaßlichen Verstoßes gegen EU-Vorschriften. [http://europa.eu/rapid/press-release\\_IP-16-2532\\_de.htm](http://europa.eu/rapid/press-release_IP-16-2532_de.htm). Zugegriffen: 27. April 2017.
- Europäische Kommission (2016c). Richtlinie (EU) 2016/681 des Europäischen Parlaments und Rates vom 27. April 2016 über die Verwendung von Fluggastdatensätzen (PNR-Daten) zur Verhütung, Aufdeckung, Ermittlung und Verfolgung von terroristischen Straftaten und schwerer Kriminalität: RL 2016/681/EU. <http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1495448623854&uri=CELEX:32016L0681>. Zugegriffen: 22. Mai 2017.

- Europäisches Parlament (2016). EU Passenger Name Record (PNR) directive: an overview. [http://www.europarl.europa.eu/news/en/news-room/20150123BKG12902/eu-passenger-name-record-\(pnr\)-directive-an-overview](http://www.europarl.europa.eu/news/en/news-room/20150123BKG12902/eu-passenger-name-record-(pnr)-directive-an-overview). Zugegriffen: 13. Januar 2017.
- Europäische Zentralbank (2014). Social Media Sentiment and Consumer Confidence. Statistics Paper Series 5. <http://www.ecb.europa.eu/pub/pdf/scpsps/ecbsp5.pdf>. Zugegriffen: 03. Februar 2017.
- Ezrahi, Y. (2012). *Imagined Democracies. Necessary Political Fictions*. Cambridge: Cambridge University Press.
- Federal Trade Commission (2016). Big Data. A Tool for Inclusion or Exclusion? Understanding the Issues – FTC Report. <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>. Zugegriffen: 13. Januar 2017.
- Feenberg, A. (2002). *Transforming Technology*. Oxford: Oxford University Press.
- Fihn, S. D., Francis, J., Clancy, C., Nielson, C., Nelson, K., Rumsfeld, J., Cullen, T., Bates, J., & Graham, G. L. (2014). Insights from advanced analytics at the Veterans Health Administration. *Health affairs (Project Hope)* 33(7), 1203-1211. doi:10.1377/hlthaff.2014.0054
- Forelle, M. C., Howard, P. N., Monroy-Hernández, A., & Savage, S. (2015). Political bots and the manipulation of public opinion in Venezuela. SSRN. doi: 10.2139/ssrn.2635800
- Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings, 1972-1977*. New York: Pantheon.
- Ganz, K. (2014). Nerd-Pride, Privilegien und Post-Privacy: Eine intersektional-hegemonietheoretische Betrachtung der Netzbewegung. *FEMINA POLITICA – Zeitschrift für feministische Politikwissenschaft* 23(2). doi:10.3224/feminapolitica.v23i2.17613
- Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Berkeley, CA: University of California Press.
- Gillespie, T. (2014). The Relevance of Algorithms. In K. A. Foot, P. J. Boczkowski & T. Gillespie (Hrsg.), *Inside technology. Media technologies. Essays on communication, materiality, and society* (S. 167-193). Cambridge, MA: MIT Press.
- Gitelman, L., & Jackson, V. (2013). Introduction. In L. Gitelman (Hrsg.), *„Raw Data“ Is an Oxymoron* (S. 1-14). Cambridge, MA: MIT Press.
- Goerge, R. M. (2014). Data for the Public Good: Challenges and Barriers in the Context of Cities. In J. Lane, V. Stodden, S. Bender & H. Nissenbaum (Hrsg.), *Privacy, Big Data, and the Public Good* (S. 153-172). New York: Cambridge University Press.
- Gothe, H. (2014). Routinedaten im Ausland. In: E. Swart, P. Ihle, H. Gothe & D. Matusiewicz (Hrsg.), *Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven* (S. 260-267). Bern: Verlag Hans Huber.
- Graef, I. (2016). Blurring Boundaries of Consumer Welfare: How to Create Synergies between Competition, Consumer and Data Protection Law in Digital Markets. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2881969](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2881969). Zugegriffen: 16. Dezember 2016.
- Graef, I. & van Alsenoy, B. (2016). Data protection through the lens of competition law. <http://blogs.lse.ac.uk/mediapolicyproject/2016/03/23/data-protection-through-the-lens-of-competition-law-will-germany-lead-the-way>. Zugegriffen: 20. September 2016.
- Graff, G. M. (2016). Wie Big-Data-Startups die Prognosen zur US-Wahl revolutionieren. WIRED Germany. <https://www.wired.de/collection/science/wie-big-data-startups-die-prognosen-zur-us-wahl-revolutionieren>. Zugegriffen: 03. Februar 2017.
- Hacker, P. (2015). Nudge 2.0: The Future of Behavioural Analysis of Law in Europe and Beyond: A Review of ‚Nudge and the Law. A European Perspective‘, edited by Alberto

- Alemanno and Anne-Lise Sibony“. *European Review of Private Law* 24(2), 297-322. doi:10.2139/ssrn.2670772
- Hacking, I. (1999). *The Social Construction of What?* Cambridge, MA: Harvard University Press.
- Hanfeld, M. (2015). Punkte für gefälliges Verhalten. Frankfurter Allgemeine Zeitung. <http://www.faz.net/video/medien/punktrichter-citizen-score-ueberwachung-in-china-13848403.html>. Zugegriffen: 10. Februar 2017.
- Hardesty, L. (2013). How hard is it to „de-anonymize“ cellphone data?. MIT News. <https://news.mit.edu/2013/how-hard-it-de-anonymize-cellphone-data>. Zugegriffen: 10. Februar 2017.
- Hardt, M. (2014). How big data is unfair. Medium. <https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de#.l3ia947tq>. Zugegriffen: 13. Januar 2017.
- Hasbrouck, E. (2016). What's in a Passenger Name Record (PNR)? <https://hasbrouck.org/articles/PNR.html>. Zugegriffen: 13. Januar 2017.
- Heintz, B. (2008). Governance by Numbers. Zum Zusammenhang von Quantifizierung und Globalisierung am Beispiel der Hochschulpolitik. In G. F. Schuppert & A. Voßkuhle (Hrsg.), *Governance von und durch Wissen* (S. 110-128). Baden-Baden: Nomos.
- Helbing, D. (2016). Why we need democracy 2.0 and capitalism 2.0 to survive. <http://ssrn.com/abstract=2769633>. Zugegriffen: 10. Februar 2017.
- Hersh, E. (2015). *Hacking the Electorate*. New York: Cambridge University Press.
- Hofmann, J. (2014). Digitisation and Democracy: The challenges of shaping the Digital Society. In European Digital Forum (Hrsg.), *Digital Minds for a New Europe* (S. 52-54). Brüssel: Lisbon Council for Economic Competitiveness and Social Renewal.
- Hofmann, J. & Bergemann, B. (2016). Informierte Einwilligung – Ein Datenschutzphantom. *Spektrum der Wissenschaft Kompakt*, 50-59.
- Hofstetter, Y. (2016). *Das Ende der Demokratie. Wie künstliche Intelligenz die Politik übernimmt und uns entmündigt*. Gütersloh: C. Bertelsmann.
- Hood, C., Scott, C., James, O., Jones, G., & Travers, T. (1999). *Regulation inside Government: Waste-Watchers, Quality Police, and Sleaze-Busters*. Oxford, New York: Oxford University Press.
- House of Lords (2007). The EU/US Passenger Name Record (PNR) Agreement. <http://www.statewatch.org/news/2007/jun/eu-pnr-hol-report.pdf>. Zugegriffen: 13. Januar 2017.
- Hoyer, N., & Schönwitz, D. (2015). In der Mitte des Rasters. *Wirtschaftswoche* 39, 16-22.
- Hull, G. (2015). Successful failure: What Foucault can teach us about privacy self-management in a world of Facebook and big data. *Ethics and Information Technology* 17(2), 89-101. doi:10.1007/s10676-015-9363-z
- IT-Planungsrat (2015). Nationale E-Government-Strategie Fortschreibung 2015. [http://www.it-planungsrat.de/SharedDocs/Downloads/DE/NEGS/NEGS\\_Fortschreibung.pdf?\\_\\_blob=publicationFile&v=4](http://www.it-planungsrat.de/SharedDocs/Downloads/DE/NEGS/NEGS_Fortschreibung.pdf?__blob=publicationFile&v=4). Zugegriffen: 06. Februar 2017.
- Jenkins-Smith, H., Nohrstedt, D., Weible, C. M., & Sabatier, P. A. (2014). The Advocacy Coalition Framework: Foundations, Evolution, and Ongoing Research. In P. A. Sabatier & C. M. Weible (Hrsg.), *Theories of the policy process* (S. 183-224). Boulder, Colorado: Westview Press.
- Johnson, B. (2010). Privacy no longer a social norm, says Facebook founder. *The Guardian*. <https://www.theguardian.com/technology/2010/jan/11/facebook-privacy>. Zugegriffen: 03. Februar 2017.
- Jungheer, A. (2015). *Analyzing political communication with digital trace data*. Cham, Switzerland: Springer.

- Just, N., & Latzer, M. (2016). Governance by algorithms: reality construction by algorithmic selection on the Internet. *Media, Culture & Society*. doi:10.1177/ 0163443716643157
- Karpf, D. (2016). This Election Violates Everything We Thought We Knew About Data. <https://backchannel.com/this-election-violates-everything-we-thought-we-knew-about-data-935605ecf1b#.8s5hr4rhy>. Zugegriffen: 05. Dezember 2016.
- Katzenbach, C. (2012). Technologies as institutions: Rethinking the role of technology in media governance constellations. In N. Just & M. Puppis (Hrsg.), *Trends in Communication Policy Research: New Theories, Methods and Subjects* (S. 117-138). Bristol, UK: Intellect Books.
- Keller, R., Hirsland, A., Schneider, W., & Viehöver, W. (2005). *Die diskursive Konstruktion von Wirklichkeit*. Konstanz: UVK Verlagsgesellschaft.
- Kimmich, D., & Schahadat, S. (2016). Diskriminierung. Versuch einer Begriffsbestimmung. *Zeitschrift für Kulturwissenschaften* 2016(2), 9-21.
- Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography* 3(3), 262-267. doi:10.1177/2043820613513388
- Kitchin R. (2014a). *The Data Revolution: Big Data, Open Data, Data Infrastructures & their Consequences*. London: Sage.
- Kitchin, R. (2014b). The real-time city? Big data and smart urbanism. *GeoJournal* 79(1), 1-14.
- Kitchin, R., Coletta, C., Evans, L., Heaphy, L., Perng, S. Y., Bradshaw, B., & Lauriault, T. P. (2015). How vulnerable are smart cities to cyberattack? <http://progcity.maynoothuniversity.ie/2015/12/how-vulnerable-are-smart-cities-to-cyberattack/>. Zugegriffen: 05. Dezember 2016.
- Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction Policy Problems. *Papers and proceedings of the annual meeting of the American Economic Association* 105(5), 491-495. doi:10.1257/aer.p20151023
- Koelwel, D. (2017). Herausforderung für den Online-Handel: Mehr Verbraucherschutz im Internet. <http://www.e-commerce-magazin.de/herausforderung-fuer-den-online-handel-mehr-verbraucherschutz-im-internet>. Zugegriffen: 06. Februar 2017.
- Korff, D., & George, M. (2015). Passenger Name Records, data mining & data protection: the need for strong safeguards. <http://docplayer.net/16673463-Passenger-name-records-data-mining-data-protection.html>. Zugegriffen: 17. Januar 2017.
- Kosinski, M., Stillwell, D. & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences* 110, 5802-5805.
- Krieg, G. (2015). No-fly nightmares: The program's most embarrassing mistakes. CNN. <http://edition.cnn.com/2015/12/07/politics/no-fly-mistakes-cat-stevens-ted-kennedy-john-lewis/>. Zugegriffen: 13. Januar 2017.
- Kreiss, D. (2012). Yes we can (profile you). A brief primer on campaigns and political data. *Stanford Law Review Online* 64(70), 70-74.
- Kreiss, D., & Howard, P.N. (2010). New challenges to political privacy: Lessons from the first US Presidential race in the Web 2.0 era. *International Journal of Communication* 2010(4), 1032-1050.
- Krogerus, M., & Grassegger, H. (2016). Ich habe nur gezeigt, dass es die Bombe gibt. Das Magazin, Heft 48. [https://www.dasmagazin.ch/aktuelles\\_heft/n-48-3/?reduced=true](https://www.dasmagazin.ch/aktuelles_heft/n-48-3/?reduced=true).
- Kuchler, H. (2016). How 'big data' analysts are counting on your vote. Financial Times. <https://www.ft.com/content/fc1879be-1ed4-11e6-b286-cddde55ca122>. Zugegriffen: 03. November 2016.

- Kucklick, C. (2014). *Die granulare Gesellschaft: Wie das Digitale unsere Wirklichkeit auflöst*. Berlin: Ullstein Hardcover.
- Kuner, C., Cate, F.H., Millard, C., Svantesson, D.J. B., & Lynskey, O. (2014). When two worlds collide: the interface between competition law and data protection. *International Data Privacy Law* 4(4), 247-248.
- Kurz, C., & Rieger, F. (2012). *Die Datenfresser*. Bonn: bpb.
- Lane, J., Stodden, V., Bender, S., & Nissenbaum, H. (Hrsg.). (2014). *Privacy, Big Data, and the Public Good*. Cambridge, UK: Cambridge University Press.
- Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016). How We Analyzed the COMPAS Recidivism Algorithm. ProPublica. <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Zugegriffen: 13. Januar 2017.
- Latour, B. (2003). What if we talked politics a little? *Contemporary Political Theory* 2, 143-164.
- Latour, B. (2005). *Reassembling the Social: An introduction to actor-network-theory*. Oxford, UK: Oxford University Press.
- Law, J. (2009). Seeing like a survey. *Cultural Sociology* 3, 239-256.
- Law, J. (2012). Collateral realities. In: F.D. Rubio & P. Baert (Hrsg.), *The Politics of Knowledge* (S. 156-178). London: Routledge.
- Law, J., Ruppert, E., & Savage, M. (2011). The Double Social Life of Method. CRESC Working Paper Series. Milton Keynes: Centre for Research on Socio-Cultural Change. <http://research.gold.ac.uk/7987/1/The%20Double%20Social%20Life%20of%20Methods%20CRESC%20Working%20Paper%2095.pdf>. Zugegriffen: 03. Februar 2017.
- Law, J., & Urry, J. (2004). Enacting the social. *Economy and society* 33, 390-410.
- Lemov, R.M. (2005). *World as laboratory: experiments with mice, mazes, and men*. New York: Hill and Wang.
- Link, J. (2006). *Versuch über den Normalismus. Wie Normalität produziert wird*. Göttingen: Vandenhoeck & Ruprecht.
- Lodge, M., & Wegrich, K. (Hrsg.). (2014). *The problem-solving capacity of the modern state: Governance challenges and administrative capacities*. Oxford: Oxford University Press.
- Lowry, S., & MacPherson, G. (1988). A Blot on the Profession. *British Medical Journal* 296(6623), 657-658.
- Lupton, D. (2014). The commodification of patient opinion: the digital patient experience economy in the age of big data. *Sociology of Health & Illness* 36(6), 856-869. doi: 10.1111/1467-9566.12109
- Lupton, D. (2016). *The Quantified Self*. Cambridge, MA: Polity Press.
- Lyon, D. (2006). Airport Screening, Surveillance, and Social Sorting: Canadian Responses to 9/11 in Context. *Canadian Journal of Criminology and Criminal Justice* 48(3), 397-411. doi:10.1353/cj.2006.0030
- Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society* 2014, 1-13.
- Mager, A. (2012). Algorithmic Ideology: How capitalist society shapes search engines. *Information, Communication & Society* 15(5), 769-787.
- Mahr, B. (2003). Modellieren. Beobachtungen und Gedanken zur Geschichte des Modellbegriffs. In S. Krämer & H. Bredekamp (Hrsg.), *Bild, Schrift, Zahl* (S. 59-86). München: Wilhelm Fink Verlag.
- Majone, G. (1997). From the Positive to the Regulatory State: Causes and Consequences of Changes in the Mode of Governance. *Journal of Public Policy* 17(2), 139-167.

- Maki, K. (2011). Neoliberal deviants and surveillance: Welfare recipients under the watchful eye of Ontario Works. *Surveillance & Society* 9(1), 47-63.
- Mantelero, A. (2017). Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data. <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806ebe7a>. Zugegriffen: 06. Februar 2017.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>. Zugegriffen: 13. Januar 2017.
- March, S., Rauch, A., Bender, S., & Ihle, P. (2014). Datenschutzrechtliche Aspekte bei der Nutzung von Routinedaten. In E. Swart, P. Ihle, H. Gothe & D. Matusiewicz (Hrsg.), *Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven* (S. 291-303). Bern: Verlag Hans Huber.
- Marcuse, H. (1965). Industrialisierung und Kapitalismus. In O. Stammer & Deutsche Gesellschaft für Soziologie (Hrsg.), *Max Weber und die Soziologie heute: Verhandlungen des 15. Deutschen Soziologentages in Heidelberg 1964* (S. 161-180). Tübingen: Mohr Siebeck.
- Margetts, H., Hale, S. A., & Yasseri, T. (2014). Big Data and Collective Action. In M. Graham & W. H. Dutton (Hrsg.), *Society and the Internet: How Networks of Information and Communication are Changing Our Lives* (S. 223-237). Oxford: Oxford University Press.
- Marres, N., & Gerlitz, C. (2016). Interface methods: Renegotiating relations between digital social research, STS and sociology. *The Sociological Review* 64, 21-46.
- McKenna, E., Richardson, I., & Thomson, M. (2012). Smart meter data: Balancing consumer privacy concerns with legitimate applications. *Energy Policy* 41, 807-814. doi:10.1016/j.enpol.2011.11.049
- Merkel, A. (2016). Rede von Bundeskanzlerin Merkel beim 10. Nationalen IT-Gipfel am 17. November 2016, Saarbrücken. <https://www.bundestkanzlerin.de/Content/DE/Rede/2016/11/2016-11-17-rede-merkel-it-gipfel.html>. Zugegriffen: 06. Februar 2017.
- Merton, R. K. (1948). The self-fulfilling prophecy. *The Antioch Review* 8, 193-210.
- Merz, C. (2016). Predictive Policing – Polizeiliche Strafverfolgung in Zeiten von Big Data (ABIDA-Dossier). [http://www.abida.de/sites/default/files/Dossier\\_Predictive\\_Policing.pdf](http://www.abida.de/sites/default/files/Dossier_Predictive_Policing.pdf). Zugegriffen: 10. Februar 2017.
- Miller, P. (2001). Governing by Numbers: Why Calculative Practices Matter. *Social Research* 68(2), 379-396.
- Mitcham, C. (2014). Agency in humans and in artifacts: A contested discourse. In P. Kroes & P.-P. Verbeek (Hrsg.), *The moral status of technical artefacts* (S. 11-29). Dordrecht: Springer Netherlands.
- Moenchel, E. (2010). Flugdaten: US-Heimatschutz belügt EU. Futurezone. <http://www.fuzo-archiv.at/artikel/1659227v2>. Zugegriffen: 13. Januar 2017.
- Morozov, E. (2014). The Rise of Data and the Death of Politics. The Guardian. <https://www.theguardian.com/technology/2014/jul/20/rise-of-data-death-of-politics-evgeny-morozov-algorithmic-regulation>. Zugegriffen: 19. Januar 2017.
- Napoli, P. M. (2014). Automated media: An institutional theory perspective on algorithmic media production and consumption. *Communication Theory* 24(3), 340-360.
- Narayanan, A., & Shmatikov, V. (2006). How To Break Anonymity of the Netflix Prize Dataset. <https://arxiv.org/abs/cs/0610105>. Zugegriffen: 14. Februar 2017.

- Nickerson, D. W., & Rogers, T. (2014). Political Campaigns and Big Data. *Journal of Economic Perspectives* 28(2), 51-74. doi:10.1257/jep.28.2.51
- Nixon, R. (2016). Visitors to the U.S. May Be Asked for Social Media Information. New York Times. [http://www.nytimes.com/2016/06/29/us/homeland-security-social-media-border-protection.html?\\_r=2](http://www.nytimes.com/2016/06/29/us/homeland-security-social-media-border-protection.html?_r=2). Zugegriffen: 13. Januar 2017.
- Noursalehi, P., & Koutsopoulos, H. N. (2016). Real-time Predictive Analytics for Improving Public Transportation Systems' Resilience. Presented at the Data For Good Exchange 2016. <https://arxiv.org/ftp/arxiv/papers/1609/1609.09785.pdf>. Zugegriffen: 13. Januar 2017.
- O'Connor, B., Balasubramanyan, R., Routledge, B.R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *The International Conference on Web and Social Media* 11, 1-2.
- OECD 2015. *Data-Driven Innovation: Big Data for Growth and Well-Being*. Paris: OECD. <http://dx.doi.org/10.1787/9789264229358-en>
- Open Government Partnership (2011). Open Government Declaration. <http://www.open-govpartnership.org/about/open-government-declaration>. Zugegriffen: 14. Februar 2017.
- Osborne, T., & Rose, N. (1999). Do the social sciences create phenomena? The example of public opinion research. *The British Journal of Sociology* 50, 367-396.
- Pariser, E. (2012). *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Reprint edition. London: Penguin Books.
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.
- Pitkin, H. F. (1967). *The concept of representation*. Berkeley and Los Angeles: University of California Press.
- Portmess, L., & Tower, S. (2015). Data barns, ambient intelligence and cloud computing: The tacit epistemology and linguistic representation of Big Data. *Ethics and Information Technology* 17(1), 1-9. doi:10.1007/s10676-014-9357-2
- Pozzato, V. (2014). 2014 Opinion of the European Data Protection Supervisor: Interplay Between Data Protection and Competition Law. *Journal of European Competition Law & Practice* 5(7), 468-470. doi:10.1093/jeclap/lpu067
- Puschmann, C., & Burgess, J. (2014). Metaphors of Big Data. *International Journal of Communication* 2014 (8), 1690-1709.
- Quetelet, A. (1914). *Soziale Physik oder Abhandlung über die Entwicklung der Fähigkeit des Menschen*. Berlin: Fischer.
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems* 2, 3-12.
- Rehder, B., & Schneider, I. (2016). *Gerichtsverbände, Grundrechte und Politikfelder in Europa*. Baden Baden: Nomos.
- Reisch, L., Büchel, D., Joost, G., & Zander-Hayat, H. (2016). Digitale Welt und Handel. Verbraucher im personalisierten Online-Handel. [https://www.bmjjv.de/SharedDocs/Downloads/DE/Artikel/01192016\\_Digitale\\_Welt\\_und\\_Handel.pdf](https://www.bmjjv.de/SharedDocs/Downloads/DE/Artikel/01192016_Digitale_Welt_und_Handel.pdf). Zugegriffen: 10. Februar 2017.
- Rip, A. (1987). Controversies as Informal Technology Assessment. In *Knowledge: Creation, Diffusion, Utilization* 8, 349-371.
- Ritzl, C. (im Erscheinen). Dezision statt Konvention. Die Politisierung von Privatheit im digitalen Zeitalter. In T. Thiel & D. Jacob (Hrsg.), *Politische Theorie und Digitalisierung*. Baden-Baden: Nomos.

- Rosanvallon, P. (2002). *Le Peuple introuvable. Histoire de la représentation démocratique en France*. Paris: Gallimard.
- Rosanvallon, P. (2006). *Democracy past and future*. New York: Columbia University Press.
- Roßnagel, A. (2013). Big Data – Small Privacy? Konzeptionelle Herausforderungen für das Datenschutzrecht. *Zeitschrift für Datenschutz* 3(11), 562-567.
- Roßnagel, A., Geminn, C., Jandt, S., & Richter, P. (2016). Datenschutzrecht 2016 „Smart“ genug für die Zukunft? Ubiquitous Computing und Big Data als Herausforderungen des Datenschutzes. <http://www.uni-kassel.de/upress/online/OpenAccess/978-3-7376-0154-2.OpenAccess.pdf>. Zugegriffen: 13. Januar 2017.
- Rouvroy, A. (2016). Council of Europe, “Of Data and Men“ Fundamental Rights and Freedoms in a World of Big Data, Bureau of the Consultative Committee of the Convention for The Protection of Individuals with Regard to Automatic Processing of Personal Data. [https://works.bepress.com/antoINETTE\\_rouvroy/64/](https://works.bepress.com/antoINETTE_rouvroy/64/). Zugegriffen: 13. Januar 2017.
- Ruppert, E. (2012). Seeing Population: Census and Surveillance by Numbers. In K. Ball, K. Haggerty & D. Lyon (Hrsg.), *Routledge International Handbook of Surveillance Studies* (S. 209-216). London: Routledge.
- Ruppert, E., Law, J., & Savage, M. (2013). Reassembling the Social Science Methods. *Theory, Culture & Society* 30(4), 22-46.
- Russell Neuman, W., Guggenheim, L., Mo Jang, S., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication* 64, 193-214.
- Sales, N. A. (2015). Big Data at the border: balancing visa-free travel and security in a digital age. *ECPR Conference Université de Montréal* 26.-29. August. <https://ecpr.eu/Events/PaperDetails.aspx?PaperID=26167&EventID=94>. Zugegriffen: 13. Januar 2017.
- Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology* 41, 885-899.
- Saward, M. (2006). The representative claim. *Contemporary Political Theory* 5, 297-318.
- Schallaböck, J. (2014). Verbraucher-Tracking. [https://www.gruene-bundestag.de/fileadmin/media/gruenebundestag\\_de/themen\\_az/digitale\\_buergerrechte/Tracking-Bilder/Verbraucher\\_Tracking.pdf](https://www.gruene-bundestag.de/fileadmin/media/gruenebundestag_de/themen_az/digitale_buergerrechte/Tracking-Bilder/Verbraucher_Tracking.pdf). Zugegriffen: 06. Februar 2017.
- Scharpf, F. W. (2006). *Interaktionsformen: Akteurzentrierter Institutionalismus in der Politikforschung*. Unveränderter Nachdruck der 1. Auflage. Wiesbaden: VS, Verlag für Sozialwissenschaften.
- Scherer, M. (2012). Inside the Secret World of the Data Crunchers Who Helped Obama Win. Time. <http://swampland.time.com/2012/11/07/inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win>. Zugegriffen: 03. Februar 2017.
- Schmidt, V. A. (2008). Discursive Institutionalism: The Explanatory Power of Ideas and Discourse. *Annual Review of Political Science* 11(1), 303-326. doi:10.1146/annurev.polisci.11.060606.135342
- Schintler, L. A., & Kulkarni, R. (2014). Big Data for Policy Analysis: The Good, The Bad, and The Ugly. *Review of Policy Research* 31(4), 343-348. doi:10.1111/ropr.12079
- Schneier, B. (2015). *Data und Goliath: Die Schlacht um die Kontrolle unserer Welt : wie wir uns gegen Überwachung, Zensur und Datenklau wehren können*. München: Redline Verlag.
- Schrems, M. (2017). Europe versus Facebook. <http://europe-v-facebook.org/EN/en.html>. Zugegriffen: 14. Februar 2017.
- Schulzki-Haddouti, C. (2016). Des Kaisers neue Kleider: Wie sieht eine angemessene Datenschutzkontrolle aus? In *DatenDebatten Bd. 1. Zukunft der informationellen Selbstbestimmung* (S. 111-126). Berlin: Erich Schmidt Verlag.

- Schwartz, P. M., & Solove, D. J. (2011). The PII Problem: Privacy and a new concept of personally identifiable information. *New York University Law Review* 86, 1814-1894.
- Scott, J. C. (1999). *Seeing Like a State*. New Haven, CT: Yale University Press.
- Scott, W. R. (2008). *Institutions and organizations: Ideas and interests*. 3 Auflage. Los Angeles: Sage.
- Seaver, N. (2013). Knowing algorithms. Presentation held at Media in Transition 8. <http://nickseaver.net/papers/seaverMiT8.pdf>. Zugegriffen: 05. Dezember 2016.
- Selke, S. (2016). *Lifelogging. Digitale Selbstvermessung und Lebensprotokollierung zwischen disruptiver Technologie und kulturellem Wandel*. Wiesbaden: Springer VS.
- Semsrott, A., zum Felde, J., & Palmeshofer, W. (2016). „Rohstoff der Zukunft“: Was bringt das Open-Data-Gesetz? <https://www.okfn.de/blog/2016/10/opendata-gesetz/>. Zugegriffen: 14. Februar 2017.
- Skeem, J. L., & Lowenkamp, C. T. (2016). Risk, Race and Recidivism: Predictive Bias and Disparate Impact. *Criminology* 54(4), 680-712. doi:10.1111/1745-9125.12123
- Smith, R. J. (2014). Missed Miracles and Mystical Connections: Qualitative Research, Digital Social Science and Big Data, In M. Hand & S. Hillyard (Hrsg.), *Big Data? Qualitative Approaches to Digital Research* (S. 181-204). Bingley, UK: Emerald Group Publishing Limited.
- Stiftung Neue Verantwortung (2016). Impuls für Open Government in Deutschland. <http://www.stiftung-nv.de/publikation/impuls-f%C3%BCr-open-government-deutschland>. Zugegriffen: 06. Februar 2017.
- Straßheim, H., Jung, A., & Korinek, R. L. (2015). Reframing Expertise: The Rise of Behavioural Insights and Interventions in Public Policy. In A. B. Antal, M. Hutter & D. Stark (Hrsg.), *Moments of Valuation. Exploring Sites of Dissonance* (S. 249-268). Oxford: Oxford University Press.
- Straßheim, H. (im Erscheinen). Behavioural Expertise and Regulatory Power in Europe. In M. Lee, A. de Ruiter & M. Weimer (Hrsg.). *Regulating Risks in the European Union*. Oxford: Hart Publishing.
- Streeck, W., & Thelen, K. A. (2005). Introduction: Institutional Change in Advanced Political Economies. In W. Streeck & K. A. Thelen (Hrsg.), *Beyond continuity: institutional change in advanced political economies*. Oxford: Oxford University Press.
- Sunstein, C. R. (2015). *Choosing Not to Choose: Understanding the Value of Choice*. Oxford: Oxford University Press.
- Sunstein, C. R. (2014). *Why nudge? The Politics of Libertarian Paternalism*. New Haven/London: Yale University Press.
- Sußmilch, J. P. (1761). *Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts aus der Geburt, dem Tode und der Fortpflanzung desselben. Erster Theil*.
- Swart, E., Ihle, P., Gothe, H., & Matusiewicz, D. (Hrsg.). (2014). *Routinedaten im Gesundheitswesen. Handbuch Sekundärdatenanalyse: Grundlagen, Methoden und Perspektiven*. 2. vollständig überarbeitete Auflage. Bern: Verlag Hans Huber.
- Swan, M. (2013). The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery. *Big Data* 1, 85-99.
- Sweeney, L. (2013). Discrimination in Online Ad Delivery. *Communications of the Association of Computing Machinery* 56(5), 44-54. doi:10.1145/2447976.2447990
- Tallacchini, M., Boucher, P., & Nascimento, S. (2014). *Emerging ICT for Citizens' Veillance: Theoretical and Practical Insights*. Brüssel: Publications Office of the European Union.
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving Decisions About health, wealth and happiness*. London: Penguin Books.

- Theile, C. (2016). Nirgendwo auf der Welt arbeiten so paranoide Menschen wie hier. Interview mit Emmanuel Mogenet. Tagesanzeiger. <http://www.tagesanzeiger.ch/digital/internet/nirgendwo-auf-der-welt-arbeiten-so-paranoide-menschen-wie-hier/story/19360540>. Zugegriffen: 05. Dezember 2016.
- Tufekci, Z. (2014). Engineering the public: Big data, surveillance and computational politics. *First Monday* 19(7). doi:10.5210/fm.v19i7.4901
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *The International Conference on Web and Social Media* 10, 178-185.
- Ulbricht, L. (2017). Big Data und Diskriminierung – Impulse aus den USA für eine deutsche Debatte. In *ads aktuell 1 – Newsletter der Antidiskriminierungsstelle des Bundes*.
- Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society* 12(2), 197-208.
- Vormbusch, U., & Kappler, K. (2016). Leibschreiben. Zur medialen Repräsentation des Körperleibes im Feld der Selbstvermessung. T. Mämecke, J.H. Passoth & J. Wehner (Hrsg.), *Bedeutende Daten. Verfahren und Praxis der Vermessung und Verdattung im Netz*. Wiesbaden: Springer VS.
- Voß, J.-P., & Amelung, N. (2016). Innovating public participation methods: Techno-scientization and reflexive engagement. *Social Studies of Science* 26, 749-772.
- Washington, A. L. (2014). Government Information Policy in the Era of Big Data. *Review of Policy Research* 31(4), 319-325. doi:10.1111/ropr.12081
- Wehner, J. (2008). Taxonomische Kollektive. Zur Vermessung des Internet. In H. Willems (Hrsg.), *Weltweite Welten: Internet-Figurationen aus wissenssoziologischer Perspektive* (S. 363-383). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Weichert, T. (2013). Big Data und Datenschutz. <https://www.datenschutzzentrum.de/bigdata/20130318-bigdata-und-datenschutz.pdf>. Zugegriffen: 06. Februar 2017.
- Weltbank (2015). World Development Report 2015: Mind, Society, and Behavior. <http://www.worldbank.org/en/publication/wdr2015>. Zugegriffen: 10. Februar 2017.
- Wewer, G. (2016). *Open Government, Staat und Demokratie*. Berlin: Edition Sigma.
- White House (2014). *Big data: seizing opportunities, preserving values*. <http://purl.fdlp.gov/GPO/gpo64868>. Zugegriffen: 06. Dezember 2016.
- White House (2016). Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights. [https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016\\_0504\\_data\\_discrimination.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf). Zugegriffen: 13. Januar 2017.
- White, M. C. (2016). Orbitz Shows Higher Prices to Mac Users. Time. <http://business.time.com/2012/06/26/orbitz-shows-higher-prices-to-mac-users/>. Zugegriffen: 13. Januar 2017.
- Wright, D., & de Hert, P. (2012). *Privacy Impact Assessment (1. Auflage)*. Law, Governance and Technology Series 6. Dordrecht: Springer Netherlands.
- Yeung, K. (2016). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, 1-19.
- Zuboff, S. (2015). Big other. Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30 (1), 75-89.
- Zuiderveen Borgesius, F., Trilling, D., Möller, J., Bodó, B., de Vreese, C., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review* 5(1). doi:10.14763/2016.1.401.