BIG DATA:
OR, THE VISION THAT WOULD NOT FADE

A dissertation submitted in compliance
with the requirements for the degree of

DOCTOR OF PHILOSOPHY (Ph.D.)

by

Gernot Rieder

IT University of Copenhagen
Submitted 14th July 2018

Supervisor:
Judith Simon

Examination Committee:
Evelyn Ruppert, Irina Shklovski, Sally Wyatt

for my mother

Table of Contents

Part II
Research Papers

Abstract English

This doctoral thesis examines the rise of Big Data as a complex sociotechnical phenomenon. How is it that the notion of Big Data could gain so much traction, becoming the subject of considerable hype and hope? Adopting a broadly interdisciplinary perspective, the article dissertation investigates the historical roots as well as contemporary causes of the recent push for data and numerical evidence. Furthermore, it critically analyses the epistemic promises of industry vendors, considers the implications of Big Data as a tool for policy and decision making, and explores fictional accounts of algorithmic regulation that may trigger a more reflexive engagement with the potential pitfalls of Big Data-based technologies. The thesis objective is to contribute to a better understanding of the politics of people's trust in numbers, but also to point towards a more responsible and democracy-preserving governance of Big Data-driven innovation. What is ultimately argued is that Big Data should not be seen as a fad destined to fade, but as the current tagline for a process of computerization and datafication that will continue to shape and transform every aspect of our daily lives.

Abstract Danish

Denne Ph.d.-afhandling undersøger fremkomsten af Big Data som et komplekst socioteknisk fænomen. Hvordan er det gået til, at Big Data har vundet så stor opmærksomhed og er omgærdet af så betydelig håb og *hype*? Ud fra et bredt tværfagligt perspektiv undersøger artikel-afhandlingen de historiske rødder og de nutidige årsager til de seneste års efterspørgsel efter data og numerisk evidens. Endvidere foretages en kritisk analyse af brancheleverandørers epistemiske løfter, implikationerne af Big Data som politisk og beslutningsmæssigt værktøj tages i betragtning, ligesom fiktive fremstillinger af algoritmisk regulering udforskes, som muligvis kan understøtte en mere refleksiv håndtering af de potentielle faldgruber ved Big Data-baserede teknologier. Tesens formål er at bidrage til en bedre forståelse af den politiske betydning af vores tillid til tal, men også at udpege nogle retninger for en mere ansvarlig og demokrati-bevarende *governance* af Big Data-drevet innovation. Ultimativt argumenteres der for det synspunkt, at Big Data ikke bør opfattes som et modefænomen, der med tiden vil falme, men som det aktuelle udtryk for en digitalisering og dataficering, som også fremover vil præge og transformere alle aspekter af vores daglige tilværelse.

Acknowledgments

This thesis represents the culmination of three years working and studying as a doctoral candidate at the IT University of Copenhagen, Denmark, a remarkable interdisciplinary institution that has proven an ideal environment for this dissertation project. Throughout the past years, I have benefited greatly from the support and opportunities provided by ITU, for which I would like to express my sincere gratitude and appreciation.

My foremost thanks go to my supervisor, Prof. Judith Simon, who has provided me with continual support, critical feedback, and valuable insight into a fast-moving research area. It was through her guidance and mentorship that this thesis was written in a spirit of curiosity and free thinking, while still balanced against the need for tangible results. For the chance to sometimes wander off, grow, and explore I am tremendously grateful.

My next thanks go to my friends and colleagues at the ITU Copenhagen, especially to the brilliant members of the Technologies in Practice (TiP) research group, who have graciously commented on parts of this dissertation on various occasions. In hindsight, I could not have wished for a more generous and inspiring research community. A special shout-out goes to the artists of wing 4E whose creative tinkering with animate and inanimate objects allowed my mind to rejoice whenever the written word became too difficult to process. There is true delight in having wonderful, thought-provoking exhibitions right outside your office door. For this and for the many coffee conversations, thank you.

I am also much indebted to Prof. Geoffrey C. Bowker, Prof. Judith Gregory, and Assoc. Prof. Christina Neumayer for their thoughtful comments and suggestions during my mid-way evaluation. They have greatly contributed to the refinement of my arguments.

Material from this dissertation has been presented at various conferences and meetings. I would like to thank all audiences for their questions, interest, and feedback. I would also like to acknowledge the anonymous reviewers who took the time to read and comment on the articles that form the core of this thesis. Their constructive suggestions have helped in fine-tuning the manuscripts and pushing each paper into a better shape. A special thanks goes to the editors who have supported the articles' publication process, especially to Ass. Prof. Andrew Iliadis and Ass. Prof. Federica Russo, who guest edited the *Big Data & Society* special theme on *Critical Data Studies*, and to Prof. Ann Rudinow Sætnan, Prof. Ingrid

Schneider, and Dr. Nicola Green, who invested great efforts in seeing the edited volume *The Politics of Big Data: Big Data, Big Brother?* successfully through the press.

In terms of finance, I am grateful to the Austrian Science Fund (grant number P-23770) for supporting the first year of this research endeavor, and to the ITU committee that awarded me with a tuition free stipend and decided to employ me for the remaining duration of the project.

Last but not least, I would like to extend my deepest thanks to family members and friends. To my brother, Bernhard, for reading through some of the material, but also for keeping conversations about the topic both lighthearted and fun. To Martin, Thomas, and Walther for their open ears and witty comments. To Dr. Thomas Völker for not only having me visit the European Commission's Joint Research Centre in beautiful Ispra, Italy, but also for co-authoring a paper on *datafictions* that I know was dear to both our hearts. To all of you, thank you very much for your friendship and efforts.

PART I
Dissertation Framework

1. Introduction

When I started my PhD studies in early 2015, there was a lot of confusion around the term and concept of "Big Data". Although various definitions had been proposed (see Press, 2014), at my university, not even the researchers from technical disciplines could agree on a common meaning; if one would add the social science and humanities scholars to the mix, the chaos was perfect and a Tower of Babel-like situation would prevail. A literature review did not provide much clarity. While some would hold that Big Data was best understood as society's ability to harness information in novel ways (see Mayer-Schönberger and Cukier, 2013), others would argue that it had to do with the volume, velocity, and variety of data available (see Laney, 2012; 2001). While some would speak of a new problem-solving philosophy (see Hartzog and Selinger, 2013), others would point to the prominence and status acquired by data as a recognized output (see Leonelli, 2014). While some would call it a movement (see Sardana and Sardana, 2013), others would label it the "buzzword of the decade" (Barocas and Selbst, 2016 [2014]: 673). These contrasting views could in part be attributed to different scholarly or professional backgrounds, but they were also indicative of a broader truth: Big Data was more than just the mountains of data generated in an increasingly computerized world, and it was more than our improving ability to collect, store, and make sense of these data. Rather, in order to gain a deeper understanding of the rise and prevalence of Big Data, it would be necessary to treat it as a complex *sociotechnical phenomenon*, one that, as boyd and Crawford (2012) had observed early on, rests on an interplay of science, technology, and culture, an amalgamation of hardware power and analytical prowess combined with the "widespread belief that large data sets offer a higher form of intelligence […], with and aura of truth, objectivity, and accuracy." (ibid.: 663) Given this messiness, no wonder there was a good deal of confusion and conceptual disagreement.

What soon became clear was that researching and mapping the debate around Big Data was a bit like aiming at a moving target, with new articles and feature stories published on a daily basis. Big Data was *en vogue*, and initially the coverage was mainly positive: "A new goldmine", proclaimed *The Economist*, celebrating the implementation of open data initiatives (Economist.com, 2013); a "revolution by numbers", promised *The Guardian*, listing sectors and fields that were bound to be changed by the oncoming data deluge (Naughton,

2012); and the German tabloid *Bild* ran a story on how the country's national football team was using Big Data to prepare for 2014 World Cup in Brazil (see Stein, 2014). A popular book at the time was Mayer-Schönberger and Cukier's (2013) *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. The book mainly focused on possible gains, promoting the idea that Big Data would "shake up everything from businesses and the sciences to healthcare, government, education, economics, the humanities, and every other aspect of society" (ibid.: 11). Yet despite its focus on the expected benefits of modern data analytics,[1] the latter part of the book already hinted at some of the problems that would soon overshadow the initial rush of excitement: the erosion of privacy under constant digital surveillance; the growing potential for systematic bias and discrimination; society's overreliance on numbers that are far more fallible than often assumed. Soon the media's tone shifted, and while there was still no shortage of Big Data success stories, they were balanced out by more critical reports: "Eight (no, nine!) problems with Big Data", found *The New York Times* (Marcus and Davis, 2014); "Big Data: are we making a big mistake?", wondered the *Financial Times* (Harford, 2014); and *The Atlantic* made sure to properly "welcome" its readers "to algorithmic prison" (Davidow, 2014). What seemed to evolve was a *climate of clash* between different stakeholders, but more generally, if one would peruse the comment sections, between people with different backgrounds, experiences, and beliefs. Vendors and programmers, regulators and NGOs, concerned citizens and affects publics – suddenly everyone appeared to have a different take on what Big Data meant and implied for society at large. What was striking was the polarization of the debate, which showed that at bottom this was not just a casual conversation, but rather a serious dispute over diverging values, interests, and priorities, and the kind of future society should aim for.

The controversial atmosphere was nourished by a nonstop series of global and regional events: from the Snowden revelations to the Cambridge Analytica scandal, from U.S. data breaches to the EU's right to be forgotten, from the British care.data initiative to China's social credit system, there seemed to be hardly a day without Big Data-related headlines making the news. Larger happenings were accompanied by smaller, often "creepy" (Shklovski et al., 2014) incidents: nosy smart TVs capturing living room chatter (see Harris, 2015), interactive Barbie dolls recording whatever a child was saying (see Halzack, 2015), social

[1] For a critical review of the book, see Armstrong (2014).

media platforms experimenting with users' emotional states (see Meyer, 2014), seemingly innocuous smartphone apps sharing sensitive user information (see Goodin, 2015), Twitter chatbots turning racist overnight (see Vincent, 2016), the list could be continued almost indefinitely. But none of this was apparently enough to diminish what Kallinikos (2013) has called "the allure of Big Data", that is, the promise that Big Data would allow to "assemble the big picture of many phenomena" and "contribut[e] to lifting, as it were, the veil of reality" (ibid.: 42). As I write these lines in the spring of 2018, Big Data, despite massive criticism on epistemic and ethical grounds, continues to be a popular buzz term, even if part of the debate has "started to shift to a focus on computation and analytics such as algorithms, machine learning, and artificial intelligence" (Ruppert et al., 2017: 2). Between January 1st and March 31st, for instance, the *New York Times* has published 33 articles that use the term either in the heading or body of text, which is less than during the same months in 2014 (54 articles) and 2015 (42), but more than in 2016 (20) and 2017 (27).[2] Equally, a Google Trends search, which charts search term interest over time, shows a popularity score of 89 for "Big Data" in March 2018, which is higher than in March 2014 (72) and about the same as in March 2015 (89) and 2016 (91). Global search interest was highest in March 2017, with a "peak popularity" score of 100.[3] The term has not disappeared from the political scene either. Searching the European Commission's official website, one can find an abundance of recent Big Data-related content and initiatives: from Big Data for migration management (see Rango and Vespe, 2017), official statistics (see ESSnet Big Data, 2018), and healthcare (see Andriukaitis, 2018) to briefings about existing regulatory challenges and possible ways forward (see EC-DTM, 2017). Thus, despite earlier predictions that Big Data would go out of fashion by 2014 (see KDnuggets, 2012), the term has proven remarkably resilient, still serving as a go-to concept well after its presumed expiration date.

But how was it that the vision of Big Data proved so successful? How did it come to be that what mainly started as a marketing strategy gained so much traction and power, especially in the political domain? This cumulative dissertation is, in part, an attempt to dig deeper and search for possible answers to these questions, exploring the rise and

[2] Number of articles was counted using the on-site search functions of nytimes.com.

[3] Of course, such indicators should be interpreted with caution. In this case, they are solely presented in support of the claim that global (search) interest in Big Data has not diminished but, as of yet, proven stable.

proliferation of Big Data in public policy. But it does not stop there. Making the case that Big Data should not be regarded as a fad destined to fade, the thesis considers ways to anticipate and prepare for the consequences of an increasingly datafied society, seeking to contribute to recent discussions about responsible research and innovation (see Stilgoe et al., 2013). Thus, rather than simply dismissing it as yet another overblown hype, the following treats the widespread Big Data excitement as a meaningful object of inquiry, examining its composition, the reasons for its persistence, and potential short- and longer-term repercussions. Because even if the term itself should eventually run out of steam, the powers and dynamics that contributed to its prominence probably will not, making it that much more important to gain a clearer sense of the promises and expectations but also the pressures and challenges that fueled its cultural impact, shaping the "dreamscapes of modernity" (Jasanoff and Kim, 2009) for at least the better part of a decade. Ultimately, though offering a number of concluding remarks, the thesis should not be read as final statement on the matter, but rather as a modest contribution to a crucial issue and debate that is set to become even more relevant as the world turns increasingly digital.

2. Thesis Outline, Methodology, and Impact

Prelude

The main part of the thesis consists of the thesis framework (Chapters 1 to 4) and five peer-reviewed papers (four published, one currently under review). Of the five papers, four are full-length research articles and one is a concise commentary. Four of the papers have been co-authored with other scholars: Three with my thesis supervisor, Dr. Judith Simon, Professor for Ethics in Information Technology at the University of Hamburg, Department of Informatics, and one with Dr. Thomas Völker, currently postdoctoral researcher at the European Commission's Joint Research Centre in Ispra, Italy. Both have kindly agreed for the papers to be incorporated and published as part of this thesis.

The papers approach Big Data from different directions and angles, seeking to treat it, as argued in the introduction, as a complex and messy phenomenon that is cultural as well as technoscientific. In order to capture the multifacetedness of the Big Data discourse, the papers make use of *sensitizing concepts*[4] from multiple fields and disciplines, including Science and Technology Studies (STS), technology assessment (TA), ethics, social epistemology, and literature studies. This plurality of perspectives is at once a strength and a weakness. On the one hand, the use and combination of different conceptual and disciplinary lenses allows for a better appreciation of the phenomenon's intricate nature, a way to illuminate the nexus of science, technology, culture, and power that *co-produces*[5] the *hopeful monster*[6] that is Big Data. On the other hand, the breadth of perspectives bears the risk that each individual approach does not go quite deep enough, merely scratching the surface of what should be a much more detailed discussion. To be sure, the themes and questions explored in the papers are rich enough to warrant further analytical investigation, potentially

[4] As Blumer (1954: 7) notes: "[A sensitizing concept] gives the user a general sense of reference and guidance in approaching empirical instances. Whereas definitive concepts provide prescriptions of what to see, sensitizing concepts merely suggest directions along which to look."

[5] In essence, the idiom of co-production can be seen as "a critique of the realist ideology that persistently separates the domains of nature, facts, objectivity, reason and policy from those of culture, values, subjectivity, emotion and politics." (Jasanoff, 2004: 3)

[6] In the edited volume *A Sociology of Monsters*, Law (1991) employs the figure of the hopeful monster to refer to the heterogeneous networks, overlaps, and collages that make up modern sociotechnical order. His argument is dialectic, holding that while "entities make history", they are "made by history too" (ibid.: 18), thus opposing positions of both technological determinism and social reductionism.

even at the scale of a separate thesis project. But this would be a very different project, one that would not, and probably could not, attend to the Big Data phenomenon's *multiplicity*[7] in a similar way. Ultimately, the thesis sought to find a balance between depth and breadth, with emphasis on the STS-guided examination of the emergence of Big Data in the political sphere (Papers I, III, and IV).

While using a number of sensitizing concepts, the empirical analysis was guided by the tenets and procedures of the *grounded theory approach*, a qualitative mode of inquiry where "theory is derived from data and then illustrated by characteristic examples of data" (Glaser and Strauss, 2006 [1967]: 5). The results of this research approach are best visible in Rieder (2018): Here, the notion of sociotechnical imaginaries was used as a conceptual framework for investigating the European Commission's vision of Big Data, while the analysis of the collected material was supported by the construction of analytic codes and categories that is central to grounded theory methodology (e.g., see Charmaz, 2006: 42ff.). In addition, the different visualization techniques of Clarke's (2005) *situational analysis*[8] served as a valuable tool to make sense of the "universes of discourse" (Strauss, 1978: 120) that had formed around Big Data. Especially the creation of arena maps (see Clarke, 2005: 109ff.) proved useful in developing a bottom-up understanding of the various perspectives and motivations at play, and of the different "social worlds" (ibid.) that would sustain them. Eventually, engagement with the data would lead to the development of broader themes and theories, which subsequently were tested against additional chunks of data. The thesis' treatment of empirical material can thus can be regarded as the outcome of a shifting process between inductive theory generation and deductive hypothesis testing, an abductive form of reasoning that resonates with the exploratory spirit of the project. In the remainder of this chapter, the individual papers will be discussed in greater detail, thereby providing a clear overview of the thesis structure, the methods applied, and the intended outcomes and impacts. In addition to concise summaries of the main findings, the aim of these discussions is to offer some insight into how the papers came to be and what each of them sought to achieve. The papers

[7] For a thorough treatment of ontological multiplicity in an STS fashion, see Mol (2002).

[8] Situational analysis can be seen as an extension of traditional grounded theory, with special attention to the analysis of discourses, the importance of nonhuman elements, and the impact of power relations (see Clarke et al., 2015).

are connected through interludes that are supposed to afford some additional continuity between the presented arguments and perspectives.

2.1. Discussion of Paper I - *Datatrust: Or, the Political Quest for Numerical Evidence and the Epistemologies of Big Data*

The *Datatrust* commentary (Rieder and Simon, 2016) was written for a *Big Data & Society* special theme on *Critical Data Studies* (CDS), co-edited by Andrew Iliadis and Federica Russo.[9] The special theme conceptualized data as a "form of power" (Iliadis and Russo, 2016: 1), inviting contributions that would "analyz[e] the ground upon which positivistic Big Data science stands" (ibid.: 2). Agreeing with the CDS aim to study the "technological, political, social and economic apparatuses" that "constitut[e] and fram[e] the generation, circulation and deployment of data" (Kitchin and Lauriault, 2014: 1), the commentary was an attempt to shed some light on the historical foundations of the Big Data phenomenon, situating the recent push for *evidence-based policy making*[10] within a broader sociocultural context. More specifically, the paper sought to illuminate the roots of people's "trust in numbers" (Porter, 1995), comparing it to other forms of trust – i.e., trust in human experts or institutions – and making the case that authorities' reliance on numbers and figures is often a sign of political weakness rather than strength, a strategy of democratically accountable administrations to maintain legitimacy and shield themselves from persistent scrutiny in times of crisis, public distrust, and uncertainty. By approaching the issue from a history of science perspective and drawing a line from early statistics in 17th-century England and 18th-century France (see Desrosières, 1998) to modern-day data analytics, the paper thus sought to highlight *patterns of continuity*, arguing that there is value in considering the *long durée* of processes of counting

[9] The commentary has also been translated into German (Rieder and Simon, 2018) and published as part of the edited volume *(Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft* (Mohabbat Kar et al., 2018).

[10] The notion of evidence-based policy is not new. Solesbury (2002), for instance, notes an "ascendancy of evidence" (90) in British public policy during the late 90's and early 2000's, finding a "new thirst for knowledge" marked by a more "instrumental view of research" (91) for social and economic development. Yet, as a Deloitte study for the European Commission indicates, the rise of Big Data "gave birth to a renaissance of evidence-based policy making practices", offering new techniques "to find insights […] and to answer questions that very previously considered beyond reach" (Barbero et al., 2016: 9). Thus, while the notion has been around for some time, it gained considerable momentum with the advent of Big Data analytics, as for example demonstrated by the recent proliferation of "data for policy" initiatives (see Poel et al., 2015).

and accounting rather than solely focusing on the proclaimed novelty of Big Data science. What such a perspective can allow is a greater appreciation of "quantification as a social technology" (Porter, 1995: 49) as well as of the origins of the strive for "mechanical objectivity" (Daston and Galison, 2010) as a preferred way of knowing. Such a reading may also provide some analytical distance, an opportunity to look beyond the marketing narrative and regard the ascent of Big Data as but the latest manifestation of the age-old *Cartesian dream* – the idea that "the emergent rational method and the deployment of science and science-based technology would deliver the truth (expressed in quantitative terms) and empower humanity to fulfil its destiny as masters and possessors of Nature" (Funtowicz and Pereira, 2015: 1). It is by thinking along these lines that the commentary aimed to encourage dialogue and critical reflection, urging for a historically grounded analysis of the quest for numerical evidence that presently culminates in the promotion and funding of Big Data monitoring and forecasting technologies (e.g., see European Commission (EC), 2017). The paper is not alone in this ambition: Ambrose (2015), for example, compares Big Data to the probabilistic revolution of the mid-1800s, arguing that this lens allows us to assess Big Data as an epistemic revolution, as opposed to simply a technological or economic one, offering insight into our attempts to govern it. In a similar vein, Aronova and co-authors (2017), in the introduction to a substantial *Osiris* issue on *Data Histories*, emphasize that Big Data can be seen as a chapter in a longer history of observation, quantification, and statistical methods, amplifying "features of data-driven science already present or latent in earlier data cultures", including "the automation of data collection, storage, and analysis; the mechanization of hypothesis testing; an increasing division of labor between data collectors and consumers; and the technological black-boxing of many data practices" (ibid.: 7). The *Datatrust* paper's call for a contextualization of Big Data practices within the broader history of social statistics has thus been echoed by others; what is fairly original, however, is its particular focus on questions of trust and data-driven forms of governance, adding a heightened sense of immediacy to the subject matter.

Interlude: In 2016, at around the same time the *Datatrust* paper was released, David Beer published an article that asked, "How should we do the history of Big Data?" (Beer, 2016: 1) In it, Beer argues that while it is important to place Big Data within the genealogical lineage of the modern state and consider it as part of a much longer series of historical developments,

it is equally important to explore the actual concept of Big Data: "where it came from, how it is used, what it is used for, how it lends authority, validates, justifies, and makes promises" (ibid.: 2). What is needed, according to Beer, are close examinations of the *discursive framings* around Big Data, providing rich contextual accounts of how the notion becomes embedded in organizational, political, and cultural life. While the *Datatrust* paper mainly sought to contribute to the former part of the project – that is, to historically situate Big Data within the expansion of numbers and metrics over the past few centuries – the next two articles published in the course of this thesis were geared towards the latter part, aiming to provide a clearer picture of the Big Data-related claims and promises of both vendors (see Rieder and Simon, 2017a) and policymakers (see Rieder, 2018). The underlying assumption was that a better understanding of the discursive framings would reveal something about the nature of the phenomenon itself, and the dynamics that sustained it. As Beer aptly notes: "The power of Big Data is not just in the data themselves, it is in how those data and their potential is imagined and envisioned" (2016: 9). As Papers II and III show, these imaginaries and visions became a central concern of the thesis project.

2.2. Discussion of Paper II - *Big Data: A New Empiricism and Its Epistemic and Socio-Political Consequences*

The second paper (Rieder and Simon, 2017a) was a research article published in an edited volume, a *Festschrift* in honor of the retirement of Prof. Klaus Mainzer, the then Chair of Philosophy and Philosophy of Science at the Technical University of Munich (see Pietsch et al., 2017). In the main, the paper followed two lines of inquiry: First, drawing on Kitchin's (2014a; b) influential work on the subject, it sought to examine some of the most prominent *epistemic claims* made by Big Data proponents, focusing on how the concept is marketed by stakeholders from business and industry. Reading through a plethora of news items, product sites, tech reports, but also field-related books and publications, it was possible to discern the composition of a very specific *epistemic imaginary*, based on four key assumptions: that Big Data is exhaustive in scope, capable of capturing entire populations or domains (N=all); that high volume data can speak for themselves, eliminating the need for a priori theory; that data are neutral and objective, offering a disinterested picture of the world that counteracts human biases; and that modern data analytics can provide certainty in an uncertain world,

accurately predicting ever more aspects of human life. The main point here was not that these assumptions are necessarily wrong – even though researchers from various disciplines have refuted them on both conceptual and empirical grounds (see Kitchin, 2014a; Leonelli, 2014) – but to emphasize that this imaginary feeds into a culture where numbers and figures are seen as infallible, unchallengeable harbingers of truth, allegedly looking at the world from an all-seeing god's-eye perspective. Consequently, the second part of the paper was devoted to examining the sociopolitical consequences of *blind data trust*, drawing on Pasquale's (2015) notion of the "black box society", indicating a system "whose workings are mysterious" (ibid.: 3) and where people are increasingly submitted to the "rule of scores and bets" (ibid.: 191). More precisely, the text offered a concise discussion of three interrelated issues and concerns: issues of *opacity*, caused by either corporate or state secrecy, high algorithmic complexity, or code volatility; problems of *accountability*, that is, the unwillingness of public and private entities to render themselves accountable as well as the inadequacy of current laws to enforce greater accountability; and questions of *responsibility*, meaning the considerable difficulty to determine a single culprit or 'smoking gun' in complex digital environments made up of a growing number of learning systems and autonomous AI agents.[11] The concluding section then shifted from a descriptive to a more normative stance, calling for a culture marked by new forms of digital "epistemic vigilance" (Sperber et al., 2010) rather than blind data trust, replacing 'trustful by default' with a model where trustees must establish their *trustworthiness* upfront, and trustors have the necessary means and competence to assess these assurances (see Simon, 2016; 2013). Ultimately, a combination of hard law (e.g., the primary rules of the European General Data Protection Regulation (GDPR)), soft law (e.g., nonbinding codes of conduct and attempts at public-private co-regulation),[12] ethically-informed software design (e.g., Apple's recent push for differential privacy),[13] and educational efforts (e.g., raising critical awareness among students but also

[11] For a recent perspective on regulators' struggle to implement better transparency and accountability mechanisms in face of ever more widespread automated decision-making processes, see Wachter et al. (2017).

[12] For more on the dynamics between hard and soft law, see Hagemann et al. (2018); for a discussion of the legal challenges of Big Data with focus on the "secondary rules of the law" at work within the GDPR, see Pagallo (2017).

[13] For a brief overview of Apple's use of differential privacy, see Greenberg (2016); for a more critical discussion, see Greenberg (2017); for wider philosophical reflections on the concept, see Nissenbaum (2016).

the public at large) would be required to meet the challenges posed by Big Data analytics. Instead of a magic silver bullet, the paper thus envisioned a combination of remedies that work on multiple fronts, treating Big Data as a vexing technopolitical problem with no easy solution. As Pasquale (2015: 218) concludes:

"Black box services are often wondrous to behold, but our black box society has become dangerously unstable, unfair, and unproductive. Neither New York quants nor California engineers can deliver a sound economy or a secure society. Those are the tasks of a citizenry, which can perform its job only as well as it understands the stakes."

Interlude: In the area of science and technology, promissory claims are often employed strategically to attract interest and investment. As is the case with Big Data, these claims tend to be overblown and sensationalized, lacking in nuance and detail. This, however, does not make them any less effective. Quite on the contrary, as scholars engaged in the sociology of expectations have pointed out, such "expectations can be seen to be fundamentally 'generative'" (Borup et al., 2006: 285). They "drive technical and scientific activity", "provide structure and legitimation", and "attract[t] the interest of necessary allies" (ibid.: 286; 289). If the promissory rhetoric is successful, the visions become inscribed into actions, texts, bodies, and machines. They "take on substance, becoming materially embedded in structures, routines, systems, matters" (ibid.: 292). They are both performative and normative, enacting specific desired futures that trigger specific real-world actions. In the case of Big Data, the epistemic promises of industry stakeholders (outlined in detail in Paper II) have successfully found their way into the political discourse, spurring the imagination of policymakers around the globe. Paper III takes a closer look at these imaginaries, and the specific political ambitions and agendas they engender.

2.3. Discussion of Paper III - *Tracing Big Data Imaginaries through Public Policy: The Case of the European Commission*

The third paper (Rieder, 2018) presented in this dissertation was a full-length research article published in the Routledge edited volume *The Politics and Policies of Big Data: Big Data, Big Brother?* (Sætnan et al., 2018) Its main aim was to develop a better understanding of the rise of Big Data in public policy, examining official EU publications to investigate the values and visions that guide European policymakers' Big Data rhetoric. More specifically, based on a

grounded theory analysis of roughly 120 documents scraped from the European Commission's website,[14] the text outlined and critically discussed three interrelated narratives that figured prominently in the policy discourse. These were: (I) Big Data as the cornerstone of a thriving data-driven economy, with data analytics expected to increase productivity and boost economic growth; (II) Big Data as a way to transform and improve public services, with data technologies seen as a way to enhance government efficiency and reduce costs; and (III) Big Data as a tool for evidence-informed policy and decision making, with Big Data expected to provide timely, actionable insights that enable authorities to tackle societal challenges with high precision and accuracy. What thus emerged from the empirical analysis was a sweepingly optimistic vision of Big Data and its presumed impacts, the Commission essentially promoting Big Data's widespread application as a major opportunity not to be missed. While this focus on commercial growth and organizational efficiency should not come as a surprise given the institution's historic roots as an "economic community […] among peoples long divided by bloody conflicts" (Treaty of Paris, 1951: 3), the paper sought to dig deeper, citing the recent financial crisis, the pressing need for public legitimation, and a certain longing for certainty in uncertain times as additional factors that may have contributed to the Commission's determined push for Big Data solutions. The Commission's vision, however, entailed a number of problems, and the remainder of the paper listed and discussed some of them: from blind trust in numerical data to the strive for trusted rather than trustworthy technologies (compare Paper II), from the idea(l) of an entrepreneurial, self-managing data citizenry to the adoption of an open-by-default principle for all public sector data. The conclusion was drawn that these concepts and convictions, underpinned by a strong free-market agenda, threaten to undermine the fundamental rights of EU citizens as well as the Commission's own call for responsible research and innovation (RRI).

Readers who have followed the EU's difficult struggle for new data protection laws may be somewhat surprised by these findings. Mager (2017), for instance, contends that in the context of the EU privacy reform, the "economic rationale of the digital single market was

[14] Google's Advanced Search feature, available at https://www.google.com/advanced_search, was used to scrape the European Commission's official website, http://ec.europa.eu/, for documents referring to Big Data. The collected material included communications, speeches, news, and blog entries published between 2012 and early 2017. The qualitative data analysis software NVivo (Version 10 and 11) was used to support the coding, management, and analysis of the data (see Bazeley and Jackson, 2013).

increasingly overshadowed by the fundamental rights discourse staging citizens' rights and freedoms as core European values" (ibid.: 9). As the Commission's 2012 proposal for a new data protection framework clearly states:

> "Data protection is a fundamental right in Europe, enshrined in Article 8 of the Charter of Fundamental Rights of the European Union, as well as in Article 16(1) of the Treaty on the Functioning of the European Union (TFEU), and needs to be protected accordingly." (EC, 2012: 2)

Given this commitment, how can it be that the Commission has nurtured a Big Data imaginary that lauds potential benefits but ignores well-documented harms and risks, that focuses on economic growth but fails to address any of the associated ethical and legal concerns? The paper offered a possible answer to this conundrum, arguing that on EU level, the Big Data discourse appears to run separate from ongoing discussions about data regulation, a certain *state of detachment* that has allowed for the continued use of Big Data as a generic marketing term, structurally divorced from the task of creating fair and sustainable data policies. Importantly, this detachment may be another reason why the term has not yet gone out of fashion, at least not in the political domain. Instead, it remains a key rhetorical resource in what Bos and co-authors (2014) have called the practice of "steering with big words", injecting a hint of tech-euphoria into almost any area of public policy.

Interlude: Much of this dissertation project was dedicated to examining the promises made by Big Data advocates, and the (political) hopes and expectations these promises would generate. More often than not, this would involve dealing with overblown marketing claims, and to consider possible reasons for how these claims could prove so effective, creating a buzz that would spread well beyond IT and business circles. But Big Data is more than just a hype without substance. It is also the umbrella term for a number of technologies and analytical practices that are already in use, applied to ever more contexts, with real consequences for people's individual lives as well as for society at large (e.g., see Ezrachi and Stucke, 2016; O'Neil, 2016; Lyon, 2014). Consequently, Paper IV was a call to arms, aimed at a particular research community that over time has garnered much experience in studying the potential impacts, but also the societal desirability, of new and emerging technologies, from bioengineering to nanotechnology. I am referring, of course, to the field of technology assessment (TA), which has not been especially fast at addressing the challenges posed by Big

Data technologies, but has recently recognized the need for deeper engagement, the theme of the 2018 S.NET conference being "Anticipatory Technologies: Data and Disorientation" (S.NET Organizing Committee, 2018).

2.4. Discussion of Paper IV - *Big Data and Technology Assessment: Research Topic or Competitor?*

The fourth paper (Rieder and Simon, 2017b) was published in the *Journal of Responsible Innovation* as part of the special issue *Into the Wild: Futures and Responsibilities in Technology Assessment*, introduced by Erik Fisher (Fisher, 2017) and guest-edited by Ulrike Bechtold, Daniela Fuchs, and Niklas Gudowsky (Bechtold et al., 2017). Its main idea was to emphasize the importance of the Big Data phenomenon for the technology assessment (TA) community, arguing that the relationship between Big Data and TA is both special and peculiar in the sense that Big Data may be the first research topic TA has dealt with that is not only an object of inquiry, but that also rivals TA in some its core functions. Regarding the former, the paper stressed that TA's experience in multi-, inter-, and transdisciplinary research, together with its expertise in deliberation and consultation at the science-society-policy interface, can help in addressing and successfully dealing with the manifold challenges posed by Big Data. Concerning the latter, the argument was made that Big Data is bound to compete with TA on multiple fronts, including the assessment of public views and visions, means and methods for exploring the future, and the provision of actionable knowledge and advice. Thereby, it was important to highlight that while Big Data and TA may increasingly compete for the same public funds, epistemologically, they tread very different paths, offering contrasting views of what the exploration of public sentiments or the engagement in future-oriented analysis actually entails. While outlining such differences and emphasizing the allure of modern data analytics for public administrations, the paper also sought to sketch a way toward more responsible data-based research and innovation, pondering how to move Big Data under the RRI umbrella. Importantly, while arguing for *critical engagement beyond ELSIfication*[15] and the option to *stop socially undesired innovations* through, for example, temporary or

[15] In short, the acronym ELSI (used in the United States) or ELSA (used in Europe) refers to a range of efforts and formats that examine the ethical, legal, and social implications/aspects of new and emerging technologies. For a concise review of the "rise of ELS programs", see Hilgartner et al. (2017).

permanent moratoria, the paper pleaded for a *coalition of mutual learning* rather than dogmatic opposition, with TA acknowledging the value of data science as a powerful epistemic practice and actively engaging in the development of digital tools and methods that can serve as best practice examples. The paper was thus a call to arms not only in the sense that it asked for TA projects to consider the potential impacts and consequences of Big Data, but also in pointing to the field's need to methodologically expand and add computational capabilities to its own research efforts.

While the methodological expansion may be a longer process with gradual implementation in future research designs, the Germany-based project ABIDA[16] has already demonstrated what a large-scale, multi-year TA study can achieve. Tackling Big Data from different disciplinary perspectives – including ethics, law, sociology, economics, and political science – the project teams have employed a number of research- and dialogue-oriented instruments to foster not only academic but also public discussions about public issues, synthesizing eventual findings into the development of options for political action.[17] Projects such as this underscore the value of technology assessment as a field and practice that in many countries has formed and maintained close relationships with the political domain, potentially playing a key role in the democratic governance and resolution of technoscientific controversies. The paper sought to emphasize this role, but also to highlight that the very same technological phenomenon that now requires TA's attention may also start to rival TA in several of its core competencies, a challenge academics and practitioners in the field would do well to earnestly consider and address. As the paper concluded:

"If done right, modern data analytics could become an ally rather than a competitor to the field of technology assessment, potentially extending the scope, speed, and quality of discourse, dispute, and trend analysis. If disregarded and left to the proprietary discretion of commercial products, however, Big Data may not only grow to challenge the TA community, but could also pose a considerable threat to the core principles of RRI." (Rieder and Simon, 2017b: 11)

[16] The project acronym ABIDA stands for "Assessing Big Data". The project consortium includes teams from six German universities and research centers, with each working group approaching the subject matter from a different disciplinary background.

[17] For a brief outline of the ABIDA project structure, see the official website at http://www.abida.de/en/content/abida-das-projekt [Accessed 20 June 2018].

Interlude: Over the past decades, TA scholars have put much thought into how to improve society's capacity to anticipate and reflexively assess the dynamics and implications of sociotechnical change (e.g., see Guston, 2014). Thereby, the use of science fiction and science fiction-inspired approaches has been discussed as a promising way to grapple with the longer-term implications of technoscientific innovation. According to Miller and Bennett (2008), such approaches can help to "meaningfully evaluate new and emerging technologies and to democratically govern the design and construction of future technological worlds" (ibid.: 598). What makes (science) fiction-based books and films such a valuable resource for broader societal conversation and debate is their considerable public appeal. While academic publications and outreach activities tend to reach only a small number of people, popular fictions may reach millions. The question then is whether disciplines that seek to foster reflexive capacities – from technology assessment (TA) to Science and Technology Studies (STS) to computer and information ethics – can take advantage of this popularity and use science fiction to engage publics more broadly in deliberations about technoscientific futures (see ibid.: 599). Paper V aimed to contribute to this cause, providing an overview of selected *datafictions* that may be used in a variety of contexts to reflect upon the textures and politics of our increasingly digital lives.

2.5. Discussion of Paper V - *Datafictions: Or How Measurements and Predictive Analytics Rule Imagined Future Worlds*

The fifth paper of this dissertation (Rieder and Völker, 2018, forthcoming) has been submitted to the SAGE journal *Futures* and is, at the time of writing, still under peer review. The initial idea for the article came from the observation that while media and academic publications would frequently refer to George Orwell's (1949) novel *Nineteen Eighty-Four* when discussing the prospect of far-reaching digital surveillance (e.g., see Condliffe, 2016), references to other works of fiction were woefully rare. Was the often-cited connection between Big Data and Big Brother really all that narrative fiction had to offer? Based on an extensive literature review, the paper sought to show otherwise, taking a closer look at fifteen *datafictions* – including novels, short stories, films, and TV shows – that may serve as worthwhile discussion pieces when contemplating both the immediate and long-term effects of increased datafication and computerization. Although different in style and story, they all focused on

questions of (algorithmic) power and control, pondering the state of humanity in ever more tech-regulated environments. The fictions were subdivided into five main categories: surveillance (I); social sorting (II); prediction (III); advertising and corporate power (IV); hubris, breakdown, and the end of Big Data (V). The categories were meant to be suggestive rather than definitive, an attempt to cluster works of fiction that are so rich and varied that they could be interpreted and grouped in many different ways. The number of reviewed titles, three per category, was chosen so as to strike a balance between breadth and depth, that is, between providing a basic overview while still being able to briefly outline and comment on each of the selected texts. Conceptually, the paper drew on the notion of *speculative fiction* as a label for a genre that generates "other worlds as a comment upon our own" (Miner, 1991: 150), telling plausible stories that explore real-life issues by pushing them to extremes in strange but imaginable alternative realities. Given their often dystopian nature, such visions can serve as dire warnings, as "dark shadows cast by the present into the future" (Atwood, 2005: 94), stimulating reflection and critical thinking by providing "rich and complex avenues for reading and rereading the world" (Thomas, 2013: 4). The concept of speculative fiction guided our literature review in important ways, allowing us to discard fictions that would not bear a sufficient resemblance to the universe as we currently know it while still providing the flexibility to include fictions that are so close to home that they could hardly be considered sci-fi at all. The notion of *datafictions* was then introduced as an umbrella term for speculative stories that deal with the datafication of society in both imaginative and imaginable ways,[18] thus adding a clear thematic focus to our analysis. As mentioned, the principal goal was to gather and present striking fictional examples of data power in action, so as to contribute to the cultivation of a *critical imaginative gaze* that can support a more proactive engagement with the potential consequences of Big Data applications. However, in addition to honing reflexive capacities in students, scholars, and practitioners, such fictions can also be used to investigate how specific technologies are culturally imagined and what kind of futures are considered plausible given current implementations and trajectories. They thus should be recognized as fruitful sites of analysis for professional technology assessment, conceptually no less credible and legitimate than, for example, mapping exercises, scenario

[18] The only exemption from this requirement was the inclusion of Mary Shelley's (1818) novel *Frankenstein*, which mainly serves as a thoughtful reminder that technologies, once created, must continually be checked and cared for in order not to run into any nasty surprises along the way.

development, or Delphi surveys. As the provided overview of datafictions was neither meant to be comprehensive nor conclusive, the article ended with a call for shared collections and repositories, a public resource for anyone with interest in the subject matter.[19]

A final thought: When working through the speculative fictions, it became increasingly clear that on a technological level their imagination was outpaced by the actual speed of innovation we are witnessing today. For instance, Google's 2018 keynote presentation of an AI assistant calling and making table reservations at a restaurant, thereby successfully conversing with a strongly-accented employee, seemed technologically more radical and disruptive than almost any of the visions encountered in the sci-fi literature (see Welch, 2018). As briefly alluded to in the paper, however, the strength of narrative fiction does not necessarily lie in the detailed depiction of cutting-edge technologies, but in creating believable worlds where the consequences of sociotechnical change are shown from a protagonist's personal perspective, thus immersing the reader in an alternative reality and allowing her to become emotionally invested in a given situation or event. As a consequence, ethical inquiries into the potential pitfalls of datafied societies may benefit from an approach where fictional stories are discussed alongside real-world tech developments, providing some insight into recent trends and developments while offering a creative toolbox for pondering related risks and implications.

Postlude

As mentioned, the five articles outlined above and included in full length below form the core of this thesis. Whilst differing in terms of their specific focus and research design, they find common ground in seeking to address Big Data as a complex sociotechnical phenomenon that is more than the sheer combination of hardware and software power. Rather, the term's long-lasting prominence connects to a culture of mechanical objectivity that has grown and expanded significantly over the past two centuries (Paper I), to a powerful industry's relentless efforts to market and sell their latest products (Paper II), and to the specific needs

[19] An example for such a resource is the *Internet Speculative Fiction Database* at http://www.isfdb.org [Accessed 20 June 2018]. This website provides an open bibliographic archive that allows anyone with an account to add and freely tag new literature entries. Unfortunately, at the time of writing, the search term "Big Data" returns only a few results, but the technology to improve upon this situation is already in place.

of a struggling political system in times of public distrust and crisis (Paper III). The vision that Big Data offers may begin to challenge established modes of prognosis and analysis (Paper IV) as well as lead to situations of algorithmic regulation and control that not long ago would have seemed like far-off fiction rather than a plausible near-future scenario (Paper V).

While alluding to a range of Big Data-related problems and issues – from blind trust to issues of opacity and accountability – and stressing the need for greater awareness of the potential social, ethical, and legal implications of widespread data analytics, the articles spent comparatively little time discussing the detrimental effects that have *already* started to manifest in the wake of the Big Data era. These tangible effects, however, are exactly why the rise of modern data analytics requires society's collective attention. Consequently, before the articles themselves take center stage and in an effort to provide a better sense of *what is actually at stake*, the thesis continues with a brief overview of the substantial risks and harms that Big Data and its application can engender (Chapter 3). This is then followed by a forward-looking conclusion that rounds off the argument for why Big Data can indeed be seen as *the vision that would not fade* (Chapter 4).

3. What Is at Stake: A Brief Overview of Big Data Risks and Harms

In recent years, much has been said about the perils and pitfalls of an ever more widespread application of Big Data analytics. From civil rights groups to data protection authorities, from investigative journalism to academia-based critical data studies, from the world of fiction to non-fiction books and documentaries, warnings of the detrimental consequences of increased datafication and algorithmic decision making have come from numerous sources and directions. Thereby, the critique has mainly focused on three problem areas: issues of privacy and surveillance; problems of bias and discrimination; and the perils and drawbacks of virtual competition. This chapter briefly outlines each of these problem areas, providing a concise overview of some of the key concerns associated with the rise of Big Data analytics. As such, the chapter does not present original research, but draws on a few selected sources that have had considerable impact on the debate. As indicated, the goal is not to provide a comprehensive review of issues and challenges, but to foster a sense of awareness that sets the scene for the ensuing articles, a testament to their timeliness and relevance in the present digital era.

3.1. Big Data: Issues of Privacy and Surveillance

In mid-2013, the documents[20] leaked by former NSA contractor Edward Snowden provided some insight into the extent to which U.S. intelligence agencies had collected personal information on both domestic and foreign citizens. While extensive state surveillance is by no means a new phenomenon and evidence of NSA practices had surfaced before (see Bamford, 2008), the Snowden revelations went beyond what was deemed possible in a Western democracy. Instead of targeted surveillance of individual suspects, the NSA and its partner organizations engaged in mass surveillance of ordinary, law-abiding citizens, obtaining vast amounts of personal data from a range of different sources. These sources include, but are not limited to, the interception of data-in-transit by tapping the fiber optic cables that run between and across countries and continents, installing spyware on personal computers,

[20] The leaked documents are archived and publicly available in the Snowden Digital Surveillance Archive at https://snowdenarchive.cjfe.org [Accessed 20 June 2018].

servers, and routers, and gleaning user data from the servers of major U.S. IT companies such as Apple, Google, Facebook, and Microsoft.[21] The latter source is particularly important because it indicates that in the 21st century it is no longer sufficient to merely speak of a top-down *surveillance state* or totalitarian *surveillance society*, but that one must take into account the development of a comprehensive *surveillance culture* (see Lyon, 2015). After all, by willingly sharing our personal information online and using the 'free' tools of private for-profit companies to do so, "[s]urveillance is not just practiced *on* us, we participate *in* it" (ibid.: 3). This is also why references to Orwell's *Nineteen Eighty-Four* are generally insufficient to grasp the very nature of Big Data surveillance.[22] As Snowden himself once stated: "The types of collection in the book – microphones and video cameras, TVs that watch us – are nothing compared to what we have available today. We have sensors in our pockets that track us everywhere we go. Think about what this means for the privacy of the average person." (cited after Pittas, 2013) Snowden's point here is, of course, technical, a reference to the rapid pace of innovation over the past few decades; but it is also cultural in the sense that it alludes to people's practice of habitually taking their smartphones everywhere they go. What is more, by interacting with the system and using services such as social media, users – sometimes knowingly, sometimes not – actively volunteer their data to providers that will treat this information in more or less privacy-protecting ways.[23] However, if the monetization of the data is part of a company's business model,[24] the result is often what Zuboff (2015) has referred to as "surveillance capitalism", that is, a "logic of accumulation" where data are routinely captured, aggregated, analyzed, packaged, and sold to the highest bidder. Yet, as Zuboff highlights, the power structure of this new surveillance architecture does not correspond to that of Orwell's Big Brother, where force is exercised through centralized command and control, but rather to what she calls *Big Other*, a "ubiquitous networked institutional regime" that records and commodifies everyday life, yielding "radically distributed opportunities for observation, interpretation, communication, influence, prediction, and ultimately modification of the totality of action" (ibid.: 82). And further: "If

[21] All these surveillance techniques are detailed in the third chapter of Glenn Greenwald's (2014) book *No Place to Hide: Edward Snowden, the NSA and the Surveillance State*.
[22] Also see Paper V in this thesis.
[23] On the concept of volunteered data, see Kitchin (2014b: 93).
[24] This is for instance the case when data are used for targeted advertising or directly sold to data brokers. For a historically grounded introduction to the topic, see Turow (2011).

power was once identified with the ownership of the means of production, it is now identified with ownership of the means of behavioral modification" (ibid.). What Zuboff describes here in more abstract terms is tech companies' ever-increasing proficiency in tracking, anticipating, and steering human action, applying this knowledge so as to generate clicks, views, and other forms of profit-making engagements. But the proliferation of such monitoring regimes poses a number of societal challenges and threats: First, as Zuboff points out, surveillance capitalism does not simply lead to an *erosion* of privacy rights, but rather to their *redistribution*. While covert data capture[25] deprives individuals of the privacy producing choice to either share or withhold information, companies such as Google or Facebook hide their operations behind a veil of secrecy, seeking exemption from both public oversight and regulation. Second, by imposing automated systems of reward and punishment,[26] surveillance capitalism changes the very nature of contractual relationships, replacing trust in people's solidarity and the rule of law with new techniques of algorithmic judgment and control. In an effort to mitigate the uncertainty caused by human fallibility, freedoms achieved by the "consensual participation in the values from which legitimate authority is derived […] are traded in for the universal equivalent of the prisoner's electronic ankle bracelet" (ibid.: 81). Insurance prices, for instance, can be tied to the monitoring of one's lifestyle choices (health care) or the scoring of one's personality based on social media data (car insurance), in an effort to calculate risk and incentivize behavioral change (e.g., see Ralph, 2017). Which leads to a third consequence. As Zuboff highlights, the invasive power of surveillance capitalism engenders a kind of *anticipatory conformity* where a "vortex of stimuli" nudges people into compliance with the "financial and, or, ideological interests that imbue Big Other" (Zuboff, 2015: 82). In such a world, human autonomy and psychological self-determination is a cruel illusion. Instead, people are "reduced to a mere animal condition, bent to serve the new laws of capital imposed on all behavior through an implacable feed of ubiquitous fact-based real-time records of all things and creatures." (ibid.) Society becomes dazed and tranquillized, thrown into submission by the orders and ideals of a new invisible hand. Although Zuboff's article mainly focuses on the practices of private tech companies, the Snowden revelations as well

[25] As the past years have shown, Big Data corporations' modus operandi is often to quietly advance and expand their activities until resistance is met. If there is public backlash, apologies are offered, but the take-without-asking mentality continues (e.g., see Fowler and Esteban, 2018).

[26] For a well-made fictional treatment of the potential ramifications of such an automated system, see the *BlackMirror* episode "Nosedive" outlined in Paper V.

as public-private collaborations such as Google's involvement in the Pentagon's Project Maven (see Conger and Cameron, 2018) indicate that there is a confluence between commercial and state interests. As Bruce Schneier (2015: 78) contends in his book *Data and Goliath*: "Corporate surveillance and government surveillance aren't separate. They are intertwined; the two support each other. It's a public-private surveillance partnership that spans the world." More recently, the abolishment of the "safe harbor" agreement between Europe and the U.S. through the European Court of Justice (see Gibbs, 2015) as well as the EU's new Data Protection Regulation have shown that this partnership has at least some legal limits,[27] which, however, does not seem to impede law enforcement and intelligence agencies' push for Big Data-enabled mass monitoring solutions. Examining the "rapid and widespread adoption" of Big Data techniques by government agencies, Lyon (2014) identifies three key ways in which the commitment to Big Data changes the nature of surveillance practices: First, surveillance is becoming more *automated* in the sense that data are collected and analyzed routinely, creating detailed profiles of individuals that are being mined for statistical patterns and correlations. The combination of powerful new software and cheaper storage makes the constant digital surveillance of entire populations not only technologically possible but, in times of tight budgetary constraints, also economically attractive. Second, Big Data practices change the "tense" of surveillance by attempting to *predict* and *preempt* future developments. In the context of predictive policing, for instance, profiled individuals are not merely assessed in terms of who they 'are' and what they have previously done, but also with respect to who they might become and what they probably will do (see Rieke et al., 2014). If paired with automated forms of decision making, such computational forecasts may impair civil liberties and undermine due process rights, especially if citizens cannot push back and contest the charges and suspicions. Third and finally, since Big Data applications are used in an ever-increasing variety of areas, one can observe processes of *adaption* where tools that were developed for one purpose are 'borrowed' and used for another.[28] However, the level of accuracy that for instance suffices in the context of online marketing may not be acceptable

[27] Moreover, in the aftermath of the Snowden disclosures, the tech giants sought to shed their image as NSA collaborators, upgrading their devices with stronger security and pushing back against government requests to unlock encrypted data (see Hautala, 2013).

[28] The predictive policing software PredPol, for example, forecasts the likelihood of future crimes in a given area based on an algorithm that was originally designed to predict earthquake aftershocks. For a critical discussion, see Lum and Isaac (2016).

in the context of law enforcement or criminal justice as the harms of false positives may be even more severe in the latter case. What is more, the fact that proprietary analytical systems are often thoroughly black-boxed and their workings not even well understood by the responsible authorities themselves casts a worrying light on the legitimacy and democratic nature of the use of Big Data technologies for surveillance and security purposes (see Ferguson, 2017: 1166). As Lyon (2015: 136) concludes: "One kind of freedom is to pursue technically supported ways of seeking human security of all kinds. But when the product of such freedom curtails the enjoyment of everyday freedoms to speak or associate with others, or to dissent from government policy, or just to live without fear, something has gone badly wrong."

3.2. Big Data: Problems of Bias and Discrimination

In 2016, mathematician and former Wall Street quant Cathy O'Neil published her widely successful book *Weapons of Math Destruction*. Having witnessed the destructive power of financial algorithms first hand, the book offers a rich exploration of what O'Neil calls "the dark side of Big Data" (2016: 13). The book goes through a series of examples – from teacher evaluations and U.S. college rankings to online advertising and crime prediction to human resources, credit scoring, health care and insurance – outlining the potential harmful effects of computational models on a case-by-case basis. In a nutshell, the main points of O'Neil's argument can be summed up in the following way:

Though often perceived as fair and objective, algorithmic models are based on the choices of fallible human beings. Rather than providing a disinterested picture of reality, they function as "opinion[s] formalized in code" (see ibid.: 53), reflecting the assumptions and ambitions of their respective creators as well as, in the case of machine learning, the biases and prejudices of society at large. Consequently, computational models do not guarantee a more even-handed treatment, but can perpetuate social injustices camouflaged by a *façade* of scientific certitude and rigor. While discrimination against protected classes – for example, sex, race, or disability – is, in theory, punishable by law, entire industries have formed that reward the rich and punish the (unprotected) poor, creating toxic feedback loops. For instance, if a responsible, hard-working person loses her job, this may negatively affect her credit score, making it more difficult and costly (due to higher interest rates) to get a loan. In

further consequence, the belief that bad credit correlates with bad job performance makes it even less likely to find new employment. As O'Neil explains: "Joblessness pushes them toward poverty, which further worsens their scores, making it even harder for them to land a job. It's a downward spiral." (ibid.: 7) Similar mechanisms can be observed in the criminal justice system. If a person is scored "high risk" due to, e.g., unemployment, living in a poor, high-crime neighborhood, or having friends and family that had run-ins with the law, this may affect the length of her sentence or whether she qualifies for probation. However, since the time spent in prison increases the chance that a person upon release cannot find work and requires social welfare, the likelihood that the person commits another crime and returns to prison increases as well, creating a destructive cycle that the algorithmic model helps to sustain. If we look more closely, it should become apparent that such a model does not only discriminate against low-income classes, but also, using the zip codes of highly segregated communities as proxy data, against ethnic minorities. As *ProPublica* investigations have shown (see Angwin et al., 2016), the result is a racially biased system that overestimates the recidivism rate of African Americans (generating comparatively more false positives) while underestimating it for Caucasians (generating comparatively more false negatives), systematically privileging one ethnic group while disadvantaging the other. If affecting the court's sentencing decision, such biased models can contribute to trapping specific demographic groups in a continuous downward spiral, according to O'Neil a "signature quality of a [Weapon of Math Destruction]" (2016: 27).

What exacerbates the situation is the fact that these technologies tend to be opaque, their algorithms protected from scrutiny by the expansive nature of trade secret law (see Moore, 2017). This secrecy, however, makes it much harder to question a result and contest a score. As O'Neil (2016: 146) emphasizes, part of the technology's fearsome power is that there is "little recourse to complain, much less correct the system's error." So even if unfair treatment is detected, which due to the veil of secrecy rarely happens, there is no chance to provide feedback and "set the system straight."[29] (ibid.) This also applies to erroneous personal information collected and stored in the massive databases of the data broker

[29] In this context, consider Article 17 ("right to erasure", also known as the "right to be forgotten") and Article 18 ("right to restriction of processing") of the new European General Data Protection Regulation (EC, 2016), which can be seen as an attempt to implement a feedback structure that bolsters Europeans' privacy rights.

industry.[30] If such records are sold to other companies, they may affect whether someone qualifies for a job or a loan or even, as O'Neil details, admittance to a senior living center (see ibid.: 151). Given the sheer number of data brokers selling customer profiles, faulty data can be difficult if not almost impossible to correct. The ones who have a way out are, once again, the financially better off. As O'Neil stresses, "privacy, increasingly, will come at a cost", a "luxury that only the wealthy can afford." (ibid.: 170) Trackers, she argues, will likely become the norm, and if someone wishes to opt out of the data economy, they will have to pay a premium. Google's services, for instance, are mostly free to use, but financed through targeted advertising based on continuously gathered user data.[31] In order to enjoy an ad-free YouTube experience without using ad-blocking technology that many Internet users are still unaware of,[32] one can only opt to purchase a YouTube Premium subscription which, after a trial period, costs €11.99 per month.[33] Something similar can be observed in the area of online news. After years of struggling with dwindling advertising revenue, the website of the Austrian newspaper *Der Standard*, for example, no longer displays any content if an activated ad blocker in the user's browser is detected, giving readers the choice to either deactivate the ad-blocking software on its site, thereby accepting both ad delivery and third-party tracking, or to pay a subscription fee of, at the time of writing, 6 Euros per month. This type of choice may serve as a reminder of the well-known adage: "If you are not paying for it, you're not the customer; you're the product being sold", and the financially better off may find it quite a bit easier to pay the fees and safeguard at least some of their privacy when browsing the Web.

Another issue that O'Neil only briefly touches upon is that an effective regulation of discriminatory data mining is difficult to achieve, at least under current law. In an often-cited paper, Barocas and Selbst (2016) have made the case that even though Big Data applications can "affect the fortunes of whole classes of people in consistently unfavorable ways" (ibid.:

[30] For critical a review of the data broker industry, see Rieke et al. (2016) and Office of Oversight and Investigations (2013).

[31] For more on YouTube's targeting methods, see the "About targeting for video campaigns" help page at https://support.google.com/youtube/answer/2454017?hl=en [Accessed 20 June 2018].

[32] According to 2016 survey by Midia Research, 41% of the 3,600 respondents from the U.S., the U.K., Brazil, Australia, France, and Sweden were aware of ad blocking technology. But among those who were aware, 80% use ad-blocking software on desktop devices and 46% do so on smartphones (see Moses, 2016).

[33] For more details on the YouTube Premium service, formerly known as YouTube Red, see https://www.youtube.com/premium [Accessed 20 June 2018].

673), the logic of the data mining process, which is "always a form of statistical (and therefore seemingly rational) discrimination" (ibid.: 677), creates several hurdles to finding disparate treatment and impact liability, leading the authors to conclude that "existing law largely fails to address the discrimination that can result from data mining" (ibid.: 675). Unfortunately, there is no simple remedy, as a "standard that holds companies liable for any amount of theoretically avoidable disparate impact is likely to ensnare all companies." (ibid.: 729) As Barocas and Selbst stress, such standards would counsel against using data mining altogether, creating a "perverse outcome, given how much even imperfect data mining can do to help reduce the very high rates of discrimination in [traditional, non-computerized decision making]." (ibid.: 730).[34] From a regulatory perspective, the challenge, then, is to create a legislative framework capable of safeguarding civil rights and liberties without discouraging data-driven innovation as such. The EU's General Data Protection Regulation constitutes such an approach, but it remains to be seen whether the principles laid out in this new set of law are capable of "fully engag[ing] with the impeding Big Data tsunami" (Rubinstein, 2013: 74), affecting not only how data are collected and stored, but also how they are processed and used for automated decision making (see Goodman and Flaxman, 2017).

3.3. Big Data: The Perils and Drawbacks of Virtual Competition

In April 2015, the European Commission announced that it had sent a Statement of Objection to Google, alleging that "the company has abused its dominant position in the markets for general internet search services in the European Economic Area (EEA) by systematically favoring its own comparison shopping product in its general search results pages." (EC, 2015) Margarethe Vestager, the European Union's Commissioner for Competition, added that she had also launched a formal antitrust investigation of Google's conduct concerning mobile operating systems, apps, and services, declaring that she wants "to make sure the markets in this area can flourish without anticompetitive constrains imposed by any company." (ibid.) In

[34] This thesis is mainly focused on the problems and challenges associated with the rise of Big Data analytics. This, however, does not mean that there are no advantages to be gained from its application. For a balanced review of Big Data's potential to increase access to credit for the financially underserved, uncover and possibly reduce employment discrimination, and create educational opportunities for the students who most need them, see the White House publication *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* (White House, 2016).

June 2017, the Commission fined Google €2.42 billion for "abusing dominance as search engine" and "breaching EU antitrust rules", with Commissioner Vestager noting that the company had "denied other companies the chance to compete", depriving "European consumers [of] a genuine choice of services and the full benefits of innovation." (EC, 2017) At the time of writing, the outcome of the Commission's antitrust investigation of Google's mobile operating system, Android, is still pending, but reports suggest that the penalty could even be higher than the previous one (see Scott, 2018). Given that the European Commission is not the only regulatory body pursuing Google for anticompetitive conduct – with similar investigations having been carried out in Russia (see Scott, 2016), India (see Kalra and Shah, 2018), Brazil (see Leahy, 2013), and, infamously leading to no formal charges, the U.S. (see Mullins et al., 2015) – what is the background of these actions against a company that only recently removed its unofficial "don't be evil" motto from its code of conduct (see Conger, 2018)?

In their book *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*, Ezrachi and Stucke (2016) contend that while digital markets have traditionally been thought to contribute to increased innovation and well-being by reducing search costs, lowering entry barriers, and improving information flows – thereby, in theory, creating a more dynamic and transparent marketplace in which the best product triumphs – at a closer look, one can identify the operations of „an increasingly well-oiled machine that can defy the free competitive forces we rely on." (ibid.: vii) Thus, rather than reducing market concentration and limiting the power of large corporations, the data-based, algorithm-driven economy has given rise to a relatively small number of new platforms and ecosystems that dominate vast parts of their respective field(s) of operation. For instance, while Google dominates the search engine market (with 86.3 percent of global desktop market share in April 2018)[35] and, together with Facebook, accounts for over 60 percent of global online ad revenues in 2017,[36] Amazon's U.S. e-commerce market share is expected to rise from 37 percent in 2017 to 50

[35] See Statista "Worldwide Desktop Market Share of Leading Search Engines from January 2010 to April 2018" at https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/ [Accessed 20 June 2018].

[36] See Statista "25 Percent of Global Ad Spend Goes to Google or Facebook" at https://www.statista.com/chart/12179/google-and-facebook-share-of-ad-revenue/ [Accessed 20 June 2016]. Excluding China, the digital duopoly's market share was expected to reach 84 percent in 2017 (Garrahan, 2017).

percent by 2021.[37] What these three companies – which, together with a few other tech giants, belong to the most valuable companies in the world[38] – have in common, is that they are *masters of Big Data*, meaning that they not only own and receive a lot of data, but that they also have the ability to capitalize on these data through analytical means. Yet the fact that large parts of the digital domain are controlled by only a small number of corporate players endangers competition by (a) enabling new forms of algorithmic collusion that threaten to leave antitrust law behind,[39] by (b) allowing for "almost perfect" price and behavioral discrimination that may not only negatively affect individual well-being (see previous section), but also thwart smaller rivals' competitiveness, and by (c) creating complex "Frenemy" dynamics where "super-platforms" such as Microsoft's Windows, Apple's iOS, or Google's Android operating system set the rules of the game, determining who can join the platform and which apps are featured in the respective app store (see ibid.: 35ff, 83ff, 145ff). Importantly, the market power of large Internet companies increases through a variety of network effects – in essence, more users create more data that can be used to improve products and services – which raises entry barriers and makes it less likely that any newcomer will be able to compete. As Ezrachi and Stucke (2016: 135) put it with reference to the search engine market:

"[M]ore users generate more search queries, which generate more trial and error, which yields better search results, which attracts more users and advertisers to the search platform, which enables better profiling of users and greater likelihood of users clicking on the ads, which generates more advertising revenue to enable the search engine to offer even more free services […]."

As a result, the "big web-aggregators become bigger, […] occupying a strategic position in the distribution channel" (ibid.), which increases their power to dictate prices and, through a

[37] See Statista "Projected Retail E-Commerce GMV Share of Amazon in the United States from 2016 to 2021" at https://www.statista.com/statistics/788109/amazon-retail-market-share-usa/ [Accessed 20 June 2018].

[38] See Statista "The 100 Largest Companies in the World by Market Value in 2018 (in Billion U.S. Dollars)" at https://www.statista.com/statistics/263264/top-companies-in-the-world-by-market-value/ [Accessed 20 June 2018].

[39] Whereas Apple's conspiracy with major publishing houses to raise and fix e-book prices can be understood as a more traditional form of collusion that is punishable by law (see Kastrenakes, 2016; Robertson, 2013), legal prosecution becomes more difficult in the case of algorithm-driven online systems where evidence of clear intent – the proverbial "smoking gun" – often cannot be found. For a detailed discussion of this problem, see chapters 6 to 8 in Ezrachi and Stucke (2016).

broad range of strategies, distort competition.[40] Thus, while superficially competition in digital markets may seem robust, it is in fact controlled by a few giants who cement their leadership by either "acquiring or blocking innovation or entry that might potentially undermine [their] dominance." (ibid.: 175) Consequently, the often-repeated dictum that the "competition is just one click away" (Schmidt, 2015) is, while technically correct, misleading. In increasingly algorithm-driven digital environments, the availability of high-quality training data has become a key resource for product development, which gives established companies with a large, active user base a significant advantage over newcomers, effectively raising market entry barriers (see Radinsky, 2015). If, however, a tech startup does find a way to grow and succeed, it is far more likely that the company will end up being bought rather than ever getting the chance of becoming a serious rival.[41]

What makes the situation even more challenging is that the market power of large Internet companies is not limited to their initial core business. As recent years have shown, the giants of the Web are constantly seeking to expand their operations by diversifying into new markets. Alphabet Inc., for instance, not only operates the world's most popular search engine (Google Search), develops the most widespread mobile operating system (Android), and runs a suite of highly profitable online advertising services (AdWords, AdSense), the company also produces hardware (from Pixel smartphones over Chromebook laptops and Google Home speakers to Nest thermostats), builds mapping (Google Maps), storage (Google Drive, Google Cloud Platform), productivity (Gmail, Google Docs), streaming (YouTube), and payment services (Google Pay), and currently heavily invests in AI research (DeepMind) including smart assistants (Google Duplex) and self-driving cars (Waymo). At first glance, it may be surprising that the company that once developed the PageRank algorithm is now one of the leading forces behind the autonomous car movement. Are these not very different areas of research and expertise? Looking a bit more closely, however, it becomes clear that much of what is needed to build a great search engine also comes in handy when building a

[40] Google's alleged favoring of its own services in search results alluded to in the beginning of this section is one example of such a competition-crushing strategy. The Amazon-Hachette dispute, in which the online retailer temporarily removed the pre-order buttons from forthcoming titles of the French publishing house, in turn illustrates the kind of negotiating power that 'gatekeeping' online platforms in two-sided markets can wield (see Doctorow, 2014).

[41] For a graphical overview of the largest acquisitions by one of the 'Big Five', that is, Alphabet (Google), Apple, Microsoft, Amazon, and Facebook, see Rowley (2018).

navigation system for self-driving cars. For instance, such cars "need up-to-the-minute maps of every conceivable roadway to move" (Bergen, 2018) and Alphabet, through Google Maps and related services, has already gained vast experience in the mapping business. As a company spokeswoman notes: "We've built a comprehensive map of the world for people and we are working to expand the utility to our maps to cars." (ibid.) If one takes into account other 'asset classes' such as algorithmic competence, tried-and-tested logistics, and various economies of scale, it becomes evident that Alphabet is not a surprising but rather a quite obvious contender in the evolving self-driving car market. As the computerization of ever more aspects of human life progresses, it seems likely that large tech firms will continue to move and expand into these new areas, using synergies between their established and their future business to further consolidate their market power. What can thus be observed is an emerging "digitized hand" that gives rise to a suite of newly possible anticompetitive behaviors that threaten to undermine free market forces and pose difficult questions for regulators around the globe (see Ezrachi and Stucke, 2016: viii).

4. Final Thoughts and Outlook

What should have become clear by now is that Big Data, as understood here, is more than a short-lived fad. While the term itself may eventually fade away, the vision that has fueled its rise will not. To gather and analyze 'data' is not just the craze of the season, it is one of the defining features of modern civilization ever since the scientific revolution. As Hacking (1990: 1) writes in *The Taming of Chance*, over the course of the 19th century, "[s]ociety became statistical." But statistics need data to work, and the recent "data explosion" (Holmes, 2017) has redefined the limits of what can be traced, tracked, and subjected to computational analysis. To put it in simple terms, over recent years, Big Data has been a crack-like infusion to a *mathematical science* that builds on the work of Carl Friedrich Gauss, Pierre-Simon Laplace, Adolphe Quetelet, and other contemporaries (e.g., see Ambrose, 2015), as well as to a *machine-based industry* that goes back to Herman Hollerith's electromechanical punched card tabulator, the bedrock technology of a company that, through a series of mergers, would later become IBM (e.g., see Driscoll, 2012). From this perspective, Big Data's story is one of continuity that begins in the late stages of the mechanical age and currently culminates in the early 21st century digital world. But this story is nowhere near over. In the upcoming years, Big Data-based applications will continue to change the face of many professions as well as people's lives in general. In purely scientific terms, the advances will probably be impressive, with data-crunching technologies able to exploit the information flood in ever more sophisticated ways. From a sociotechnical point of view, however, these innovations pose a number of problems and challenges, threatening privacy, social justice, and consumer welfare (see previous chapter). Unregulated, these problems are bound to increase. In the future, people will leave even more digital footprints that can be mined and analyzed, revealing minute details about our daily lives. Clearly, such data mining techniques can also be used to reveal patterns of injustice and unfairness, potentially helping to prevent discrimination in an already biased world (e.g., see White House, 2016). But this is often neither the goal nor the outcome of this kind of analysis. As historian Melvin Kranzberg (1986: 545) famously maintained: "Technology is neither good nor bad; nor is it neutral." While technologies have certain affordances – that is, specific possibilities for action – it is society that ultimately determines their application and impact. Unfortunately, in the case of Big Data analytics, one cannot only see the potential for harm, but actual harms already manifesting themselves

within and across society. From this perspective, Big Data's story is one of rupture and discontinuity because changes in the amount and nature of data being collected have engendered a "qualitative shift in how data can be analyzed, to what ends data and analytics can be employed, and who has the power to utilize and extract value from data." (Kitchin, 2014b: 176) As a result, Big Data poses a series of threats to people's fundamental rights and freedoms[42] and extends processes of algorithmic decision making to an ever-greater number of application areas. It alters the power relations between, first, government and industry, and, second, between citizens, on the one hand, and both public and private entities, on the other. From a societal and democratic perspective, the question then remains how best to respond to the many challenges posed by Big Data. This is what the remainder of this chapter aims to address.

First off, I believe there needs to be acknowledgment of the fact that when it comes to the problems and challenges of Big Data, no single solution is likely to suffice. Big Data is not simply a technological issue, but a phenomenon that is deeply entwined with human history and culture. As mentioned – and argued in greater detail in Papers I and III – the appeal of quantification and mechanical objectivity is not new, but has gradually grown with the needs, pressures, and complexities of modern civilization (see Daston and Galison, 2010; Desrosières, 1998; Porter, 1995). Thus, contemporary society's trust *in* and reliance *on* numbers and figures is tied to a continually increasing demand for efficiency, predictability, and control. But to aim for such goals is not a value-free aspiration. Quite on the contrary, in the context of Big Data, notions such as efficiency, predictability, and control are bound to clash with established rights and values such as privacy, autonomy, and non-discrimination. Which leads to an important point: at the most fundamental level, issues around Big Data are grounded in specific economic and political ideologies, their respective norms and values. But if a system is geared toward profit and efficiency rather than privacy and non-discrimination,

[42] Article 8 of the *Charter of Fundamental Rights of the European Union*, for instance, holds that "everyone has the right to the protection of personal data concerning him or her" and that such "data must be processed fairly for specified purposes and on the basis of consent of the person concerned" (European Union, 2012). Ever since the Safe Harbor hearings, when a European Commission lawyer told an EU judge to "consider closing [his] Facebook account" because Brussel executives were unable "to guarantee 'adequate' safeguards" (Nielsen, 2015), it should have become clear that the rights guaranteed in the Charter are very much in jeopardy. For a more detailed discussion of how Big Data threatens to undermine European law, see Rubinstein (2013).

no "technological fix"[43] will make much of a difference. What is needed, instead, is serious *public debate* about ethical acceptability, social desirability, and the key values a community wishes to uphold (see von Schomberg, 2013; Habermas, 1964). This, however, must be paired with a governmental commitment to protect these values, put them into legislation, and enforce them. The recent European General Data Protection Regulation, for instance, is the outcome of a long and difficult negotiation process where political powers have agreed upon a legal framework that – by providing new rights to consumers and imposing additional obligations on data controllers – may disrupt and transform parts of the sprawling data market (e.g., see Manthorpe, 2018; Rose, 2017). Certain EU member states such as Austria, however, have already passed laws that could undermine the enforcement of the new Regulation (see Nielsen, 2018; Sokolov, 2018), which from an analytic point of view only adds to the insight that a responsible governance of Big Data is ultimately a question of sufficient *political and institutional will*. Academia can – and should – of course contribute to this conversation in multiple ways:

Raising Awareness. In a recent *New York Times* op-ed, Cathy O'Neil (2017) critically notes that "our main source of information on the downside of bad technology […] is the media" while the ivory towers of academia "keep ignoring tech". Having spent several years working with scholars from fields such as Science and Technology Studies (STS), computer and information ethics, media studies, and human-computer interaction (HCI), I do not share this impression. Quite the contrary, I would argue that recent times have seen a sharp increase in research on the social and cultural consequences of modern data analytics, as evidenced by dedicated new journals (e.g., SAGE's open access journal *Big Data & Society*) and conferences (e.g., the bi-yearly *Data Power Conference*). At the same time, and as a reaction to the growing importance of computer technology in everyday life, technical universities have started to incorporate courses that foster critical reflection on IT into their curricula, a wave of STS-based hirings at the TU Munich and the establishment of the Research Group for Ethics in Information Technology (EIT) at the Departments of Informatics at Universität Hamburg being two telling examples. Educational efforts beyond the traditional ambit of the social sciences and the humanities has thus become an important means of awareness creation among future generations of data scientists and software developers but

[43] For more on the origins of the notion of "technological fix" as "social cure-all", see Johnston (2018).

also, as seen at the IT University of Copenhagen, among students of business informatics and digital innovation. In addition, researchers concerned with the potential pitfalls of Big Data analytics and algorithmic decision making often seek to communicate beyond the confines of the ivory tower, actively striving to engage with representatives from industry and government as well as with civil society at large. Last but not least, social media has arguably contributed to scholars' public visibility, with academics such as Frank Pasquale, Professor of Law at the University of Maryland and author of above-cited *The Black Box Society* (2015), maintaining prolific Twitter feeds with many thousand followers.

In-Depth Research. One reason why Cathy O'Neil is critical about the media's role in shaping the Big Data discourse is that "much of what should concern us is more nuanced and small scale – and much less understood – than what we see in the headlines." (O'Neil, 2017) While there are plenty of great examples of excellent data journalism (e.g., see Angwin et al., 2016), and Diakopoulos' (2015) work on "algorithmic accountability reporting" makes a strong case for the importance of journalistic watchdogging, I would agree with O'Neil that much of the news coverage on algorithmic power has been sensationalistic and partisan, often lacking the kind of nuance and depth that the complex subject matter deserves. Deep and nuanced analysis, however, is something that academia is ideally able to provide, with articles such as Barocas and Selbst's (2016) *Big Data's Disparate Impact*, which examines whether discriminatory data mining is likely to generate liability under Title VII of the US Civil Rights Act of 1964, being a prime example of how measured, detailed, and refined a critical discussion of algorithmic bias can actually be. Yet in order for such rigorous research to be produced, academia needs sufficient resources and suitable institutional structures. In the context of Big Data analytics, cross-disciplinary work will arguably become increasingly important, not only between fields such as ethics, law, and sociology, but also between these fields and the computer sciences. As a result, interdisciplinary research centers exploring the interplay between technology and society have become more common – see for example the newly established *Weizenbaum Institute for the Networked Society* in Berlin[44] – indicating that new forms of research and collaboration are required to fully engage with the challenges of

[44] Emphasizing its interdisciplinary approach, the *Weizenbaum Institute* promises to "unite all relevant disciplines in a single research program and develop a holistic perspective on the process of digitalization of society", see https://vernetzung-und-gesellschaft.de/english/ [Accessed 20 June 2018]. For an effort at "Mode 2" knowledge production in the Dutch humanities, see Wyatt (2016) and Wyatt and Millen (2014).

digital transformation. A final remark: In-depth studies that add to our understanding of the digital domain in general and platform politics in particular may crucially depend on the level of insight into the workings of usually proprietary, blacked-boxed algorithmic systems and their wider societal embeddedness (see Kitchin, 2017). Consequently, regulatory efforts to increase the transparency of such systems not only to ensure fair market competition (see Khan and Toplensky, 2018), but also to facilitate academic inquiry into how these systems shape and restructure everyday life, should be welcomed as a step in the right direction.

Providing Guidance. While the critique of Big Data practices is well developed within the SSH literature, attempts to provide guidance and chart possible ways forward are far less common. Emphasizing the need for responsible Big Data research, Zook and co-authors (2017), for instance, have presented "ten simple rules" for addressing ethical issues and minimizing potential harms. Though the authors' recommendations are fairly general – reaching from the acknowledgment that even seemingly innocuous data can possibly do harm to calls for codes of conduct and greater auditability – they specify a concise set of principles that may counter negligence and naïvety when it comes to ethics. And while manifestos such as these certainly do not yield a complete solution to the complex ethical challenges of modern data analytics, they nevertheless may contribute to sensitizing scholars and practitioners to the wider implications of their craft. In some cases, this might lead to a rethinking – or even an abandonment – of certain tools and products. As Muller (2018: 3) notes:

> "There are things that can be measured. There are things that are worth measuring. But what can be measured may have no relationship to what we really want to know. The costs of measuring may be greater than the benefits. The things that get measured may draw effort away from the things we really care about. And measurement may provide us with distorted knowledge – knowledge that seems solid but is actually deceptive."

To my mind, this quote epitomizes the kind of guidance SSH research is probably best suited to provide: not only offering feedback as to how to do things, but also, and perhaps even more importantly, how and when not to do them. The ability to unpack technologies and contextualize them within their wider sociotechnical assemblage is, I would argue, one of the main ways how the social sciences and the humanities can support a proper and responsible

conduct of Big Data research and innovation.[45] As indicated, however, there also needs to be a sincere willingness of public and private stakeholders to consider and act upon such input, even if this could mean missing out on potential short-term gains.

From a cursory glance, it may appear that this thesis paints an overly pessimistic picture of Big Data. But that is neither the intention nor the case. The problem is not Big Data as such, it is how computerized measurements are perceived, used, and governed. Surrounded by an "aura of truth, objectivity, and accuracy" (boyd and Crawford, 2012: 663) and embedded in a culture driven by "metric fixation" (Muller, 2018: 17), Big Data technologies are bound to do more harm than good. Behind the façade of mechanical precision, they may incorporate biases and perpetuate discrimination. Worse still, by black boxing the analytical process, algorithmic software may curtail people's ability to contest their verdict. The result could be a world ruled by automated law, regulated by "weapons of math destruction" (O'Neil: 2016) that offer no appeal. But this is not the only way forward. Data science is a craft and practice. It demands human judgement. It serves certain interests and objectives. It offers views from 'somewhere' rather than 'nowhere'. Data science has great potential: it can be used to unveil patterns of prejudice and discrimination, it can be employed to find and support those who need it most, it can contribute to improving fairness in a world that is not very fair to begin with. As a practice, however, data science can be as fallible and flawed as the people behind the screen. This thesis is not about the evils of Big Data. Instead, it is about the need for an algorithmic culture that sees the potential, acknowledges the pitfalls, and uses the power of data in a responsible manner. The following articles are an attempt to contribute to this endeavor.

[45] To elucidate what is meant by "unpack" and "contextualize", consider the following quote by Kitchin (2017: 25): "[A]lgorithms are not formulated or do not work in isolation, but form part of a technological stack that includes infrastructure/hardware, code platforms, data and interfaces, and are framed and conditioned by forms of knowledge, legalities, governmentalities, institutions, marketplaces, finance and so on. A wider understanding of algorithms then requires their full socio-technical assemblage to be examined, including an analysis of the reasons for subjecting the system to the logic of computation in the first place."

5. References

Ambrose, M. L. (2015). Lessons from the Avalanche of Numbers: Big Data in Historical Perspective. *Journal of Law and Policy for the Information Society*, 11(2), pp. 201-277.

Andriukaitis, V. (2018). *Untitled Speech at the European Alliance for Personalised Medicine 6th Annual Presidency Conference, Brussels*. [online] Available at: http://europa.eu/!vu34Nk [Accessed 20 June 2018].

Angwin, J., Larson, J., Mattu, S., Kirchner, L. (2016). Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks. *ProPublica*. [online] Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed 20 June 2018].

Armstrong, K. (2014). Book Reviews. Big Data: A Revolution that Will Transform How We Live, Work, and Think. *Information, Communication & Society*, 17(10), pp. 1300-1302.

Aronova, E., von Oertzen, C. and Sepkoski, D. (2017). Introduction: Historicizing Big Data. In: E. Aronova, C. von Oertzen and D. Sepkoski, eds., Data Histories, *Osiris*, 32(1), pp. 1-17.

Atwood, M. (2005). *Writing with Intent. Essays, Reviews, Personal Prose: 1983-2005*. New York: Caroll & Graf Publishers.

Bamford, J. (2008). *The Shadow Factors. The Ultra-Secret NSA from 9/11 to the Eavesdropping on America*. New York: Doubleday.

Barbero, M., Coutuer, J., Jackers, R., Moueddene, K., Renders, E., Stevens, W., Toninato, Y., van der Peijl, S. and Versteele, D. (2016). *Big Data Analytics for Policy Making*. Report. A study prepared for the European Commission DG INFORMATICS (DG DIGIT). [pdf] Available at: https://joinup.ec.europa.eu/sites/default/files/document/2016-07/dg_digit_study_big_data_analytics_for_policy_making.pdf [Accessed 20 June 2016].

Barocas, S. and Selbst, A. D. (2016 [2014]). Big Data's Disparate Impact. *California Law Review*, 104, pp. 671-732.

Bazeley, P. and Jackson, K. (2013). *Qualitative Data Analysis with NVivo*. Second Edition. Los Angeles; London; New Delhi: SAGE.

Beer, D. (2016). How Should We Do the History of Big Data? *Big Data & Society*, 3(1), pp. 1-10.

Bechtold, U., Fuchs, D. and Gudowsky, N. (2017). Imagining Socio-Technical Futures – Challenges and Opportunities for Technology Assessment. *Journal of Responsible Innovation*, 4(2), pp. 85-99.

Bergen, M. (2018). Nobody Wants to Let Google Win the War for Maps All Over Again. *Bloomberg*. [online] Available at: https://www.bloomberg.com/news/features/2018-02-21/nobody-wants-to-let-google-win-the-war-for-maps-all-over-again [Accessed 20 June 2018].

Blumer, H. (1954). What Is Wrong with Social Theory? *American Sociological Review*, 19(1), pp. 3-10.

Bos, C., Walhout, B., Peine, A. and van Lente, H. (2014). Steering with Big Words: Articulating Ideographs in Research Programs. *Journal of Responsible Innovation*, 1(2), pp. 151-170.

boyd, d. and Crawford, K. (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, 15(5), pp. 662-679.

Charmaz, K. (2006). *Constructing Grounded Theory. A Practical Guide Through Qualitative Analysis*. London; Thousand Oaks; New Delhi: SAGE Publications.

Clarke, A. E. (2005). *Situational Analysis. Grounded Theory After the Postmodern Turn*. Thousand Oaks; London; New Delhi: SAGE.

Clarke, A. E., Friese, C. and Washburn, R. (2015). Chapter 1: Introducing Situational Analysis. In: A. E. Clarke, C. Friese and R. Washburn, eds., *Situational Analysis in Practice: Mapping Research with Grounded Theory*. Walnut Creek, CA: Left Coast Press, pp. 11-75.

Condliffe, J. (2016). China Turns Big Data into Big Brother. *MIT Technology Review*. [online] Available at: https://www.technologyreview.com/s/602987/china-turns-big-data-into-big-brother/ [Accessed 20 June 2018].

Conger, K. (2018). Google Removes 'Don't Be Evil' Clause from Its Code of Conduct. *Gizmodo*. [online] Available at: https://gizmodo.com/google-removes-nearly-all-mentions-of-dont-be-evil-from-1826153393 [Accessed 20 June 2018].

Conger, K. and Cameron, D. (2018). Google Is Helping the Pentagon Build AI for Drones. *Gizmodo*. [online] Available at: https://gizmodo.com/google-is-helping-the-pentagon-build-ai-for-drones-1823464533 [Accessed 20 June 2018].

Daston, L. and Galison, P. (2010 [2007]). *Objectivity*. Brooklyn, NY: Zone Books.

Davidow, B. (2014). Welcome to Algorithmic Prison. *The Atlantic*. [online] Available at: https://www.theatlantic.com/technology/archive/2014/02/welcome-to-algorithmic-prison/283985/ [Accessed 20 June 2018].

Desrosières, A. (1998). *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge; London: Harvard University Press.

Diakopolous, N. (2015). Algorithmic Accountability: Journalistic Investigation of Computational Power Structures. *Digital Journalism*, 3(3), pp. 398-415.

Doctorow, C. (2014). How Amazon Is Holding Hachette Hostage. *The Guardian*. [online] Available at: https://www.theguardian.com/technology/2014/jun/20/how-amazon-holding-hachette-hostage [Accessed 20 June 2018].

Driscoll, K. (2012). From Punched Cards to "Big Data": A Social History of Database Populism. *communication +1*, 1(1), pp. 1-33.

Dutcher, J. (2014). What Is Big Data? *Berkeley School of Information Blog*. [online] Available at: https://datascience.berkeley.edu/what-is-big-data/ [Accessed 20 June 2018].

EC-DTM (European Commission - Digital Transformation Monitor) (2017). *Big Data: A Complex and Evolving Regulatory Framework*. [pdf] Available at: https://ec.europa.eu/growth/tools-

databases/dem/monitor/sites/default/files/DTM_Big%20Data%20v1_0.pdf [Accessed 20 June 2018].

Economist.com (2013). *A New Goldmine: Making Official Data Public Could Spur Lots of Innovation*. [online] Available at: https://www.economist.com/news/business/21578084-making-official-data-public-could-spur-lots-innovation-new-goldmine [Accessed 20 June 2018].

ESSnet Big Data (2018). *Welcome to ESSnet Big Data*. [website] Available at: https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/Main_Page [Accessed 20 June 2018].

European Commission (EC) (2017). *Antitrust: Commission Fines Google €2.42 billion for Abusing Dominance as Search Engine by Giving Illegal Advantage to Own Comparison Shopping Service*. Press Release. [online] Available at: http://europa.eu/rapid/press-release_IP-17-1784_en.htm [Accessed 20 June 2018].

European Commission (EC) (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)*. Official Journal of the European Union, L 119/1. [online] Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN [Accessed 20 June 2018].

European Commission (EC) (2015). *Antitrust: Commission Sends Statement of Objections to Google on Comparison Shopping Service; Opens Separate Formal Investigation on Android*. Press Release. [online] Available at: http://europa.eu/rapid/press-release_IP-15-4780_en.htm [Accessed 20 June 2018].

European Commission (EC) (2012). *Safeguarding Privacy in a Connected World: A European Data Protection Framework for the 21st Century*. Communication. [pdf] Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012DC0009&from=en [Accessed 20 June 2018].

European Union (EU) (2012). *Charter of Fundamental Rights of the European Union*. [pdf] Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012P/TXT&from=EN [Accessed 20 June 2018].

Ezrachi, A. and Stucke, M. E. (2016). *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*. Cambridge, MA: Harvard University Press.

Ferguson, A. G. (2017). Policing Predictive Policing. *Washington University Law Review*, 94(5), pp. 1109-1189.

Fisher, E. (2017). Entangled Futures and Responsibilities in Technology Assessment. *Journal of Responsible Innovation*, 4(2), pp. 83-84.

Fowler, G. A. and Esteban, C. (2018). 14 Years of Mark Zuckerberg Saying Sorry, Not Sorry. *The Washington Post*. [online] Available at: https://www.washingtonpost.com/graphics/2018/business/facebook-zuckerberg-apologies/?utm_term=.cb53864debb3 [Accessed 20 June 2018].

Funtowicz, S. and Pereira, Â. G. (2015). Cartesian Dreams. In: Â. G. Pereira and S. Funtowicz, eds., *Science, Philosophy and Sustainability: The End of the Cartesian Dream*. London; New York: Routledge, pp. 1-9.

Garrahan, M. (2017). Google and Facebook Dominance Forecast to Rise. *Financial Times*. [online] Available at: https://www.ft.com/content/cf362186-d840-11e7-a039-c64b1c09b482 [Accessed 20 June 2018].

Gibbs, S. (2015). What Is 'Safe Harbour' and Why Did the EUCJ Just Declare It Invalid? *The Guardian*. [online] Available at: https://www.theguardian.com/technology/2015/oct/06/safe-harbour-european-court-declare-invalid-data-protection [Accessed 20 June 2018].

Glaser, S. and Strauss, A. L. (1967 [2006]). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New Brunswick; London: AldineTransaction.

Goodin, D. (2015). User Data Plundering by Android and iOS Apps Is as Rampant as You Suspected. *ArsTechnica*. [online] Available at: https://arstechnica.com/information-technology/2015/11/user-data-plundering-by-android-and-ios-apps-is-as-rampant-as-you-suspected/ [Accessed 20 June 2018].

Goodman, B. and Flaxman, S. (2017). European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". *AI Magazine*, 38(3), pp. 50-57.

Greenberg, A. (2017). How One of Apple's Key Privacy Safeguards Falls Short. *Wired*. [online] Available at: https://www.wired.com/story/apple-differential-privacy-shortcomings/ [Accessed 20 June 2018].

Greenberg, A. (2016). Apple's 'Differential Privacy' Is about Collecting Your Data – But Not *Your* Data. *Wired*. [online] Available at: https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/ [Accessed 20 June 2018].

Greenwald, G. (2014). *No Place to Hide: Edward Snowden, the NSA, and the U.S. Surveillance State*. London: Hamish Hilton.

Guston, D. H. (2014). Understanding 'Anticipatory Governance'. *Social Studies of Science*, 44(2), pp. 218-242.

Hacking, I. (1990). *The Taming of Chance*. Cambridge, UK: Cambridge University Press.

Habermas, J. (1964). The Public Sphere: An Encyclopedia Article. *New German Critique*, 3 (Autumn, 1974), pp. 49-55.

Hagemann, R., Skees, J. and Thierer, A. D. (2018, forthcoming). Soft Law for Hard Problems: The Governance of Emerging Technologies in an Uncertain Future. *Colorado Technology Law Journal*. [online] Available at: https://ssrn.com/abstract=3118539 [Accessed 20 June 2018].

Halzack, S. (2015). Privacy Advocates Try to Keep "Creepy," "Eavesdropping" Hello Barbie from Hitting Shelves. *The Washington Post*. [online] Available at: http://wapo.st/18AbUsw?tid=ss_tw&utm_term=.b51c09c91dcb [Accessed 20 June 2018].

Harford, T. (2014). Big Data: Are We Making a Big Mistake? *Financial Times*. [online] Available at: https://www.ft.com/content/21a6e7d8-b479-11e3-a09a-00144feabdc0 [Accessed 20 June 2018].

Harris, S. (2015). Your Samsung SmartTV Is Spying on You, Basically. *The Daily Beast*. [online] Available at: https://www.thedailybeast.com/your-samsung-smarttv-is-spying-on-you-basically [Accessed 20 June 2018].

Hartzog, W. and Selinger, E. (2013). Big Data in Small Hands. *Stanford Law Review Online*, 66, pp. 81-88.

Hautala, L. (2016). The Snowden Effect: Privacy Is Good for Business. *CNET*. [online] Available at: https://www.cnet.com/news/the-snowden-effect-privacy-is-good-for-business-nsa-data-collection/ [Accessed 20 June 2018].

Hilgartner, S., Prainsack, B. and Hurlbut, J. B. (2017). Ethics as Governance in Genomics and Beyond. In: U. Felt, R. Fouché, C. A. Miller and L. Smith-Doerr, eds., *The Handbook of Science and Technology Studies*. Fourth Edition. Cambridge, MA; London: The MIT Press, pp. 823-851.

Holmes, D. E. (2017). *Big Data: A Very Short Introduction*. Oxford, UK: Oxford University Press.

Iliadis, A. and Russo, F. (2016). Critical Data Studies: An Introduction. *Big Data & Society*, 3(2), pp. 1-7.

Jasanoff, S. (2004). The Idiom of Co-Production. In: S. Jasanoff, ed., *States of Knowledge: The Co-Production of Science and Social Order*. London, New York: Routledge, pp. 1-12.

Johnston, S. F. (2018). The Technological Fix as Social Cure-All. *IEEE Technology and Society Magazine*, 37(1), pp. 47-54.

Kallinikos, J. (2013). The Allure of Big Data. *Mercury Magazine*, 2(3), pp. 41-43.

Kalra, A. and Shah, A. (2018). Competition Commission of India fines Google for Abusing Dominant Position. *Reuters*. [online] Available at: https://in.reuters.com/article/india-google-antitrust/competition-commission-of-india-fines-google-for-abusing-dominant-position-idINKBN1FS29Z [Accessed 20 June 2018].

Kastrenakes, J. (2016). Apple Is Returning Money to Ebook Buyers After Losing Price-Fixing Lawsuit. *The Verge*. [online] Available at: https://www.theverge.com/2016/6/21/11991526/apple-ebook-price-fixing-refunds-begin-for-consumers [Accessed 20 June 2018].

KDnuggets (2012). *Poll Results: What Will Replace Big Data and When?* [online] Available at: https://www.kdnuggets.com/2012/09/what-will-replace-big-data-when.html [Accessed 20 June 2018].

Khan, M. and Toplensky, R. (2018). Google Targeted Under EU Plan to Regulate Search Engines. *Financial Times*. [online] Available at: https://www.ft.com/content/c6932a32-2797-11e8-b27e-cc62a39d57a0 [Accessed 20 June 2018].

Kitchin, R. (2017). Thinking Critically About and Researching Algorithms. *Information, Communication & Society*, 20(1), pp. 14-29.

Kitchin, R. (2014a). Big Data, New Epistemologies and Paradigm Shifts. *Big Data & Society*, 1(1), pp. 1-12.

Kitchin, R. (2014b). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London; Thousand Oaks; New Delhi: SAGE.

Kitchin, R. and Lauriault, T. P. (2014). Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work. The Programmable City Working Paper 2. Pre-print version of chapter to be published in: J. Thatcher, A. Shears and J. Eckert, eds., *Thinking Big Data in Geography: New Regimes, New Research*. Lincoln; London: University of Nebraska Press (2018, forthcoming).

Kranzberg, M. (1986). Technology and History: "Kranzberg's Laws". *Technology and Culture*, 27(3), pp. 544-560.

Laney, D. (2012). Deja VVVu: Others Claiming Gartner's Construct for Big Data. *Gartner Blog Network*. [online] Available at: https://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/ [Accessed 20 June 2018].

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. *Meta Group. Application Delivery Strategies*. [pdf] Available at: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf [Accessed 20 June 2018].

Law, J. (1991). Introduction: Monsters, Machines and Sociotechnical Relations. In: J. Law, ed., *A Sociology of Monsters: Essays on Power, Technology and Domination*. Sociological Review Monograph 38. London; New York: Routledge, pp. 1-23.

Leahy, J. (2013). Google Search Battle Moves to Brazil. *Financial Times*. [online] Available at: https://www.ft.com/content/cb929a86-32af-11e3-b3a7-00144feab7de [Accessed 20 June 2018].

Leonelli, S. (2014). What Difference Does Quantity Make? On the Epistemology of Big Data in Biology. *Big Data & Society*, 1(1), pp. 1-11.

Lum, K. and Isaac, W. (2016). To Predict and Serve? *Significance*, 13(5), pp. 14-19.

Lyon, D. (2015). *Surveillance after Snowden*. Cambridge; Malden: Polity Press.

Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, Consequences, Critique. *Big Data & Society*, 1(2), pp. 1-13.

Mager, A. (2017). Search Engine Imaginary: Visions and Values in the Co-Production of Search Technology and Europe. *Social Studies of Science*, 47(2), pp. 240-262.

Manthorpe, R. (2018). Congress Won't Hurt Facebook and Zuck, But GDPR and Europe Could. *Wired*. [online] Available at: http://www.wired.co.uk/article/facebook-mark-zuckerberg-congress-testimony-cambridge-analytica-gdpr-privacy [Accessed 20 June 2018].

Marcus, G. and Davis, E. (2014). Eight (No, Nine!) Problems with Big Data. *The New York Times*. [online] Available at: https://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html [Accessed 20 June 2018].

Mayer-Schönberger, V. and Cukier, K. (2013). *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt.

Meyer, R. (2014). Everything We Know about Facebook's Secret Mood Manipulation Experiment. *The Atlantic*. [online] Available at: https://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/ [Accessed 20 June 2018].

Miller, C. A. and Bennett, I. (2008). Thinking Longer Term about Technology: Is There Value in Science Fiction-Inspired Approaches to Constructing Futures? *Science and Public Policy*, 35(8), pp. 597-606.

Miner, M. (1991). "Trust Me": Reading the Romance Plot in Margaret Atwood's The Handmaid's Tale. *Twentieth Century Literature*, 37(2), pp. 148-168.

Mohabbat Kar, R., Basanta, T. and Parycek, P., eds., *(Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft*. Berlin: Kompetenzzentrum Öffentliche IT.

Mol, A. (2002). *The Body Multiple: Ontology in Medical Practice*. Durham; London: Duke University Press.

Moore, T. R. (2017). *Trade Secrets and Algorithms as Barriers to Social Justice*. Resource of the Center for Democracy & Technology. [pdf] Available at: https://cdt.org/files/2017/08/2017-07-31-Trade-Secret-Algorithms-as-Barriers-to-Social-Justice.pdf [Accessed 20 June 2018].

Moses, L. (2016). Survey: 80 Percent of Those Who Know about Ad Blocking Use It. *Digiday UK*. [online] Available at: https://digiday.com/media/survey-80-percent-know-ad-blocking-use/ [Accessed 20 June 2018].

Muller, J. Z. (2018). *The Tyranny of Metrics*. Princeton; Oxford: Princeton University Press.

Mullins, B., Winkler, R. and Kendall, B. (2015). Inside the U.S. Antitrust Probe of Google. *The Wall Street Journal*. [online] Available at: https://www.wsj.com/articles/inside-the-u-s-antitrust-probe-of-google-1426793274 [Accessed 20 June 2018].

Naughton, J. (2012) Big Data: Revolution by Numbers. *The Guardian*. [online] Available at: https://www.theguardian.com/technology/2012/nov/18/data-analysis-applied-business-science [Accessed 20 June 2018].

Nielsen, N. (2018). Austria Accused of Undermining New EU Data Law. *EUobserver*. [online] Available at: https://euobserver.com/justice/141746 [Accessed 20 June 2018].

Nielsen, N. (2015). EU-US Data Packed Skewered in Court Hearings. *EUobserver*. [online] Available at: https://euobserver.com/justice/128131 [Accessed 20 June 2018].

Nissenbaum, H. (2016). *Differential Privacy in Context: Conceptual and Ethical Considerations*. Differential Privacy Symposium, Institute for Advanced Study, Princeton, NJ. [online video] Available at: https://youtu.be/Nefc4ey9XMU [Accessed 20 June 2018].

Office of Oversight and Investigations (2013). *A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes*. Staff Report for Chairman Rockefeller, December 18, 2013. [pdf] Available at:

http://educationnewyork.com/files/rockefeller_databroker.pdf [Accessed 20 June 2018].

O'Neil, C. (2017). The Ivory Tower Can't Keep Ignoring Tech. *The New York Times*. [Online] Available at: https://www.nytimes.com/2017/11/14/opinion/academia-tech-algorithms.html [Accessed 20 June 2018].

O'Neil, C. (2016). *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*. New York: Broadway Books.

Orwell, G. (1949). *Nineteen Eighty-Four. A Novel*. London: Secker & Warburg.

Pagallo, U. (2017). The Legal Challenges of Big Data: Putting Secondary Rules First in the Field of EU Data Protection. *European Data Protection Law Review*, 3(1), pp. 36-46.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge; London: Harvard University Press.

Pietsch, W., Wernecke, J. and Ott, M., eds., (2016). *Berechenbarkeit der Welt? Philosophie und Wissenschaft im Zeitalter von Big Data*. Wiesbaden: Springer.

Pittas, C. (2013). Snowden Warns of Loss of Privacy in Christmas Message. *Reuters*. [online] Available at: https://www.reuters.com/article/us-usa-snowden-privacy-idUSBRE9BO09020131225 [Accessed 20 June 2018].

Poel, M., Schroeder, R., Treperman, J., Rubinstein, M., Meyer, E., Mahieu, B., Scholten, C. and Svetachova, M. (2015). *Data for Policy: A Study of Big Data and Other Innovative Data-Driven Approaches for Evidence-Informed Policymaking*. Draft report about the State-of-the-Art: invitation for reflection. [pdf] Available at: https://ofti.org/wp-content/uploads/2015/05/dataforpolicy.pdf [Accessed 20 June 2018].

Porter, T. M. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, New Jersey: Princeton University Press.

Press, G. (2014). 12 Big Data Definitions: What's Yours? *Forbes*. [online] Available at: https://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/#2898e7a413ae [Accessed 20 June 2018].

Radinsky, K. (2015). Data Monopolists Like Google are Threatening the Economy. *Harvard Business Review*. [online] Available at: https://hbr.org/2015/03/data-monopolists-like-google-are-threatening-the-economy [Accessed 20 June 2018].

Ralph, O. (2017). Insurance and the Big Data Technology Revolution. *Financial Times*. [online] Available at: https://www.ft.com/content/bb9f1ce8-f84b-11e6-bd4e-68d53499ed71 [Accessed 20 June 2018].

Rango, M. and Vespe, M. (2017). *Big Data and Alternative Data Sources on Migration: From Case-Studies to Policy Support*. Workshop Summary Report. [pdf] Available at: https://bluehub.jrc.ec.europa.eu/bigdata4migration/uploads/attachments/cjdelbdgo00hnqazv3u7xi6pd-big-data-workshop-draft-summary-report.pdf [Accessed 20 June 2018].

Rieder, G. (2018). Tracing Big Data Imaginaries through Public Policy: The Case of the European Commission. In: A.R. Sætnan, I. Schneider and N. Green, eds., *The Politics*

and Policies of Big Data: Big Data, Big Brother? New York; London: Routledge, pp. 89-109.

Rieder, G. and Simon, J. (2018). Vertrauen in Daten oder: Die politische Suche nach numerischen Beweisen und die Erkenntnisversprechen von Big Data. In: R. Mohabbat Kar, T. Basanta and P. Parycek, eds., *(Un)berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft*. Berlin: Kompetenzzentrum Öffentliche IT, pp. 159-178.

Rieder, G. and Simon, J. (2017a). Big Data: A New Empiricism and its Epistemic and Socio-Political Consequences. In: W. Pietsch, J. Wernecke and M. Ott, eds., *Berechenbarkeit der Welt? Philosophie und Wissenschaft im Zeitalter von Big Data*. Wiesbaden: Springer, pp. 85-105.

Rieder, G. and Simon, J. (2017b). Big Data and Technology Assessment: Research Topic or Competitor? *Journal of Responsible Innovation*, 4(2), pp. 234-253.

Rieder, G. and Simon, J. (2016). Datatrust: Or, the Political Quest for Numerical Evidence and the Epistemologies of Big Data. *Big Data and Society*, 3(1), pp. 1-6.

Rieder, G. and Völker, T. (2018, forthcoming). *Datafictions: Or, How Measurements and Predictive Analytics Rule Imagined Future Worlds*. Manuscript currently under review.

Rieke, A., Robinson, D. and Yu, H. (2014). *Civil Rights, Big Data, and Our Algorithmic Future: A September 2014 Report on Social Justice and Technology*. [pdf] Available at: https://bigdata.fairness.io/wp-content/uploads/2015/04/2015-04-20-Civil-Rights-Big-Data-and-Our-Algorithmic-Future-v1.2.pdf [Accessed 20 June 2018].

Rieke, A., Yu, H., Robinson, D. and von Hoboken, J. (2016). *Data Brokers in an Open Society*. An Upturn Report, prepared for the Open Society Foundations. [pdf] Available at: https://www.opensocietyfoundations.org/sites/default/files/data-brokers-in-an-open-society-20161121.pdf [Accessed 20 June 2018].

Robertson, A. (2013). Apple Guilty of Ebook Price Fixing, Rules Federal Court. *The Verge*. [online] Available: https://www.theverge.com/2013/7/10/4510338/apple-found-guilty-of-ebook-price-fixing [Accessed 20 June 2018].

Rose, J. (2017). GDPR and the Death of Third-Party Data. *Gigya*. [online] Available at: https://www.gigya.com/blog/gdpr-death-third-party-data/ [Accessed 20 June 2018].

Rowley, J. D. (2018). The Largest Buys of Tech's Big Five: A Look at M&A Deals. *Crunchbase*. [online] Available at: https://news.crunchbase.com/news/the-largest-buys-of-techs-big-five-a-look-at-ma-deals/ [Accessed 20 June 2018].

Rubinstein, I. (2013). Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law*, 3(2), pp. 74-87.

Ruppert, E., Isin, E. and Bigo, D. (2017). Data Politics. *Big Data & Society*, 4(2), pp. 1-7.

Sætnan, A. R., Schneider, I. and Green, N., eds., (2016). *The Politics and Policies of Big Data: Big Data, Big Brother?* New York; London: Routledge.

Sardana, S. and Sardana, S. (2013). Big Data: It's Not a Buzzword, It's a Movement. *Forbes*. [online] Available at:

https://www.forbes.com/sites/sanjeevsardana/2013/11/20/bigdata/#59d776611129 [Accessed 20 June 2018].

Schmidt, E. (2015). The Tinkerer's Apprentice. *Project Syndicate*. [online] Available at: https://www.project-syndicate.org/commentary/google-european-commission-and-disruptive-technological-change-by-eric-schmidt-2015-01 [Accessed 20 June 2018].

Schneier, B. (2015). *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. New York; London: W. W. Norton & Company.

von Schomberg, R. (2013). A Vision of Responsible Innovation. In: R. Owen, J. Bessant and M. Heintz, eds., *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*. London: Wiley, pp. 51-74.

Scott, M. (2018). Decision in Google's Android Antitrust Case Likely Due in July. *Politico*. [online] Available at: https://www.politico.eu/article/google-android-antitrust-competition-fine-margrethe-vestager/ [Accessed 20 June 2018].

Scott, M. (2016). Google Fined for Breaking Russian Antitrust Rules with Android. *The New York Times*. [online] Available at: https://www.nytimes.com/2016/08/12/business/international/google-fine-russia-antitrust-android.html [Accessed 20 June 2018].

Shelley, M. (1818). *Frankenstein; Or, The Modern Prometheus*. London: Lackington, Hughes, Harding, Mavor & Jones.

Shklovski, I., Mainwaring, S. D., Skúladóttir, H. H. and Borgthorsson, H. (2014). Leakiness and Creepiness in App Space: Perceptions of Privacy and Mobile App Space. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, April 26 – Mai 01, Toronto, Ontario, Canada, pp. 2347-2356.

Simon, J. (2016). Data Is Not Agnostic. Interview by Sara Kjellberg. *Medium*. [online] Available at: https://medium.com/big-data-small-meaning-and-global-discourses/data-is-not-agnostic-7276ca5781fb [Accessed 20 June 2018].

Simon, J. (2013). Trust. In: D. Pritchard, ed., *Oxford Bibliographies in Philosophy*. New York: Oxford University Press.

Sokolov, D. A. (2018). Keine Strafen: Österreich Zieht Neuem Datenschutz die Zähne. *Heise Online*. [online] Available at: https://www.heise.de/newsticker/meldung/Keine-Strafen-Oesterreich-zieht-neuem-Datenschutz-die-Zaehne-4031217.html [Accessed 20 June 2018].

Solesbury, W. (2002). The Ascendancy of Evidence. *Planning Theory & Practice*, 3(1), pp. 90-96.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. and Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, 25(4), pp. 359-393.

S.NET Organizing Committee (2018). *2018 Annual S.NET Meeting - Registration Now Open & Tentative Program Outline Available*. [website] Available at: http://www.maastrichtsts.nl/snet/ [Accessed 20 June 2018].

Stein, S. (2014). Nationalelf bereitet sich mit Big Data auf WM vor. *Bild.de*. [online] Available at: https://www.bild.de/digital/cebit/cebit/bierhoff-setzt-auf-daten-nationalelf-bereitet-sich-mit-big-data-auf-wm-vor-35010948.bild.html [Accessed 20 June 2018].

Stilgoe, J., Owen, R., Macnaghten, P. (2013). Developing A Framework for Responsible Research and Innovation. *Research Policy*, 42(9), pp. 1568-1580.

Strauss, A. (1978). A Social World Perspective. *Studies in Symbolic Interactionism*, 1, pp. 119-128.

Thomas, P. L. (2013). Introduction Science Fiction and Speculative Fiction. In: P. L. Thomas, ed., *Science Fiction and Speculative Fiction. Challenging Genres*. Rotterdam: Sense Publishers, pp. 1-13.

Treaty of Paris (1951). *Treaty establishing the European Coal and Steel Community (Paris, April 1951)*. [online] Available at: https://www.cvce.eu/obj/treaty_establishing_the_european_coal_and_steel_community_paris_18_april_1951-en-11a21305-941e-49d7-a171-ed5be548cd58.html [Accessed 20 June 2018].

Turow, J. (2011). *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. New Haven; London: Yale University Press.

Vincent, J. (2016). Twitter Taught Microsoft's AI Chatbot to Be a Racist Asshole in Less Than a Day. *The Verge*. [online] Available at: https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist [Accessed 20 June 2018].

Wachter, S., Mittelstadt, M. and Floridi, L. (2017). Transparent, Explainable, and Accountable AI for Robotics. *Science Robotics*, 2(6).

Welch, C. (2018). Google Just Gave a Stunning Demo of Assistant Making an Actual Phone Call. *The Verge*. [online] Available at: https://www.theverge.com/2018/5/8/17332070/google-assistant-makes-phone-call-demo-duplex-io-2018 [Accessed 20 June 2018].

White House (2016). *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*. [pdf] Available at: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf [Accessed 20 June 2018].

Wyatt, S. (2015). Mode 2 in Action: Working Across Sectors to Create a Center for Humanities and Technology. *Scholarly Research and Communication*, 6(4), pp. 1-8.

Wyatt, S. and Millen, D., eds., (2014). *Meaning and Perspective in the Digital Humanities. A White Paper for the Establishment of a Center for Humanities and Technology (CHAT)*. Amsterdam, NL: Royal Netherlands Academy of Arts & Sciences (KNAW).

Zook, M., Barocas, S., boyd, d., Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B. A., Metcalf, J., Narayanan, A., Nelson, A., Pasquale, F. (2017). Ten Simple Rules for Responsible Bit Data Research. *PLoS Computational Biology*, 13(3), pp. 1-10.

Zuboff, S. (2015). Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*, 30, pp. 75-89.

PART II
Research Papers

Paper I

Datatrust: Or, the Political Quest for Numerical Evidence and the Epistemologies of Big Data

by Gernot Rieder and Judith Simon

Presented Version:

Accepted, peer-reviewed manuscript corrected for typographical errors and citation style

Published Version:

Rieder, G. and Simon, J. (2016). Datatrust: Or, the Political Quest for Numerical Evidence and the Epistemologies of Big Data. *Big Data and Society*, 3(1), pp. 1-6.

Datatrust: Or, the Political Quest for Numerical Evidence and the Epistemologies of Big Data

by Gernot Rieder and Judith Simon

Abstract

Recently, there has been renewed interest in so-called evidence-based policy making. Enticed by the grand promises of Big Data, public officials seem increasingly inclined to experiment with more data-driven forms of governance. But while the rise of Big Data and related consequences has been a major issue of concern across different disciplines, attempts to develop a better understanding of the phenomenon's historical foundations have been rare. This short commentary addresses this gap by situating the current push for numerical evidence within a broader socio-political context, demonstrating how the epistemological claims of Big Data science intersect with specific forms of trust, truth, and objectivity. We conclude by arguing that regulators' faith in numbers can be attributed to a distinct political culture, a representative democracy undermined by pervasive public distrust and uncertainty.

Keywords

Big Data, evidence-based policy making, quantification, trust in numbers, mechanical objectivity, epistemology

Over the past few years, there has been growing interest in so-called "evidence-based policy making". While the concept is not new (see Solesbury, 2002), the latest push for more data-driven modes of governance has been considerable (see Haskins, 2014). Against the backdrop of multiple crises, policymakers seem ever more inclined to legitimize specific ways of action by referring to 'hard' scientific evidence suggesting that a particular initiative will eventually yield the desired outcomes (see Urahn, 2015). Across many areas of public service – be it healthcare, education, or law enforcement – a steady influx of "data for policy" (EC 2015a) is meant to offer guidance in a moment marked by high levels of complexity and uncertainty (see Nowotny et al., 2001).

Legislators' current emphasis on evidence and results correlates with a recent techno-scientific development – the advent of Big Data (see Mayer-Schönberger and Cukier, 2013). While state bureaucracies have relied on statistics and numerical information for centuries (see Cohen, 2005), new analytical techniques promise to improve upon former methods in several ways: whereas data analysis has traditionally been costly and time-consuming, it is now fast and cheap; whereas previously one had to settle for samples, the ongoing computerization of society makes it possible to glean data from entire populations; whereas once there was need for theory, through sheer volume the data now speak for themselves; whereas in the past measurements were tainted by human bias, agnostic algorithms now guarantee an impartial view from nowhere. Together, the alleged qualities of Big Data technologies feed into what Rob Kitchin (2014a: 5) has described as the "articulation of a new empiricism", which operates as a "discursive rhetorical device" designed to promote the utility and value of new analytical services.

Policymakers on either side of the Atlantic have bought into the hype, usually without much regard for nuance or subtlety. In official documents and speeches, Big Data is referred to as the "new oil for the digital age" (Kroes, 2012), the next "industrial revolution" (Kroes, 2014), "gold" (ibid.), a game-changing "key asset" (EC, 2015b) for creating value, increasing productivity, and boosting growth. The technology is not only expected to improve public administration by "advanc[ing] government efficiency" (White House, 2014: 67) and enabling "better services" (EC, 2013: 2), but also to support "evidence-informed decision making" (EC, 2015a) by providing real-time feedback, generating solutions, and predicting outcomes, always ensuring that "regulation is empirically justified in advance" (Sunstein, 2012: 2). Although this focus on technology-driven benefits has in some cases expanded to include consideration of potential risks and pitfalls, political leaders remain firmly committed to "harness[ing] the power of Big Data" (Kalil and Zhao, 2013).

Much effort has already gone into challenging the buzz-laden assumptions of modern-day "data-ism" (Brooks, 2013). Investigating both the politics and power of contemporary data practices, scholars from different disciplinary backgrounds have identified a range of social, ethical, and legal issues – from privacy and security (see Ohm, 2009) to transparency and accountability (see Pasquale, 2015), bias and discrimination (see Barocas and Selbst, 2016 [2014]) – emphasizing that Big Data's presumed benefits may come at a cost. But while there has been a steady stream of critical reactions across academia and the media, attempts to

gain a better understanding of the socio-historical foundations of policymakers' push for numerical evidence have been rare. Put differently, even though the rise of Big Data and related consequences has been a major issue of concern, its significance for and embeddedness in a long-standing culture of measurement and quantification has not. As Barnes (2013) poignantly states: "Big Data, little history."

One reason for this lack of historical contextualization can be attributed to the dynamics of Big Data discourse: Presented as a rupture and revolution with no ties to the past, discussions about Big Data have focused on the modalities of change rather than forms of continuity. The 'now' is said to be fundamentally different from what came before, the 'new' supersedes the 'old'. This narrative of novelty and disruption, exemplified in notions such as Anderson's (2008) "Petabyte Age", is both powerful and convenient, but discourages appreciation of Big Data as a specific amalgamation, a "conjuncture of different elements, each with their own history, coming together at this present our moment" (Barnes 2013: 298). Yet it is precisely the recognition of Big Data's diverse roots, its connection to prior epistemic practices, that may provide greater insight into the current excitement's underlying norms and values.

Such exploratory analysis requires some conceptual rethinking: Instead of narrowly defining Big Data in mere technical terms – e.g., Laney's (2001) popular 'three Vs', which reductively characterize Big Data as an increase in (data) volume, velocity, and variety – it seems more productive to think of it as the terminologically contingent manifestation of a complex socio-technical phenomenon that rests on an interplay of technological, scientific, and cultural factors (cf. boyd and Crawford, 2012). While the *technological dimension* alludes to advances in hardware, software, but also infrastructure and the *scientific dimension* comprises both mining techniques and analytical skills, the *cultural dimension* refers to (a) the pervasive use of ICTs in contemporary society and (b) the growing significance and authority of quantified information in many areas of everyday life, including public administration and decision making. Ultimately, this broader interpretative approach may assist in "deconstructing the black boxes of Big Data" (Pasquale, 2015: 6) by paying attention not only to the mechanical, but also the mental workings of an otherwise opaque phenomenon.

Investigations into the roots and antecedents of Big Data may take different paths: Barnes and Wilson (2014), for instance, examine the origins of the social physics movement, whose monistic urge – that is, the assumption that the laws of physics apply to both natural

and social worlds – was later incorporated into spatial analysis, shaping the use of Big Data in present-day geography. Morozov (2014), drawing on Medina's (2011) *Cybernetic Revolutionaries*, details the Allende administration's Project Cybersyn to highlight the intellectual affinities between socialism, cybernetics, and Big Data, while Grandin (2014), citing Dinges (2005), reports on the Pinochet regime's Condor data bank to locate the "anti-socialist origins of Big Data", a juxtaposition of historical events illustrating that the idea of data-facilitated control may in fact appeal to different ends of the political spectrum. Last but not least, Mackenzie (2013) provides an empirical account of how recent shifts in programming practice relate to what Adams and co-authors (2009) have labeled "regimes of anticipation", demonstrating how the current emphasis on machine learning and predictive modelling is entangled with a concerted cultural effort to reduce uncertainty by fostering the continuous assessment of the 'not yet'.

While these examples offer unique perspectives, each focusing on particular cases and ideas, they are also similar in that they seek to situate Big Data discourse within a larger historical context, attributing meaning to what all too often takes the form of pure marketing. We suggest that such attempts to historicize and contextualize are crucial as they may (a) provide better insight into the epistemological foundations of contemporary data science, (b) deepen our understanding of the norms, values, and expectations driving the current climate of hope and hype, and (c) indicate potential social and ethical ramifications, serving as a guiding compass at a time when technical innovation continues to outpace government regulation (see Rubinstein, 2013). We would like to contribute to this research agenda by suggesting what may prove another fruitful avenue of investigation: The data hype's reliance on specific forms of trust, truth, and objectivity.

As boyd and Crawford (2012: 663) have argued, Big Data is not just about technological progress, but about a "widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible". Leonelli (2014: 1) makes a similar argument, stressing that the novelty of Big Data science does not lie in the sheer quantity of data involved, but in the "prominence and status acquired by data as commodity and recognized output." But where does this prominence and status come from and what exactly are the roots of the belief that more data equals better insight?

An initial answer would be that data are often perceived as raw, objective, and neutral – the "stuff of truth itself" (Gitelman, 2013: 2). But, as historians of science and technology have repeatedly shown, conceptions of objectivity, truth and truthfulness, trust and trustworthiness may vary, they are "situated and historically specific" (ibid.: 4). Therefore, it is important to clarify which particular version of these concepts manifests within Big Data discourse. One way to identify such differences is through comparison, which may involve tracing conceptual shifts and changes over time.

In his book *The Social History of Truth*, Shapin (1994) emphasizes the central role of trust in building and maintaining social order. Societies are made through acts of trust – without trust, they may falter and collapse. The allocation of trust and trustworthiness can thus be understood as the "great civility" (ibid.: 3), granting the conditions in which people can colonize each others' mind. Although often rendered invisible, trust as the "cement of society" (ibid.: 10; also see 35f.) is also essential to the construction and establishment of epistemic systems. The production of scientific knowledge, for instance, rests on myriads of social and material interactions, which take for granted the reliability of numerous stabilized norms and relationships. As a result, scientific distrust and skepticism only takes place "on the margins of trusting systems." (ibid.: 19)

But such systems of trust are not fixed – conceptions of whom to trust, what to trust, and in what circumstances, are subject to change: While in premodern society it was the politically and economically independent gentleman who was generally conceived as a credible truth-teller, modern society accorded trust to the "abstract capacities" (Giddens 1990: 26) of "faceless institutions" (Shapin, 1994: xxxi). The veracity of testimony was no longer underwritten by personal virtue, but by an elaborate system of institutionalized norms and standards, rigorously policed in a great "panopticon of truth" (ibid.: 413). A different form of trust first accompanied and then superseded the premoderns' faith in the integrity of the solitary knower and the moderns' confidence in the rigor of institutionalized expertise, a type of trust that has gained considerable traction with the arrival of Big Data: People's trust in numbers.

While the general history of quantification can be traced back much further, Desrosières (1998: 23) identifies 17th-century English political arithmetic as the "basic act of all statistical work (in the modern sense of the term), implying definite, identified, and stable

unities." Whereas early records of baptisms, marriages, and burials were meant to attest to the existence of individuals and their family relations, later statistical surveys such as the one underlying the 18th-century French "adunation" were intended to support the unification of national territory in order to establish a "politico-cognitive construction of a space of common measurement" (ibid.: 33). Examples such as these highlight the close relationship between statistics and state-making: Numbers allowed for coherence and generality, enabling central governments to exercise administrative control over matters of taxation and economic development at a time when the familiarity of face-to-face interactions gradually gave way to the anonymity and complexity of expanding trade and business networks.

But behind those numbers still stood individual experts and prestigious institutions – numbers did not speak for themselves. Quite to the contrary, it was the cultivated judgment of an administrative elite that guaranteed the trustworthiness of numerical information; deployed by outsiders, statistics counted for little. As Porter (1995: 138) explains, numbers could only "provide a modest supplement to institutional power." Their credibility rested on the authority and integrity of a bureaucracy whose members believed that measurements only became useful when subject to expert interpretation. For them, nothing could be reduced to inflexible laws, abstract formulas, or technical routines. Agreements were reached through informal discussion rather than formal procedures. In general, decisions were rarely entrusted to the numbers.

The demand for quantitative rigor increased during the first half of the 20th century: Instead of expert judgment, the pursuit of technical discipline required an "ideal of self-sacrifice" (ibid.: 89); instead of professional autonomy, the desire for precision-imposed adherence to a strict "regime of calculation" (102); instead of elite discretion, it became necessary to "manage by the numbers" (92). The result was what Porter refers to as the "cult of impersonality" (90), a specific culture of quantification that seeks to reduce the human element as much as possible, preferring formalized principles to subjective interpretation, uniform standards to methodological tinkering, the rule of law to the rule of men. The goal was to attain "mechanical objectivity" (Daston and Galison, 1992), a disinterested science that "eradicates all that is personal, idiosyncratic, perspectival." (ibid.: 123) In this brave new world, trust no longer resides in the integrity of individual truth-tellers or the veracity of prestigious institutions, but is placed in highly formalized procedures enacted through disciplined self-restraint. Numbers cease to be supplements. They are couched in a rhetoric

of factuality, imbued with an ethos of neutrality, and presented with an aura of certainty. They step out of the shadows of their human creators, enter center stage, and, in the arguments and claims of countless profiteers, start to speak for themselves.

What are the reasons for this shift toward mechanical objectivity? On the one hand, technological progress played a significant role. The growing availability of ever more capable machinery changed the face of the accounting profession. The idea was powerful: The more mechanized a process, the more automated a procedure, the less the need for – and danger of – subjective human intervention (see Venturini et al., 2014). In the words of Daston and Galison (2010: 123), "instead of freedom of will, machines offered freedom from will". The virtuous machine was conceived as the "ultimate outsider", and it was not long until "it became the greatest in the kingdom of quantification" (Porter, 1995: 85). Consequently, the "honest instrument" (Daston and Galison, 1992: 120) with its "glow of veracity" (ibid.: 111) both served as a means to and symbol of mechanical objectivity.

On the other hand, there was a social dimension: The pursuit of quantitative rigor was seen as a strategy to adapt to new external pressures in a rapidly changing political environment. War and economic crisis had left their marks, and the dynamics of democracy increased the need for hard evidence and professional accountability. Confronted with public distrust, invasive auditing, and competing political demands, bureaucratic agencies and scientific communities sought to withstand scrutiny and minimize responsibility by adhering to rigid protocols and explicit decision criteria. This willingness for personal restraint is a sign of professional weakness rather than strength: The more permeable the boundaries of a discipline, the higher its vulnerability to outside criticism, the more tempting the language of mechanical objectivity becomes. Consequently, the appeal of standardized methods is especially great in cultures where the faith in other forms of trust has been shattered. As Porter (1995) notes, methodological strictness and objective rules may serve as an alternative to trust and shared beliefs. Where trust is missing and suspicion prevails, numbers are meant to fill the gap: Regarded as carefully measured matters of fact, they are expected to offer a sense of fairness and justice, a way of making decisions without having to decide, a chance to de-politicize legislation. This push for impersonal numerical evidence is however not so much rooted in the inner workings of quantitative professions, but in the needs and demands of a specific socio-political culture, a democratic system undermined by pervasive distrust and uncertainty. It is on these grounds that the Big Data phenomenon continues to blossom.

The epistemic promises of Big Data connect to the ideal of mechanical objectivity in several ways, not only fortifying but also expanding the appeal of the doctrine:

First, a child of new analytical techniques and the progressing computerization of society, Big Data pledges to extend the reach of automation, from data collection to matters of storage, curation, and analysis. The virtuous machine emerges as ever more powerful as it covers increasingly large parts of the analytical and decision-making process.

Second, by capturing massive amounts of data and focusing on correlations rather than causes, Big Data claims to reduce the need for theory, models, and, in extension, human expertise. In addition, modern data analysis software is often thoroughly opaque, with a phenomenology that emphasizes both uniformity and impersonality.

Third, Big Data promises to expand the realm of what can be measured. Trackers, social media, and the Internet of Things allow to trace and gauge movements, actions, and behaviors in ways that were previously unfeasible. Fully quantified and free from bias, Big Data pushes the tenets of mechanical objectivity into ever more areas of application.

Fourth and finally, settling for neither the present nor the past, Big Data aspires to calculate what is yet to come. Smart, fast, and cheap predictive techniques are meant to support decision making and optimize resource allocation across many government sectors, applying a mechanical mindset to the colonization of the future.

The limitations of these "sociotechnical imaginaries" (Jasanoff and Kim, 2009) have been discussed elsewhere (e.g., see Kitchin, 2014b), but the point here is to develop a better understanding of how the current language of Big Data-related hope and hype intersects with and relies on particular forms of trust and objectivity, which, in turn, can be conceived as products of a specific socio-political culture. In a climate of distrust, crisis, and uncertainty, officials' adherence to supposedly impartial numbers may be regarded as a strategy of defense, an attempt to shield themselves from increased public and judicial scrutiny. It is not by coincidence that the European Commission, whose authority continues to be challenged by citizens and national governments alike, has emerged as the one of the most zealous political quantifiers.

Big Data has been repeatedly criticized for its positivist epistemology and its support of techno-capitalism, and while such criticism has its merits, it pays little attention to the circumstances and dynamics that contribute to the creation and internalization of

corresponding norms and values. Our proposition is simple: Instead of focusing exclusively on the potential consequences of the Big Data phenomenon, we can gain additional insight from examining its social and political, but also its technical and epistemic roots. Such an approach may foster more, not less, critical engagement as it shifts the perspective and situates Big Data discourse within a broader historical narrative. As Barnes and Wilson (2014: 11) argue:

"By showing that Big Data is historical, we show the assumptions that were built into it, as well as the contestations around them. Big Data becomes no longer a black box, self-contained, sealed and impregnable, but is opened up, available for verbalist discussion and contestation."

We wholeheartedly agree.

Acknowledgments

The authors wish to acknowledge the financial support of the Austrian Science Fund (P-23770).

References

Adams, V., Murphy, M. and Clarke, A. E. (2009). Anticipation: Technoscience, Life, Affect, Temporality. *Subjectivity*, 28(1), pp. 246-265.

Anderson, C. (2008). The End of Theory: The Big Data Deluge Makes the Scientific Method Obsolete. *Wired*. [online]. Available at: http://www.wired.com/2008/06/pb-theory/ [Accessed 20 June 2018].

Barnes, T. J. (2013). Big Data, Little History. *Dialogues in Human Geography*, 3(3), pp. 297-312.

Barnes, T. J. and Wilson, M. W. (2014). Big Data, Social Physics, and Spatial Analysis: The Early Years. *Big Data & Society*, 1(1), pp. 1-14.

Barocas, S. and Selbst, A. D. (2016 [2014]). Big Data's Disparate Impact. *California Law Review*, 104, pp. 671-732.

boyd, d. and Crawford, K. (2012). Critical Question for Big Data. Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, 15(5), pp. 662-679.

Brooks, D. (2013). The Philosophy of Data. *The New York Times*. [online] Available at: http://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html [Accessed 20 June 2018].

Cohen, B. I. (2005). *Triumph of Numbers: How Counting Shaped Modern Life*. New York: W. W. Norton & Company.

Daston, L. and Galison, P. (2010 [2007]). *Objectivity*. New York: Zone Books.

Daston, L. and Galison, P. (1992). The Image of Objectivity. *Representations*, 40, pp. 81-128.

Desrosières, A. (1998). *The Politics of Large Numbers. A History of Statistical Reasoning*. Cambridge, MA: Harvard University Press.

Dingens, J. (2005). *The Condor Years: How Pinochet and His Allies Brought Terrorism to Three Continents*. New York: The New Press.

European Commission (EC) (2015a). *Data for Policy: When the Haystack Is Made of Needles. A Call for Contributions*. Consultation. [online] Available at: http://ec.europa.eu/digital-agenda/en/news/data-policy-when-haystack-made-needles-call-contributions [Accessed 20 June 2018].

European Commission (EC) (2015b). *Making Big Data Work for Europe*. [online] Available at: https://ec.europa.eu/digital-single-market/en/making-big-data-work-europe [Accessed 20 June 2018].

European Commission (EC) (2014). *Towards a Thriving Data-Driven Economy*. Communication. [pdf] Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0232&from=EN [Accessed 20 June 2018].

European Commission (EC) (2013). *EU Leaders Call for Action on Digital Economy, Innovation and Services*. [pdf] Available at:

http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=3369 [Accessed 20 June 2018].

Giddens, A. (1990). *The Consequences of Modernity*. Cambridge: Polity Press.

Gitelman, L., ed., (2013). *'Raw Data' Is an Oxymoron*. Cambridge, MA: MIT Press.

Grandin, G. (2014). The Anti-Socialist Origins of Big Data. *The Nation*. [online] Available at: http://www.thenation.com/article/anti-socialist-origins-big-data/ [Accessed 20 June 2018].

Haskins, R. (2014). *Show Me the Evidence: Obama's Fight for Rigor and Results in Social Policy*. Washington, D.C.: The Brookings Institution.

Jasanoff, S. and Kim, S.-H. (2009). Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea. *Minerva*, 47(2), pp. 119-146.

Kalil, T. and Zhao, F. (2013). Unleashing the Power of Big Data. *The White House Blog*. [online] Available at: https://www.whitehouse.gov/blog/2013/04/18/unleashing-power-big-data [Accessed 20 June 2018].

Kitchin, R. (2014a). Big Data, New Epistemologies and Paradigm Shifts. *Big Data & Society*, 1(1), 1-12.

Kitchin, R. (2014b). *The Data Revolution. Big Data, Open Data, Data Infrastructures & Their Limitations*. London: Sage.

Kroes, N. (EC) (2014). *The Data Gold Rush*. [online] Available at: http://europa.eu/rapid/press-release_SPEECH-14-229_en.htm [Accessed 20 June 2018].

Kroes, N. (EC) (2012). *From Crisis of Trust to Open Governing*. [online] Available at: http://europa.eu/rapid/press-release_SPEECH-12-149_en.htm [Accessed 20 June 2018].

Laney, D. (2001). *3D Management: Controlling Data Volume, Velocity and Variety*. Available at: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf [Accessed 20 June 2018].

Leonelli, S. (2014). What Difference Does Quantity Make? On the Epistemology of Big Data in Biology. *Big Data & Society*, 1(1), 1-11.

Mackenzie, A. (2013). Programming Subjects in the Regime of Anticipation: Software Studies and Subjectivity. *Subjectivity*, 6(4), pp. 391-405.

Mayer-Schönberger, V. and Cukier, K. (2013). *Big Data. A Revolution That Will Transform How We Live, Work, And Think*. New York: Houghton Mifflin Harcourt.

Medina, E. (2011). *Cybernetic Revolutionaries: Technology and Politics in Allende's Chile*. Cambridge, Mass.: MIT Press.

Morozov, E. (2014). The Planning Machine. *The New Yorker*. [online] Available at: http://www.newyorker.com/magazine/2014/10/13/planning-machine [Accessed 20 June 2018].

Nowotny, H., Scott, P. and Gibbons, M. (2001). *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge: Polity Press.

Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, 57, pp. 1701-1777.

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.

Porter, T. M. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Porter, T. M. (1994). Objectivity as Standardization: The Rhetoric of Impersonality in Measurement, Statistics, and Cost-Benefit Analysis. In: A. Megill, ed., *Rethinking Objectivity*. Durham; London: Duke University Press.

Pentland, A. (2014). *Social Physics. How Good Ideas Spread – The Lessons from a New Science*. New York: The Penguin Press.

Rubinstein, I. (2013). Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law*, 3(2), pp. 74-87.

Shapin, S. (1994). *A Social History of Truth: Civility and Science in Seventeenth-Century England*. Chicago: The University of Chicago Press.

Solesbury, W. (2002). The Ascendancy of Evidence. *Planning Theory & Practice*, 3(1), pp. 90-96.

Sunstein, C. R. (2012). *Regulation in an Uncertain World*. National Academy of Sciences. [pdf] Available at: https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/speeches/regulation-in-an-uncertain-world-06202012.pdf [Accessed 20 June 2018].

Urahn, S. K. (2015). A Tipping Point on Evidence-Based Policymaking. *Governing*. [online] Available at: http://www.governing.com/columns/smart-mgmt/col-state-local-government-tipping-point-evidence-based-policymaking.html [Accessed 20 June 2018].

Venturini, T.; Laffite, N. B.; Cointet, J.-P.; Gray, I.; Zabban, V. and de Pryck, K. (2014). Three Maps and Three Misunderstandings: A Digital Mapping of Climate Diplomacy. *Big Data & Society*, 1(1), 1-19.

White House (2014). *Big Data: Seizing Opportunities, Preserving Values*. [pdf] Available at: https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [Accessed 20 June 2018].

Paper II

Big Data: A New Empiricism and Its Epistemic and Socio-Political Consequences

by Gernot Rieder and Judith Simon

Presented Version:

Accepted, peer-reviewed manuscript corrected for typographical errors and citation style

Published Version:

Rieder, G. and Simon, J. (2017). Big Data: A New Empiricism and its Epistemic and Socio-Political Consequences.
In: W. Pietsch, J. Wernecke and M. Ott, eds., *Berechenbarkeit der Welt? Philosophie und Wissenschaft im Zeitalter von Big Data*. Wiesbaden: Springer, pp. 85-105.

Big Data: A New Empiricism and Its Epistemic and Socio-Political Consequences

by Gernot Rieder and Judith Simon

Abstract

The paper investigates the rise of Big Data in contemporary society. It examines the most prominent epistemological claims made by Big Data proponents, calls attention to the potential socio-political consequences of blind data trust, and proposes a possible way forward. The paper's main focus is on the interplay between an emerging new empiricism and an increasingly opaque algorithmic environment that challenges democratic demands for transparency and accountability. It concludes that a responsible culture of quantification requires epistemic vigilance as well as a greater awareness of the potential dangers and pitfalls of an ever more data-driven society.

1. Introduction

The "Age of Big Data" (Lohr, 2012) is firmly upon us, and it promises to change not only how "we live, work, and think" (Mayer-Schönberger and Cukier, 2013), but also, and perhaps most fundamentally, *how we know*.

People, sensors, and systems generate increasingly large amounts of data. The networking company Cisco (2015) estimates that global Internet traffic has increased fivefold over the past five years, and will have tripled again by 2019. In the same year, driven by new users, products, and the quickly expanding Internet of Things, the number of Internet-connected devices is expected to reach 24 billion, compared to about 14 billion in 2015 (Ericsson, 2015).[1] Already there are "more data […] being generated every week than in the last millennia" (OECD, 2015: 1), and at a rate that is likely to accelerate.

But Big Data as a complex *techno-scientific phenomenon* is not just about growth in data "volume, velocity, and variety" (Laney, 2001), it is also seen as the ability to mine and manipulate data in ways that allow to "extract meaning and insight" and "reveal [hidden]

[1] Such forecasts should be regarded with caution. The numbers vary not only between research firms, but also between individual reports issued by the same company, and thus are quite volatile.

trends and patterns" (IBM, 2014: 11; 15). Consequently, and heavily lobbied by industry stakeholders, there have been considerable investments in analytical capabilities (i.e., hardware, software, and skills) across both public and private sectors. The European Commission, for instance, has only recently announced a €2.5 billion public-private partnership in an effort to put "Europe at the forefront of the global data race" and "master Big Data" (EC, 2014a).

Enthusiasts and advocates from research and industry have argued that Big Data presents a new scientific paradigm (see Hey et al., 2009), a data-intensive exploratory science with the "dream of establishing a 'sensors everywhere' data infrastructure" (Bell, 2009: xv), enabling us to "measure more, faster, than ever before" (Wilbanks, 2009: 214). From such a perspective, the impact of advanced data analytics is nothing short of revolutionary: In addition to transforming a wide array of areas such as health care (see Science Europe, 2014), education (see Dede, 2015), or law enforcement (see Bachner, 2013), Big Data is said to produce a new kind of knowledge, one that is more *comprehensive*, more *objective*, and more *predictive*.

Against this backdrop, scholars from the social sciences and humanities have warned of an emerging "new empiricism" (Kitchin, 2014a), a certain belief that "the volume of data, accompanied by techniques that can reveal their inherent truth, enables data to speak for themselves free of theory" (ibid.: 3). Taken to the extreme, such unbridled "trust in numbers" (Porter, 1995) is said to lead to "data-ism" (Brooks, 2013), a "deification of data" (Jenkins, 2013) that promotes forms of "algorithmic governance" (Williamson, 2014) and "digital age […] Taylorism" (Lohr, 2015) where a focus on performance and efficiency metrics replaces other norms and values.

Following a brief discussion of the rise of Big Data in contemporary society (2), this paper examines some of the most prominent epistemological claims made by Big Data proponents (3), calls attention to the potential socio-political ramifications of unrestrained 'datatrust' (4), and, in a concluding section, points to potential conceptual alternatives (5).

2. Big Data: Genesis, Definitions, Trust

While the exact origins of the term remain a matter of debate (see Lohr, 2013), bibliometric studies have documented a growing number of Big Data articles since the early 2000s, with a

sharp increase in publications since 2008 (see Halevi and Moed, 2012). At first, references to Big Data can mainly be found in the engineering and computer science literature; more recently, the use of the term has become widespread across a host of disciplines – from business and management to physics, biology, and medicine to the social sciences and the humanities. But the popularization of the notion has not been restricted to academia. A Google Trends search for "Big Data" indicates a strong surge in interest[2] since 2011, the tool's forecast feature predicting a further increase from 2015 onwards. At present, there appears to be hardly any major news outlet that has not dealt with the phenomenon in one way or another, and its heavy use as an advertising and marketing term has solidified its reputation as the "buzzword of the decade" (Barocas and Selbst, 2016: 673).

Many definitions of Big Data have been given, but no consensus has been reached. As Schroeder (2014: 5) notes, there are "no definitive, academic definitions of data and of Big Data". Probably best known, however, are the so-called "3Vs" suggested by former META Group (now Gartner) analyst Doug Laney (2012; 2001). According to this model, Big Data can be characterized as growth in data *volume* (i.e., a change in the depth and breadth of data available), *velocity* (i.e., an acceleration of data generation), and *variety* (i.e., a greater heterogeneity of data types and formats). But even when extended to four (see IBM, 2013), five (see Marr, 2014), or seven Vs (see van Rijmenam, 2013), frameworks of this kind mainly focus on *measures of magnitude* and related challenges, thus providing a very narrow view of what constitutes Big Data. Other approaches have shifted attention from data properties to new forms of analysis, conceptualizing Big Data as a "problem-solving philosophy" (Hartzog and Selinger, 2013) that "link[s] seemingly disparate disciplines" (Berman, 2013: xv), enabling researchers to "discover relationships" (Schaeffer and Olson, 2014: 44) and "harness information in novel ways to produce useful insights" (Mayer-Schönberger and Cukier, 2013: 2). While such a perspective offers a richer understanding of Big Data as a complex techno-scientific phenomenon, it cannot account for the current *climate of hype* that surrounds modern data analytics. This is where a third type of definition provides further insight. As boyd and Crawford (2012: 663) have argued, Big Data is not only about technological progress and advances in analytical techniques, but also about a "widespread belief that large data

[2] Google's "interest over time" graph shows "total searches for a term relative to the total number of searches done on Google", see: https://support.google.com/trends/answer/4355164?hl=en&rd=1 [Accessed 20 June 2018].

sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible". Leonelli (2014: 1) makes a similar argument, stressing that the novelty of Big Data science does not lie in the sheer quantity of data involved, but in the "prominence and status acquired by data as commodity and recognized output." It is arguably this *trust* in the authority of data, this *faith* and *belief* in numerical evidence, that has greatly contributed to the buzz-laden narrative of Big Data discourse.

The rhetoric of hope and hype is not limited to industry and business circles, it also features prominently in political realms. Policymakers have referred to Big Data as, e.g., the "new oil of the digital age" (Kroes, 2012), a game-changing "key asset" (EC, 2015a), or the next "industrial revolution" (Kroes, 2014). There appears to be agreement that "Big Data drives big benefits" (White House, 2014a: v), and that one must "seize the opportunities afforded by this new, data-driven revolution" (NSF, 2012). On the one hand, the conviction that data science can "change this […] world for the better" (Obama, 2015) is deeply rooted in the history of modern Western statecraft (see Rieder and Simon, 2016); on the other hand, it is sustained and nurtured by a very specific epistemic imaginary, a core set of knowledge claims that, although not entirely new, has (re)gained significant traction over the past few years. In the next section, we shall take a closer look at the composition of this imaginary, examining how it contributes to the current data hype.

3. The Rise of a New Empiricism

One of the most comprehensive critical overviews of Big Data epistemologies to date has been provided by Kitchin (2014a; b). In essence, Kitchin (2014a: 5) argues that in the context of Big Data one can observe the "articulation of a new empiricism", which "operates as a discursive rhetorical device designed to […] convince vendors [and other stakeholders] of the utility and value of Big Data analytics." But what exactly are the promises and claims associated with this emerging new paradigm? If Big Data serves as a potent rhetorical device, what are the ascribed powers? While an extensive discussion of the Big Data imaginary[3] is beyond the scope of this paper, a brief outline of certain key assumptions may allow for a

[3] For other mentions of such a distinct "Big Data imaginary", see Housley (2015) and Williamson (2015).

better understanding of the particular kind of knowledge Big Data practices are said to produce. It thus seems worthwhile to consider these assumptions in some more detail.

First, there is the notion that Big Data is exhaustive in scope, capable of capturing entire populations or domains (N=all) rather than being limited to sample-based surveys, allowing researchers to "get the complete picture" (Oracle, 2012: 3) and "make sense […] without traditional reduction" (Strawn, 2012: 34). This belief in what Lagoze (2014: 2) critically refers to as the "allness" of Big Data is, on the one hand, driven by the proliferation of digital data in today's increasingly computerized society and, on the other hand, the result of improved capacities to retrieve, store, and analyze those data. The expectations are high: At its best, Big Data is supposed to "give a view of life in all its complexity" (Pentland, 2014: 11), combining "millions, if not billions, of individual data points" to get "the full resolution on worldwide affairs" (Steadman, 2013). In addition, this vision of completeness is assumed to be (a) more inclusive and representative than other forms of research, "encompass[ing] thousands of times more people than a Gallup or Pew study" (Rudder, 2014: 20); (b) analytically superior in the sense that "the more data available the better and more accurate the results" (EC-BIO, 2013a: 13); and (c) more direct and unmediated as it reveals "what people actually *do* rather than what they *say they do*" (Strong, 2015: 2). This last aspect points to another central claim.

A second key imaginary holds that with enough volume the data speak for themselves, replacing "the narrative with the empirical" (Brooks, 2013) and eliminating any need for a priori theory. Instead, meaning is thought to emerge from the data "without human involvement" (Szal, 2015), rendering established forms of scientific inquiry – i.e., hypothesize, model, test – obsolete (see Anderson, 2008). Advanced algorithms are said to "find patterns where science cannot" (ibid.), shifting the focus from causal explanations to the discovery of statistical correlations that "inherently produce […] insightful knowledge about social, political and economic processes" (Kitchin, 2014b: 130). What follows is a reduction in the perceived relevance of context since "knowing *what*, not *why*, is good enough" (Mayer-Schönberger and Cukier, 2013: 52) to make "human systems […] run better and smarter" and "engineer a safer and healthier world" (Eagle & Greene, 2014: 153). Similarly, subject matter expertise and domain-specific knowledge is believed to "matter less" when "probability and correlation are paramount" (Mayer-Schönberger and Cukier, 2013: 16), suggesting that the pioneers and innovators of the Big Data era will "often come from fields outside the domain

where they make their mark" (ibid.: 142). Consequently, computer and data scientists rather than, e.g., physicians, biologists, or sociologists are considered the main protagonists of this new research paradigm (see Davenport and Patil, 2012).

The third imaginary overlaps with the second, but extends the argument even further: Not only are data seen as speaking for themselves, free of human intervention, they are also, by their very nature, expected to be fair and objective. More specifically, "agnostic statistics" (Anderson, 2008) are said to "eliminate human bias" (Richtel, 2013) from decision-making processes, offering an impartial "view from nowhere" that reveals new truths and provides a "disinterested picture of reality" (Jurgenson, 2014). Replacing "gut and intuition" with "numbers and metrics" (Gutierrez, 2015), fully automated software is supposed to deliver fact-based recommendations, acting as a neutral corrective to people's "subjective judgments and hunches" (Clinton, 2016). While confidence in the virtues of quantification is no new phenomenon (e.g., see Cohen, 2005; Porter, 1995; Hacking, 1990), the promise of "algorithmic objectivity" (Gillespie, 2014: 168) has raised hopes that modern data analytics may serve as a "powerful weapon in the fight for equality" (Castro, 2014), battling discrimination across a broad range of sectors – from employment and education to law enforcement and health care to housing and credit – "empower[ing] vulnerable groups" and "ensur[ing] equal opportunity for all" (FPF, 2014). The idea is both simple and compelling: The more mechanized the process, the higher the chance that the results won't be tainted by researchers' interpretive subjectivity (see Venturini et al., 2014) – after all, "it's humans, not algorithms, that have a bias problem" (New, 2015).

Fourth and finally, the application of sophisticated Big Data techniques is meant to provide "certainty in an uncertain world" (AppDynamics, 2014), generating reliable knowledge for robust, evidence-informed decision making. As Hardy (2013) notes, the "promise of certainty has been a hallmark of the technology industry for decades", and the ability to reduce ambiguity, establish clarity, and determine risk is often touted as a major benefit of advanced data analytics. Thriving on rather than drowning in information overload, Big Data methods are expected to find the 'signal in the noise', the 'needle in the haystack', 'connecting the dots' with high precision and accuracy. What is particularly noteworthy is the scope of the claim: Governing principles – much like physical laws – are believed to "undergird virtually every interaction in society" (Silver, 2012: 53), and the arrival of dense, continuous data together with modern computation makes it possible to detect "statistical regularities

that […] are true of almost everyone almost all of the time" (Pentland, 2014: 189), offering valuable "insights about human nature" (203). Once 'reality' has been mined[4] and patterns are found, "more and more aspects of our lives […] become predictable" (EC-BIO, 2013b: 3), allowing us to tame uncertainty and "sens[e] the future before it occurs" (Fitzgerald, 2012). If to "know ahead [and] act before" (Quantacast, 2013) is the industry's trademarked mantra, the "end of chance" (Müller et al., 2013) is its ultimate goal.

There are, of course, a number of other epistemological beliefs and assumptions that feed into the Big Data imaginary – such as the idea that "unless something can be measured, it cannot be improved" (Kelly, 2007) or that "the law of large numbers […] evens out the errors of any individual data point" (Phillips Mandaville, 2014) – but these are arguably an extension of, or at least closely related to, the four central propositions outlined above: Namely, that Big Data represents nothing less than a computational means to know everything (I), of anything (II), free from bias (III), with a high degree of certainty (IV).

Scholars from various disciplines have challenged the bold claims of Big Data empiricism, arguing that modern analytical techniques neither provide a complete picture of entire populations (see McFarland and McFarland, 2015) nor eradicate the need for models and a priori theory (see Pigliucci, 2009), are neither neutral or free of bias (see Hardt, 2014) nor have the ability to predict with certainty (see Silver, 2012). Instead, Big Data collections can be "small" and "partial" (Leonelli, 2014); algorithms may "perpetuate the prejudices of their creators" (Centre for Internet and Human Rights, 2015: 5) or "learn bias from the data fed into them" (Kun, 2015); and forecasts are never certain, but deal in probabilities, possibilities, and uncertainties that may be "narrow in reach, scope, and perspective" (Ekbia et al., 2013: 1539). Yet despite signs of growing awareness – e.g., the recent White House report *Big Data: Seizing Opportunities, Preserving Values* (White House, 2014b: 64) emphasizes the need for a "national conversation on big data, discrimination, and civil liberties" – the ideal of "impersonal rationality achieved through technical methods" (Porter, 2011: 46) continues to act as an important techno-political leitmotif. As Hildebrandt (2013: 28f) observes, "we have trouble resisting the seemingly clean, objective knowledge [Big Data] produces", and turn it,

[4] For more on the concept of reality mining, see Eagle and Pentland (2006) and Eagle and Greene (2014).

by making ourselves dependent upon its oracles, "into a new pantheon, filled with novel gods." But what are the dangers of being enthralled by the power and possibilities of modern computing, of placing our faith in the veracity of ever more widespread predictive analytics? If society were to follow the path toward digital serfdom, what are the costs of becoming "slaves to Big Data" (ibid.)?

4. A Black Box Society

Discussions about the social and political ramifications of Big Data have predominantly focused on two interrelated sets of issues: questions of *privacy* (e.g., regarding surveillance, profiling, or data protection/security) on the one hand and instances of *discrimination* (e.g., through differential access or treatment) on the other. Journalists and academics have examined a variety of analytical tools and techniques – from Target's pregnancy-prediction model (see Duhigg, 2012) and the City of Boston's StreetBump app (see Crawford, 2013) to Google's Flu Trend service (see Lazer et al., 2014) and Facebook's 'emotional contagion' study (see Meyer, 2014) – exposing both methodological biases and ethical transgressions.

While reports of this kind have contributed to public awareness and dialogue, the focus on a limited number of high-profile cases may give the impression that there is only cause for concern if things have either gone 'wrong' or 'too far', that is, if the analytical process is flawed or there is clear indication of unethical or illegal conduct. But this is not the case. In fact, even if done 'right', data mining may reflect "widespread biases that persist in society at large" (Barocas and Selbst, 2016: 671), and can thus "affect the fortunes of whole classes of people in consistently unfavorable ways" (673). Similarly, the strong reactions to Facebook's emotional contagion study, in which researchers altered the News Feeds of almost 700,000 users (see Kramer et al., 2014), fail to take into account that "manipulating the News Feed is Facebook's entire business" (Patel, 2014). To be clear, the point here is not to defend any particular study or experiment; rather, it is to emphasize that the issues and concerns run much deeper, are systemic rather than case-specific, cultural rather than attributable to a few outliers. The main impact of the Big Data phenomenon is not that smart TVs may be listening to "everything you say" (Harris, 2015), that social media posts "may damage your credit score" (McLannahan, 2015), or that the police might pay you a visit because your name appeared on a software-generated "heat list" (Gorner, 2013); much

rather, it is that Big Data with its "aura of truth [and] objectivity" (see boyd & Crawford, 2012: 663) contributes to a specific "algorithmic culture" (Striphas, 2015) that renders such practices not only "the new normal" (Andrejevic, 2013: 30), but also socially acceptable.[5]

The consequences may be quite severe. Pasquale (2015), for instance, cautions of an emerging "black box society", i.e., a "system whose workings are mysterious" (3), where the "distinction between state and market is fading" (10), and where people submit to the "dictate of salient, measurable data" (10), the "rule of scores and bets" (191). Others have voiced similar concerns, arguing that unquestioned faith in the seemingly impartial workings of the machine may undermine civil liberties (see Al-Rodhan, 2014), threaten social and economic justice (see Newman, 2015), and prove "toxic to democratic governance and […] democracy itself" (Howard, 2014). But what exactly are the problems and challenges associated with this brave new data world? Though there are overlaps, it is possible to distinguish between three main issues and concerns: opacity, accountability, responsibility.

Opacity – While the use of algorithmic decision-making tools by governments and private entities has grown progressively (see Zarsky, 2016), analytical processes are often opaque, operating according to rules that are hidden, using input data that remain unknown. This lack of transparency has several reasons. For one, as Burrell (2016) highlights, opacity can be the result of intentional *corporate or state secrecy*, e.g., in order to secure a competitive advantage, to shield an algorithm from being 'gamed', or to avoid regulation and control. Second, opacity may be a consequence of high algorithmic *complexity*, especially in the case of machine learning applications. While it might be possible to "untangle the logic of the code within a complicated software system" (ibid.: 5), being able to understand the algorithm in action as it operates on data may not be feasible as such machine optimizations "do not naturally accord with human semantic explanations" and hence "escape full understanding and interpretation" (10). As Gillespie (2014: 192) argues, there appears to be something "impenetrable about algorithms", even for well-trained programmers and computer scientists. Finally, there is the problem of *volatility*. As Facebook engineers explain, "code is not a fixed artifact but an evolving system, updated frequently and concurrently by

[5] Such social acceptance is actually quite contradictory: While a study of the Pew Research Center (2014) finds that 91% of the surveyed adults "agree" or "strongly agree" that consumers have lost control over how personal information is collected and used by companies, Internet and technology giants such as Amazon, Google, or Apple regularly rank not only amongst the most valuable (Ember, 2015), but also amongst the most reputable corporations (Adams, 2015).

many developers" (Calcagno et al., 2015). Google, for instance, is known to change its search algorithm around 500-600 times a year, including both minor and major updates (see Moz, 2016). In that sense, code may not only be secret and opaque, it may also be quite elusive.

Accountability – In a world where Big Data software takes command, "accountability is often murky" (Rosenblat et al., 2014: 1). Internet companies, in particular, seek to avoid scrutiny and deflect concerns regarding their services and products. Whenever there is public outcry – as in the case of Facebook's emotional contagion study – the giants of the Web are quick to apologize, acknowledging that they "did a bad job", "really messed up", or "missed the mark" (Isaac, 2014). But the tinkering and testing continues – without much oversight, meaningful public deliberation, or serious legal or financial consequences.[6] New legislative instruments, such as the EU's proposed General Data Protection Regulation (EC, 2012), are supposed to clarify industry obligations by introducing "binding corporate rules" (ibid.). Legal scholars, however, have warned that current reform efforts do not go far enough as Big Data "defeats traditional privacy law by undermining core principles and regulatory assumptions" (Rubinstein, 2013: 75), including, for instance, the informed choice model, the distinction between personal and non-personal data, or requirements of data minimization. Thus, the overall picture is somewhat bleak: While current laws seem unable to enforce greater corporate accountability, data-mining firms show little interest in disclosing details about their business practices, a sentiment not unique to the private sector as the Snowden revelations have made abundantly clear (see Greenwald, 2014).

Responsibility – Professional software systems are usually not created by a single person, but by groups of people with different skills and expertise working in institutional settings. If a system fails and causes harm, determining individual responsibility may prove difficult since "responsibility […] does not easily generalize to collective action" (Nissenbaum, 1996: 29). The question of who is to blame becomes even more challenging now that 'smart' technologies take over a growing number of knowledge and decision-making processes. Especially in machine learning, where "computers [are given] the ability to learn without being explicitly programmed" (Samuel, 1959; cited after Simon, 2013: 89), the "influence of

[6] The monetary fines imposed on large technology companies have often been insignificant compared to the revenues their services generate (Pasquale, 2015). More recently, however, EU regulators have increased pressure by filing antitrust charges against Google, which could lead to fines of more than €6 billion – about 10 percent of the company's 2014 revenue (Kanter and Scott, 2015).

the creator over the machine decreases [while] the influence of the operating environment increases" (Matthias, 2004: 182), giving rise to so-called autonomous agents whose actions can be difficult to predict. Companies have sometimes emphasized this autonomy to shirk responsibility. Google, for example, maintains that its autocomplete suggestions, which have repeatedly been criticized for discriminating against protected classes (see UN Women, 2013) and defaming individuals (see Niggemeier, 2012), are "generated by an algorithm without any human involvement" and merely "reflect what other people are searching for" (Google, 2016). Courts around the globe have arrived at different verdicts – from imposing fines and ordering the removal of specific autocomplete suggestions (see Valinsky, 2013) to discharging the search company from liability (see Bellezza and De Santis, 2013) – demonstrating that in Big Data contexts, questions of responsibility are neither obvious nor easy.

Government and corporate secrecy paired with technical inscrutability; obsolescent legal safeguards that are no match for new forms of "digital feudalism" (Clark, 2011); algorithmic scapegoating to avoid responsibility and curtail agency – these are the main ingredients of a thoroughly black-boxed data economy, in which "opaque technologies are spreading, unmonitored and unregulated" (Pasquale, 2015: 14). Such technologies do not just describe, but actively create social realities: They rank and recommend, classify and score, predict and prescribe, exercising social control through numerical judgment. What they offer are seemingly innocuous automated results; what they produce are specific systems of order. Algorithmic regulation is often described as fair and objective, and the epistemic claims of Big Data empiricism reinforce this image: By claiming to capture everything in great detail, Big Data diverts attention from what remains unseen; by claiming general applicability, it discourages critical debate about where data analytics should or should not be applied; by claiming unbiased neutrality, it discursively impedes public scrutiny; by claiming empirical certainty, it fends off doubts about analytical veracity. The epistemological promises of Big Data empiricism are both powerful and seductive, and if not challenged will contribute to an algorithmic culture that may make us abandon all checks and balances.

5. Conclusions

As outlined, the bold claims of Big Data empiricism can be contested both on epistemological grounds and for their socio-political consequences. Given the amount of criticism these claims have received not only from SSH scholars but also from computer and natural scientists (e.g., see O'Neil, 2016; Hardt, 2014), their continued prominence in business and policy contexts may come as a surprise. But while the proliferation of Big Data excitement within the business sector may be attributed to vested commercial interests (see IDC, 2015), the uncritical reception in certain political domains can be understood as but the latest manifestation of a more generic "trust in numbers" (Porter, 1995). Numbers provide authority and justify decision making, especially in times of crisis and uncertainty. The European Commission's recent push for "evidence-informed policy making" (EC, 2014b) is a telling example.

While we acknowledge the potential benefits of "data for policy" (EC, 2015b), we hold that a responsible *culture of quantification* requires greater awareness of the possible dangers and pitfalls of an increasingly data-driven society. As "governance by algorithms" (Musiani, 2013) is becoming ever more commonplace, a public conversation about the politics of algorithmic regulation and what it means to be a "citizen of a software society" (Manovich, 2013) is crucial.

What is needed is "epistemic vigilance" (Sperber et al., 2010) instead of blind data trust. Such vigilance requires, on the one hand, access to both the data used and the algorithms applied and, on the other hand, competencies to understand the analytical process as an embedded socio-technical practice, a specific way of producing knowledge that is neither inherently objective nor unquestionably fair. Thus, while legally guaranteed access to data and algorithmic code is important, it is not sufficient; what must also be cultivated are capabilities to interpret and possibly contest algorithmic data practices. The more such automated processes shape our society, the greater the need to invest in the education of researchers and policymakers, but also of journalists and ordinary citizens, to facilitate a better understanding of the epistemic foundations and socio-political ramifications of pervasive data analytics. Without doubt, this will be a difficult task that may require not only educational measures, but also technological intervention. On the one hand, we are talking about tools and services that can support epistemic vigilance by making visible what is often thoroughly black-boxed and concealed. On the other hand, technology design can also be

used as a means to implement ethically sound software solutions as a form of *governance by design*.

The importance of epistemic vigilance cannot be overstated: Any political, ethical, or legal assessment of Big Data practices hinges upon a proper understanding of the epistemological foundations – and limitations – of algorithmic knowledge production.

While transparency and education are aimed at tackling the problem of opacity, there is also a need to address issues of distributed agency and the challenges they pose for accountability and responsibility attribution. As noted, algorithmic systems are complex assemblages of human and non-human actors and identifying a single culprit – be it man or machine – can be difficult, if not impossible. Moreover, if disparate impact is not the result of intentional discrimination, but of widespread biases that exist in society at large (see Barocas and Selbst, 2016), who can ultimately be held responsible for the potentially severe real-life consequences of Big Data analytics? To deal with these distributed forms of agency, new concepts to understand and enforce socio-technically distributed responsibility must be developed (see Simon, 2015; Floridi, 2013).

In conclusion, if we wish to tackle the epistemological and socio-political challenges of a new empiricism, we must act on several fronts: Hard law, soft law, education and ethically-informed software design all need to be employed for a "good enough governance" (Pagallo, 2015: 164) of Big Data practices.

Acknowledgments

The authors wish to acknowledge the financial support of the Austrian Science Fund (P-23770).

References

Adams, S. (2015). The World's Most Reputable Companies in 2015. *Forbes*. [online] Available at: https://www.forbes.com/sites/susanadams/2015/04/21/the-worlds-most-reputable-companies-in-2015/#ae7c7a1425c6 [Accessed 20 June 2018].

Al-Rodhan, N. (2014). The Social Contract 2.0: Big Data and the Need to Guarantee Privacy and Civil Liberties. *Harvard International Review*. [online] Available at: http://hir.harvard.edu/the-social-contract-2-0-big-data-and-the-need-to-guarantee-privacy-and-civil-liberties/ [Accessed 20 June 2018].

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. [online] Available at: https://www.wired.com/2008/06/pb-theory/ [Accessed 20 June 2018].

Andrejevic, M. (2013). *Infoglut: How Too Much Information Is Changing the Way We Think and Know*. New York: Routledge.

AppDynamics (2014). *AppDynamics Announces New Application Intelligence Platform*. [online] Available at: https://www.appdynamics.com/press-release/appdynamics-announces-new-application-intelligence-platform/ [Accessed 20 June 2018].

Bachner, J. (2013). Predictive Policing: Preventing Crime with Data and Analytics. *IBM Center for the Business of Government. Improving Performance Series*. [pdf] Available at: http://www.businessofgovernment.org/sites/default/files/Predictive%20Policing.pdf [Accessed 20 June 2018].

Barocas, S. and Selbst, A. D. (2016 [2014]). Big Data's Disparate Impact. *California Law Review*, 104, pp. 671-732.

Bell, G. (2009). Foreword. In: T. Hey, S. Tansley and K. Tolle, eds., *The Fourth Paradigm. Data-Intensive Scientific Discovery*. Redmond: Microsoft Research, pp. XI-XV.

Bellezza, M. and De Santis, F. (2013). Google Not Liable for Autocomplete and Related Search Results, Italian Court Rules. *CGCS Media Wire*. [online] Available at: http://global.asc.upenn.edu/google-not-liable-for-autocomplete-and-related-search-results-italian-court-rules/ [Accessed 20 June 2018].

Berman, J. J. (2013). *Principles of Big Data. Preparing, Sharing, and Analyzing Complex Information*. Amsterdam: Elsevier, Morgan Kaufmann.

boyd, d. and Crawford, K. (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, 15(5), pp. 662-679.

Brooks, D. (2013). The Philosophy of Data. *The New York Times*. [online] Available at: https://www.nytimes.com/2013/02/05/opinion/brooks-the-philosophy-of-data.html [Accessed 20 June 2018].

Burrell, J. (2016). How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society*, 3(1), pp. 12.

Calcagno, C., Distefano, D. and O'Hearn, P. (2015*). Open-Sourcing Facebook Infer: Identify Bugs Before You Ship*. [online] Available at: https://code.fb.com/developer-tools/open-sourcing-facebook-infer-identify-bugs-before-you-ship/ [Accessed 20 June 2018].

Castro, D. (2014). Big Data Is a Powerful Weapon in the Fight for Equality. *The Hill*. [online] Available at: http://thehill.com/blogs/pundits-blog/technology/221583-big-data-is-a-powerful-weapon-in-the-fight-for-equality [Accessed 20 June 2018].

Centre for Internet and Human Rights (2015). *Should Algorithms Decide Your Future?* [pdf] Available at: https://cihr.eu/wp-content/uploads/2015/07/EoA_web.pdf [Accessed 20 June 2018].

Cisco (2015). *Cisco Visual Networking Index: Forecast and Methodology, 2014–2019*. [pdf] Available at: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html [Accessed 20 June 2018].

Clark, J. (2011). Facebook, Google: Welcome to the New Feudalism. *ZDNet*. [online] Available at: http://www.zdnet.com/article/facebook-google-welcome-to-the-new-feudalism/ [Accessed 20 June 2018].

Clinton, R. (2016). What's the Difference between Business Intelligence and Predictive Analytics? *Smart Vision Europe*. [online] Available at: https://www.sv-europe.com/blog/whats-the-difference-between-business-intelligence-and-predictive-analytics/ [Accessed 20 June 2018].

Cohen, B. I. (2005). *Triumph of Numbers: How Counting Shaped Modern Life*. New York: W. W. Norton & Company.

Crawford, K. (2013). The Hidden Biases in Big Data. *Harvard Business Review*. [online] Available at: https://hbr.org/2013/04/the-hidden-biases-in-big-data [Accessed 20 June 2018].

Davenport, T. H. and Patil, D. J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*. [online] Available at: https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century [Accessed 20 June 2018].

Dede, C. (2015). *Data-Intensive Research in Education: Current Work and Next Steps*. [pdf] Available at: https://cra.org/wp-content/uploads/2015/10/CRAEducationReport2015.pdf [Accessed 20 June 2018].

Duhigg, C. (2012). How Companies Learn Your Secrets. *The New York Times*. [online] Available at: https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html [Accessed 20 June 2018].

Eagle, N. and Greene, K. (2014). *Reality Mining: Using Big Data to Engineer a Better World*. Cambridge, MA: MIT Press.

Eagle, N., and Pentland, A. (2006). Reality Mining: Sensing Complex Social Systems. *Personal and Ubiquitous Computing*, 10(4), pp. 255-268.

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V. R., Tsou, A., Weingart, S. and Sugimoto, C. R. (2015). Big Data, Bigger Dilemmas: A Critical Review. *Journal of the Association for Information and Technology*, 66(8), pp. 1523-1545.

Ember, S. (2015). Tech Giants Top Best Global Brands List. *The New York Times*. [online] Available at: http://www.nytimes.com/2015/10/05/business/media/tech-giants-top-best-global-brands-list.html [Accessed 20 June 2018].

Ericsson (2015). *Ericsson Mobility Report. On the Pulse of the Networked Society*. [pdf] Available at: http://www.ericsson.com/res/docs/2015/ericsson-mobility-report-june-2015.pdf [Accessed 20 June 2018].

European Commission (EC) (2015a). *Big Data*. [website] Available at: https://ec.europa.eu/digital-single-market/big-data [Accessed 20 June 2018].

European Commission (EC) (2015b). *Data for Policy: When the Haystack Is Made of Needles. A Call for Contributions*. Consultation. [online] Available at: https://ec.europa.eu/digital-single-market/news/data-policy-when-haystack-made-needles-call-contributions [Accessed 20 June 2018].

European Commission (EC) (2014a). *European Commission and Data Industry Launch €2.5 Billion Partnership to Master Big Data*. [online] Available at: http://europa.eu/rapid/press-release_IP-14-1129_en.htm [Accessed 20 June 2018].

European Commission (EC) (2014b). *Data Technologies for Evidence-Informed Policy Making (Including Big Data)*. [online] Available at: https://ec.europa.eu/digital-single-market/news/data-technologies-evidence-informed-policy-making-including-big-data-smart-20140004 [Accessed 20 June 2018].

European Commission (EC) (2012). Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data (General Data Protection Regulation). [pdf] Available at: http://web.archive.org/web/20140321203049/http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf [Accessed 20 June 2018].

European Commission – Business Innovation Observatory (EC-BIO) (2013a). *Big Data: Analytics and Decision Making*. [pdf] Available at: http://ec.europa.eu/DocsRoom/documents/13411/attachments/1/translations/en/renditions/native [Accessed 20 June 2018].

European Commission – Business Innovation Observatory (EC-BIO) (2013b). *Big Data: Artificial Intelligence*. [pdf] Available at: http://ec.europa.eu/DocsRoom/documents/13411/attachments/2/translations/en/renditions/native [Accessed 20 June 2018].

European Data Protection Supervisor (EDPS) (2015). *Meeting the Challenges of Big Data. A Call for Transparency, User Control, Data Protection by Design and Accountability*. [pdf] Available at: https://edps.europa.eu/sites/edp/files/publication/15-11-19_big_data_en.pdf [Accessed 20 June 2018].

Fitzgerald, M. (2012). Sensing the Future Before it Occurs. *MIT Sloan Management Review*. [online] Available at: https://sloanreview.mit.edu/article/sensing-the-future-before-it-occurs/ [Accessed 20 June 2018].

Floridi, L. (2013). Distributed Morality in an Information Society. *Science and Engineering Ethics* 19(3): pp. 727-743.

Future of Privacy Forum (FPF) (2014). *Big Data: A Tool for Fighting Discrimination and Empowering Groups*. [online] Available at: https://fpf.org/2014/09/11/big-data-a-tool-for-fighting-discrimination-and-empowering-groups/ [Accessed 20 June 2018].

Gillespie, T. (2014). The Relevance of Algorithms. In: T. Gillespie, P. J. Boczkowski and A. F. Kirsten, eds., *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: MIT Press, pp. 167-194.

Google (2016). *Search Help: Autocomplete*. [website] Available at: https://support.google.com/websearch/answer/106230?hl=en [Accessed 20 June 2018].

Gorner, J. (2013). Chicago Police Use 'Heat List' as Strategy to Prevent Violence. *Chicago Tribune*. [online] Available at: http://articles.chicagotribune.com/2013-08-21/news/ct-met-heat-list-20130821_1_chicago-police-commander-andrew-papachristos-heat-list [Accessed 20 June 2018].

Greenwald, G. (2014). *No Place to Hide: Edward Snowden, the NSA and the Surveillance State*. London: Penguin Books.

Gutierrez, D. (2015). Will Big Data Kill the Art of Marketing? *Inside Big Data*. [online] Available at: http://insidebigdata.com/2015/01/16/will-big-data-kill-art-marketing/ [Accessed 20 June 2018].

Hacking, I. (1990). *The Taming of Chance*. Cambridge: Cambridge University Press.

Halevi, G. and Moed, H. (2012). The Evolution of Big Data as a Research and Scientific Topic. *Research Trends*, 30, [online] Available at: https://www.researchtrends.com/issue-30-september-2012/the-evolution-of-big-data-as-a-research-and-scientific-topic-overview-of-the-literature/ [Accessed 20 June 2018].

Hardt, M. (2014). How Big Data Is Unfair: Understanding Sources of Unfairness in Data Driven Decision Making. *Medium*. [online] Available at: https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de#.8ix0w5v59 [Accessed 20 June 2018].

Hardy, Q. (2013). Why Big Data is not Truth. *The New York Times*. [online] Available at: https://bits.blogs.nytimes.com/2013/06/01/why-big-data-is-not-truth/ [Accessed 20 June 2018].

Harris, S. (2015). Your Samsung SmartTV Is Spying on You, Basically. *The Daily Beast*. [online] Available at: https://www.thedailybeast.com/your-samsung-smarttv-is-spying-on-you-basically [Accessed 20 June 2018].

Hartzog, W. and Selinger, E. (2013). Big Data in Small Hands. *Stanford Law Review Online*. [online] Available at: https://www.stanfordlawreview.org/online/privacy-and-big-data-big-data-in-small-hands/ [Accessed 20 June 2018].

Hey, T., Tansley, S. and Tolle, K., eds., (2009). *The Fourth Paradigm. Data-Intensive Scientific Discovery*. Redmond: Microsoft Research.

Hildebrandt, M. (2013). Slaves to Big Data. Or Are We? *IDP. Revista De Internet, Derecho Y Política*, 17, pp. 7-44.

Housley, W. (2015). Focus: The Emerging Contours of Data Science. *Discovery Society*. [pdf] Available at: http://discoversociety.org/2015/08/03/focus-the-emerging-contours-of-data-science/ [Accessed 20 June 2018].

Howard, A. (2014). Data-Driven Policy and Commerce Requires Algorithmic Transparency. *TechRepublic*. [online| Available at: https://www.techrepublic.com/article/data-driven-policy-and-commerce-requires-algorithmic-transparency/ [Accessed 20 June 2018].

IBM (2014). *IBM Watson Explorer: Search, Analyze, and Interpret to Enable Cognitive Exploration*. [pdf] Available at: https://www.ibm.com/developerworks/community/files/basic/anonymous/api/library/1b22311f-c1ab-4a8c-8b27-ec9f39115adf/document/476814cc-636a-47fe-93c5-63c4b66c3b9b/media [Accessed 20 June 2018].

IBM (2013). The Four V's of Big Data. *Big Data & Analytics Hub*. [website] Available at: http://www.ibmbigdatahub.com/infographic/four-vs-big-data#comment-1011454161 [Accessed 20 June 2018].

International Data Corporation (IDC) (2015). New IDC Forecast Sees Worldwide Big Data Technology and Services Market Growing to $48.6 Billion in 2019, Driven by Wide Adoption Across Industries. *Business Wire*. [online] Available at: https://www.businesswire.com/news/home/20151109005070/en/New-IDC-Forecast-Sees-Worldwide-Big-Data [Accessed 20 June 2018].

Isaac, M. (2014). Facebook Says It's Sorry. We've Heard That Before. *The New York Times*. [online] Available at: http://nyti.ms/1mBqRKp [Accessed 20 June 2018].

Jenkins, T. (2013). Tiffany Jenkins: Don't Count on Big Data for Answers. *The Scotsman*. [online] Available at: https://www.scotsman.com/news/opinion/tiffany-jenkins-don-t-count-on-big-data-for-answers-1-2785890 [Accessed 20 June 2018].

Jurgenson, N. (2014). View From Nowhere. *The New Inquiry*. [online] Available at: https://thenewinquiry.com/view-from-nowhere/ [Accessed 20 June 2018].

Kanter, J. and Scott, M. (2015). Europe Challenges Google, Seeing Violations of Its Antitrust Law. *The New York Times*. [online] Available at: http://nyti.ms/1D0lnRX [Accessed 20 June 2018].

Kelly, K. (2007). What Is the Quantified Self? *Quantified Self. Self Knowledge Through Numbers*. [online] Available at: https://web.archive.org/web/20150408202734/http://quantifiedself.com/2007/10/what-is-the-quantifiable-self/ [Accessed 20 June 2018].

Kitchin, R. (2014a). Big Data, New Epistemologies and Paradigm Shifts. *Big Data & Society*, 1(1), pp. 1-12.

Kitchin, R. (2014b). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London; Thousand Oaks; New Delhi: SAGE.

Kramer, A. D. I., Guillory, J. E. and Hancock, J. T. (2014). Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks. *Proceedings of the National Academy of Science of the United States of America*, 111(24), 8788-8790.

Kroes, N. (2014). *The Data Gold Rush*. [online] Available at: http://europa.eu/rapid/press-release_SPEECH-14-229_en.htm [Accessed 20 June 2018].

Kroes, N. (2012). *From Crisis of Trust to Open Governing*. [online] Available at: http://europa.eu/rapid/press-release_SPEECH-12-149_en.htm [Accessed 20 June 2018].

Kun, J. (2015). Big Data Algorithms Can Discriminate, and It's Not Clear What to Do About It. *The Conversation*. [online] Available at: https://theconversation.com/big-data-algorithms-can-discriminate-and-its-not-clear-what-to-do-about-it-45849 [Accessed 20 June 2018].

Lagoze, C. (2014). Big Data, Data Integrity, and the Fracturing of the Control Zone. *Big Data & Society*, 1(2), pp. 1-11.

Laney, D. (2012). Deja VVVu: Others Claiming Gartner's Construct for Big Data. *Gartner Blog Network*. [online] Available at: https://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/ [Accessed 20 June 2018].

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. *META Group. Application Delivery Strategies*. [pdf] Available at: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf [Accessed 20 June 2018].

Lazer, D., Kennedy, R., King, G. and Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 343(6175), pp. 1203-1205.

Leonelli, S. (2014). What Difference Does Quantity Make? On the Epistemology of Big Data in Biology. *Big Data & Society*, 1(1), pp. 1-11.

Lohr, S. (2015). *Data-ism. The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else*. New York: Harper Collins.

Lohr, S. (2013). The Origins of 'Big Data': An Etymological Detective Story. *The New York Times*. [online] Available at: https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/ [Accessed 20 June 2018].

Lohr, S. (2012). The Age of Big Data. *The New York Times*. [online] Available at: https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html [Accessed 20 June 2018].

Manovich, L. (2013). The Algorithms of Our Lives. *The Chronicle Review*. [online] Available at: https://www.chronicle.com/article/The-Algorithms-of-Our-Lives-/143557 [Accessed 20 June 2018].

Marr, B. (2014). Big Data: The 5 Vs Everyone Must Know. *LinkedIn*. [online] Available at: https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know [Accessed 20 June 2018].

Matthias, A. (2004). The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. *Ethics and Information Technology*, 6(3), pp. 175-183.

Mayer-Schönberger, V. and Cukier, K. (2013) *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. New York: Houghton Mifflin Harcourt.

McFarland, D. A. and McFarland, R. H. (2015). Big Data and the Danger of Being Precisely Inaccurate. *Big Data & Society*, 2(2), pp. 1-4.

McLannahan, B. (2015). Being 'Wasted' on Facebook May Damage Your Credit Score. *The Financial Times*. [online] Available at: http://www.ft.com/cms/s/0/d6daedee-706a-11e5-9b9e-690fdae72044.html [Accessed 20 June 2018].

Meyer, R. (2014). Everything We Know About Facebook's Secret Mood Manipulation Experiment. *The Atlantic*. [online] Available at: https://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/ [Accessed 20 June 2018].

Moz (2016). *Google Algorithm Change History*. [website] Available at: https://moz.com/google-algorithm-change [Accessed 20 June 2018].

Müller, M. U., Rosenbach, M. and Schulz, T. (2013). Living by the Numbers: Big Data Knows What Your Future Holds. *Spiegel Online International*. [online] Available at: http://www.spiegel.de/international/business/big-data-enables-companies-and-researchers-to-look-into-the-future-a-899964.html [Accessed 20 June 2018].

Musiani, F. (2013). Governance by Algorithms. *Internet Policy Review*, 2(3), pp. 1-8.

National Science Foundation (NSF) (2012). NSF Leads Federal Efforts in Big Data. (online) Available at: https://www.nsf.gov/news/news_summ.jsp?cntn_id=123607&org=NSF&from=news [Accessed 20 June 2018].

New, J. (2015). It's Humans, Not Algorithms, That Have a Bias Problem. *Center for Data Innovation*. [online] Available at: https://www.datainnovation.org/2015/11/its-humans-not-algorithms-that-have-a-bias-problem/ [Accessed 20 June 2018].

Newman, N. (2015). Data Justice: Taking on Big Data as an Economic Justice Issue. [pdf] Available at: http://www.datajustice.org/sites/default/files/Data%20Justice-%20Taking%20on%20Big%20Data%20as%20an%20Economic%20Justice%20Issue.pdf [Accessed 20 June 2018].

Niggemeier, S. (2012). Autocompleting Bettina Wulff: Can a Google Function Be Libelous? *Spiegel Online*. [online] Available at: http://www.spiegel.de/international/zeitgeist/google-autocomplete-former-german-first-lady-defamation-case-a-856820.html [Accessed 20 June 2018].

Nissenbaum, H. (1996). Accountability in a Computerized Society. *Science and Engineering Ethics*, 2(1), pp. 25-42.

Obama, B. (2015). *President Barack Obama's Big Data Keynote*. [video] Available at: https://www.youtube.com/watch?v=vbb-AjiXyh0 [Accessed 20 June 2018].

O'Neil, C. (2016). The Ethical Data Scientist. *Slate*. [online] Available at: http://www.slate.com/articles/technology/future_tense/2016/02/how_to_bring_better_ethics_to_data_science.html [Accessed 20 June 2018].

Oracle (2012). *Integrate for Insight: Combining Big Data Tools with Traditional Data Management*. [pdf] Available at: http://www.oracle.com/us/technologies/big-data/big-data-strategy-guide-1536569.pdf [Accessed 20 June 2018].

Organisation for Economic Co-Operation and Development (OECD) (2015). *Data-Driven Innovation for Growth and Well-Being*. STI Policy Note. [pdf] Available at: http://www.oecd.org/sti/ieconomy/PolicyNote-DDI.pdf [Accessed 20 June 2018].

Pagallo, U. (2015). Good Onlife Governance: On Law, Spontaneuous Orders, and Design. In: L. Floridi, ed., *The Onlife Manifesto: Being Human in a Hyperconnected Era*, Cham: Springer, pp. 161-177

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Patel, N. (2014). Screwing with Your Emotions Is Facebook's Entire Business. *Vox*. [online] Available at: https://www.vox.com/2014/7/3/5865279/the-truth-about-facebook [Accessed 20 June 2018].

Pentland, A. (2014). *Social Physics: How Good Ideas Spread – The Lessons from a New Science*. New York: Penguin Press.

Pew Research Center (2014). *Public Perceptions of Privacy and Security in the Post-Snowden Era*. [pdf] Available at: http://www.pewinternet.org/files/2014/11/PI_PublicPerceptionsofPrivacy_111214.pdf [Accessed 20 June 2018].

Phillips Mandavill, A. (2014). The Revolution Is Not Everywhere Yet – and That's a Challenge for Global Development. *Millenium Challenge Corporation*. [online] Available at: https://www.mcc.gov/blog/entry/blog-011514-the-revolution-is [Accessed 20 June 2018].

Pigliucci, M. (2009). The End of Theory in Science? *EMBO Reports*, 10(6), pp. 534.

Porter, T. M. (2011). Statistics and the Career of Public Reason: Engagement and Detachment in a Quantified World. In: T. Crook and G. O'Hara, eds., *Statistics and the Public Sphere: Numbers and People in Modern Britain c. 1800-2000*. New York: Routledge, pp. 32-48.

Porter, T. M. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Quantacast (2013). *Know Ahead. Act Before*. [pdf] Available at: https://www.quantcast.com/wp-content/uploads/2013/11/Quantcast-Advertise-Mediakit_2013.pdf [Accessed 20 June 2018].

Richtel, M. (2013). How Big Data Is Playing Recruiter for Specialized Workers. *The New York Times*. [online] Available at: https://nyti.ms/2ktNd7I [Accessed 20 June 2018].

Rieder, G. and Simon, J. (2016). Datatrust: Or, the Political Quest for Numerical Evidence and the Epistemologies of Big Data. *Big Data and Society*, 3(1), pp. 1-6.

van Rijmenam, M. (2013). Why the 3V's Are Not Sufficient to Describe Big Data. *Datafloq*. [online] Available at: https://datafloq.com/read/3vs-sufficient-describe-big-data/166 [Accessed 20 June 2018].

Rosenblat, A., Kneese, T. and boyd, d. (2014). *Algorithmic Accountability*. A workshops primer Produced for: The Social, Cultural & Ethical Dimensions of "Big Data". [pdf] Available at: https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2543576_code1210838.pdf?abstractid=2535540&mirid=1 [Accessed 20 June 2018].

Rubinstein, I. (2013). Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law*, 3(2), pp. 74-87.

Rudder, C. (2014). *Dataclysm: Who We Are (When We Think No One's Looking)*. New York: Crown Publishers.

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal and Development*, 3(3), pp. 210-229.

Schaeffer, D. M. and Olson, P. (2014). Big Data Options for Small and Medium Enterprises. *Review of Business Information Systems*, 18(1), pp. 41-46.

Schroeder, R. (2014). Big Data and the Brave New World of Social Media Research. *Big Data & Society*, 1(2), pp. 1-11.

Science Europe (2014). How to Transform Big Data into Better Health: Envisioning a Health Big Data Ecosystem for Advancing Biomedical Research and Improving Health Outcomes in Europe. [pdf] Available at: https://www.scienceeurope.org/wp-content/uploads/2015/12/Workshop-Report_MED_Big_Data_web.pdf [Accessed 20 June 2018].

Silver, N. (2012). *The Signal and the Noise. Why So Many Predictions Fail – But Some Don't.* New York: The Penguin Press.

Simon, J. (2015). Distributed Epistemic Responsibility in a Hyperconnected Era. In: L. Floridi, ed., *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Cham: Springer, pp. 145-159.

Simon, P. (2013). *Too Big to Ignore: The Business Case for Big Data*. Hoboken, New Jersey: Wiley & Sons.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G. and Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, 25(4), pp. 359-393.

Steadman, I. (2013). Big Data and the Death of the Terrorist. *Wired*. [online] Available at: http://www.wired.co.uk/article/big-data-end-of-theory [Accessed 20 June 2018].

Strawn, G. O. (2012). Scientific Research: How Many Paradigms? *Educause Review*, 47(3), pp. 26-34.

Striphas, T. (2015). Algorithmic Culture. *European Journal of Cultural Studies*, 18(4-5), pp. 395-412.

Strong, C. (2015). *Humanizing Big Data: Marketing at the Meeting of Data, Social Science and Consumer Insight*. London: Kogan Page.

Szal, A. (2015). MIT System Successfully Processes Big Data Without Human Involvement. *Manufacturing.net*. [online] Available at: http://www.manufacturing.net/news/2015/10/mit-system-successfully-processes-big-data-without-human-involvement [Accessed 20 June 2018].

UN Women (2013). UN Women Ad Series Reveals Widespread Sexism. [online] Available at: http://www.unwomen.org/en/news/stories/2013/10/women-should-ads [Accessed 20 June 2018].

Valinsky, J. (2013). Japanese Court Orders Google to Pay Fine for Embarrassing Autocomplete Results. *Observer*. [online] Available at: http://observer.com/2013/04/japanese-court-search-ruling-google/ [Accessed 20 June 2018].

Venturini, T.; Laffite, N. B.; Cointet, J.-P.; Gray, I.; Zabban, V. and de Pryck, K. (2014). Three Maps and Three Misunderstandings: A Digital Mapping of Climate Diplomacy. *Big Data & Society*, 1(1), 1-19.

White House (2014a). *Big Data and Privacy: A Technological Perspective*. [pdf] Available at: https://bigdatawg.nist.gov/pdf/pcast_big_data_and_privacy_-_may_2014.pdf [Accessed 20 June 2018].

White House (2014b). *Big Data: Seizing Opportunities, Preserving Values*. [pdf] Available at: https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [Accessed 20 June 2018].

Wilbanks, J. (2009). I Have Seen the Paradigm Shift and It Is Us. In: T. Hey, S. Tansley and K. Tolle, eds., *The Fourth Paradigm. Data-Intensive Scientific Discovery*. Redmond: Microsoft Research, pp. 209-214.

Williamson, B. (2015). Smarter Learning Software: Education and the Big Data Imaginary. Conference Paper. [pdf] Available at: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2014/7-In-depth_review_of_big_data.pdf [Accessed 20 June 2018].

Williamson, B. (2014). Knowing Public Services: Cross-Sector Intermediaries and Algorithmic Governance in Public Sector Reform. *Public Policy and Administration*, 29(4), pp. 292-312.

Zarsky, T. (2016). The Trouble with Algorithmic Decisions: An Analytical Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science, Technology, & Human Values*, 41(1), pp. 118-132.

Paper III

Tracing Big Data Imaginaries through Public Policy: The Case of the European Commission

by Gernot Rieder

Presented Version:

Accepted, peer-reviewed manuscript corrected for typographical errors and citation style

Published Version:

Rieder, G. (2018). Tracing Big Data Imaginaries through Public Policy:
The Case of the European Commission.
In: A. R. Sætnan, I. Schneider and N. Green, eds., *The Politics and Policies of Big Data: Big Data, Big Brother?* New York; London: Routledge, pp. 89-109.

Tracing Big Data Imaginaries through Public Policy:
The Case of the European Commission

by Gernot Rieder

Abstract

Across the globe, the notion of Big Data has received much attention, not only in technology and business circles but also among political authorities. Public officials in Europe, the U.S., and beyond have formulated Big Data strategies that will steer I(C)T development towards certain goals and aspirations. Drawing on official European Commission documents and using the notion of sociotechnical imaginaries as a sensitising concept, this chapter investigates the values, beliefs, and interests that guide European policymakers' Big Data rhetoric, making the argument that while the Commission's embrace of a strong free-market position can be partly explained in terms of vexing economic, institutional, and epistemic challenges, its push for Big Data solutions threatens to undermine democratic rights and principles as well as efforts towards responsible research and innovation. The article concludes with recommendations for further research, emphasising the need for cross-disciplinary dialogue and scholarship.

1. Introduction

In recent years, the term Big Data has emerged as a major buzzword, widely used by both public and private actors. A precise definition, however, remains elusive, as various stakeholders have offered different views – pointing, for instance, to the volume, velocity, and variety of data produced (see Laney, 2012), new and improved ways to collect, store, process, and analyse those data (see Ward and Barker, 2013), or profound changes in how people think, work, and live (see Mayer-Schönberger and Cukier, 2013).

Others have been more reluctant to buy into the hype, arguing that the current excitement is driven by inflated expectations rather than actual shifts in operational reality.[1] But while claims that the "Big Data bubble" is bound to burst sooner rather than later have been around for years (see Franks, 2012), reports indicate that investments in Big Data

[1] This is perhaps best visualized in Gartner's (2013) Hype Cycle for Emerging Technologies, which shows Big Data right at the "Peak of Inflated Expectations", gradually making its way toward the "Trough of Disillusionment".

solutions have only been increasing, with decision makers considering the ability to extract value from data as critical to future success (see Columbus, 2015). Media and public interest, too, remains high, with the *New York Times* publishing 118 articles mentioning Big Data over the course of 2016[2] and a Google Trends analysis attesting to the continued popularity of the term in global search queries.[3] The reasons for the persistence of what has repeatedly been described as a "fad" destined to fade (see Woody, 2012) are arguably twofold: On the one hand, the ongoing computerisation of ever more areas of human life – from social interaction and commerce to health care, law enforcement, and education – has provided ample opportunity for Big Data small talk. The notion has therefore become a convenient umbrella term, broad enough to be applicable to almost anything technology-related, while imparting a sense of urgency and importance. Big Data's conceptual vagueness is thus very much part of the term's appeal, serving as a common point of reference in today's fast-changing digital environments. On the other hand, what may have started as a technical discussion[4] has since developed into a much more complex cultural phenomenon. Big Data is not just a fashionable catchphrase; it is a modern myth that has inspired an almost religious following (see boyd and Crawford, 2012). This mythology, as will be shown below, is structured around a logic of promise and obligation that deals in metaphors and visions, hopes, dreams, and ambitions. As a result, Big Data no longer functions as a mere shorthand for a set of computational problems and methods, but acts as a powerful rhetorical device designed to boost support and ensure public consent.

Many of these narratives originated in the marketing departments of hard- and software vendors (e.g., see IBM, 2012), only to be echoed in the reports of research and consultancy firms (e.g., see IDC et al., 2013). It was not long, however, until they found their way into political discourse: For several years now, policymakers on both sides of the Atlantic have stressed the great potential of Big Data, framing it as a potent antidote to a wide range of societal issues. The excitement and confidence expressed in numerous political speeches

[2] Articles have been identified and counted using the *New York Times*' on-site search service.

[3] Google's Trends graph measures interest over time by showing the number of searches for a particular term relative to the total number of searches performed on Google.

[4] One of Big Data's origin stories leads to a series of presentations in the middle and late 1990s by John R. Mashey, a former chief scientist at Silicon Graphics who pointed to the challenge of increasing data traffic for IT infrastructures, for instance in his talk "Big Data and the Next Wave of InfraStress" (Mashey, 1999).

and communications is considerable, rivalling even the grand claims of industry stakeholders. Officials' widespread endorsement of Big Data has translated into concrete funding actions: In 2012, the Obama Administration announced a $200 million Big Data Research and Development Initiative to "help solve some [of] the Nation's most pressing challenges" (OSTP, 2012: 1); in 2014, the European Commission (EC) launched a €2.5 billion public-private partnership in an effort to "master Big Data" and put the Union at the "forefront of the global data race" (EC, 2014b: 1). In view of such large-scale commitments, the (dis)qualification of Big Data as a short-lived buzz or fad seems both analytically wanting and conceptually inadequate, unable to account for the actual scale and longevity of the phenomenon.

Against this background, the chapter at hand aims to develop a better understanding of the rise of Big Data in public policy. It will do so in two important ways: First, it will examine key *sociotechnical imaginaries* as they manifest in official policy documents, with recent publications of the European Commission serving as the main empirical material. Second, the chapter seeks to contextualise these narratives and visions by adopting a broader analytical perspective, pointing to a number of economic, institutional, and epistemic challenges that contribute to the production and perpetuation of the Big Data imaginary. Based on these observations, the chapter develops a pointed critique of the Commission's Big Data rhetoric, arguing that it incorporates a set of values and beliefs that threaten to undermine key democratic principles as well as efforts towards responsible and sustainable ICT development.

2. Interlude: Sociotechnical Imaginaries

The concept of sociotechnical imaginaries, introduced in Jasanoff and Kim (2009) and further elaborated in Jasanoff and Kim (2015), has been influential within the field of Science and Technology Studies (STS). Loosely tied to earlier work on social imaginaries by Anderson (1983), Appadurai (1996), and Taylor (2004), but with particular emphasis on science and technology as key sites of modern imagination, it has inspired a growing number of case studies to look for and investigate "collectively held, institutionally stabilised, and publicly performed visions of desirable futures" (Jasanoff, 2015: 4). Crucially, such visions are considered to be both *situated*, meaning that they are culturally and temporally particular, embedded within specific socio-political environments, and *materially grounded* in the sense that they are co-produced within heterogeneous networks of both human and non-human

actors. The concept's focus, however, is not only on the *formation* of sociotechnical imaginaries but also on their *performative power*: Once imaginaries become widely accepted and used, they may shape trajectories of research and innovation, steering technological progress as well as public/private expenditure. Thus, while imaginative work can be understood as an important cultural practice that creates common narratives and enables shared interpretations of social reality, it can have serious *normative implications*: What starts as a description of potentially attainable futures may soon turn into a prescription of futures that ought to be attained (see Jasanoff and Kim, 2009). Rather than mere fantasy, sociotechnical imaginaries thus constitute a "crucial reservoir of power and action [that] lodges in the hearts and minds of human agents and institutions" (Jasanoff, 2015: 17).

How then can the concept of sociotechnical imaginaries contribute to the kind of qualitative policy analysis proposed in this chapter? There are a couple of points to be made here. First and foremost, the very invocation of the imagination as an object of study "rejects the idea of politics as consisting simply of purposive, rational action" (ibid.: 7). Instead, it emphasises that political agendas are driven by culturally-specific belief and value systems that produce different forms of techno-political order. The governance of nanotechnologies, for instance, has been shaped by contrasting visions and ideals in Germany and the United States (see Burri, 2015), and narratives of national nuclear identity differ greatly between atom-free Austria (see Felt, 2015) and radiating France (see Hecht, 2009 [1998]). In this sense, the notion of sociotechnical imaginaries invites a close reading of the various expectations and concerns, the diverse norms, mores, and ideologies that guide and inform the articulation of national policies. Second, and equally important, the concept encourages a critical examination of dominant imaginaries in the sense that it takes into consideration the topographies of power that help these imaginaries to form, stabilise, proliferate, and endure. Again, this is a hybrid process that involves not only humans and their respective aspirations, but also technical artefacts, institutions, industrial practices, and regulatory frameworks, to list but a few. Thus, in addition to its focus on values and beliefs, the concept calls for a broader assessment of techno-political regimes, their actors, structures, and embodiments, and how these participate in the cultivation and maintenance of specific collective imaginations. Third, and on a more methodological note, the concept explicitly refers to official documents – e.g. policy texts, political speeches, press releases – as materials providing "some of the most accessible and ubiquitous resources for analysing sociotechnical

imaginaries" (STS-RPSI, 2017). Studied carefully, such documents may reveal "recurrent themes or tropes in references to national and cultural practices", including "articulations of the public good, risk, and responsibility" (ibid.). This emphasis on language as a "crucially important medium for the construction of imaginaries" (Jasanoff and Kim, 2009: 122) makes the concept a suitable tool when seeking to identify and discuss prominent visions of techno-scientific futures in official policy discourse. Last but not least, the concept advocates comparative investigations as well as research on changes over time, suggesting that imaginations of particular sociotechnical innovations can vary between different nations or stakeholders, and that shared visions may change with shifting circumstances. Diverging opinions, contrasting ideas, and signs of controversy should therefore be regarded as pointers that may help to illuminate imaginaries together with their inscribed values, goals, and politics. Ultimately, it is through such comparisons that the distinctive features of prevalent imaginaries become apparent, while simultaneously providing a clearer sense of possible alternatives.

Drawing on these insights, this chapter uses the notion of sociotechnical imaginaries as a sensitising concept to study visions of Big Data as they manifest in the official policy discourse of the European Commission. The aim is to gain a better understanding of the hopes and fears that drive these narratives, the ways in which they are related to particular constructions of the public good, and the broader historical and cultural contexts in which they are embedded. The European Commission presents a particularly interesting case as the institution has become an important regulator of digital services, its policies and directives not only affecting the national legislations of EU member states, but also impacting data-related laws, rights, and markets around the globe (see Buttarelli, 2016). Tracing the Commission's Big Data imaginary through respective policy documents[5] thus provides insight into the values and aspirations of a key political actor whose normative power can be expected to play a formative role in the creation and design of our digital future.

[5] For this chapter's analysis, Google's Advanced Search feature was used to scrape the European Commission's official website, https://ec.europa.eu, for documents referring to Big Data. Using this term alone and in combination with other search queries (e.g. names of commissioners, specific topics), roughly 120 publications (incl. communications, speeches, individual web pages) were collected and subjected to qualitative analysis, providing a decidedly partial but nevertheless rich picture of the Commission's Big Data imaginary.

3. Tracing Big Data Imaginaries: The Case of the European Commission

In early 2014, the then Vice-President of the European Commission responsible for the Digital Agenda, Neelie Kroes, declared at the European Data Forum in Athens that about 200 years after railways started to connect people and energise the economy, society now faces a new industrial revolution: a digital one, with Big Data as its fuel. Her dream, Kroes continued, was that Europe would take full part in this revolution, with "European industry able to supply, European citizens and businesses able to benefit, [and] European governments able and willing to support" (Kroes, 2014b: 1). A year later, in a speech at Hannover Messe, Kroes' successor in office, Guenther Oettinger, presented a similar vision. Europe's future is digital, he stated, with the availability and use of Big Data crucial for maintaining the Union's competitiveness. In order to not fall behind and realise the potential of digital technologies, Europe would need to act fast, becoming the avant-garde of digital manufacturing (see Oettinger, 2015c).

Framings such as these, which emphasise Big Data's great economic potential and the necessity to exploit it, have for years dominated the Commission's Big Data narrative. At its core, this narrative is characterised by a twofold dynamic: On the one hand, the frequent use of buzz-laden metaphors imparts a sense of novelty and excitement. Going through official EC publications, one can, for instance, find Big Data referred to as the "new oil" (Kroes, 2013c: 2), the "motor and foundation of the future economy" (EC, 2014b: 1), a "key asset" (EC, 2017a), a "game-changer" (Šefčovič, 2016), some "magic material" (Kroes, 2013a: 2), a "goldmine" (EC, 2015b: 1), the "lifeblood of digital markets" (Oettinger, 2015a). In such cases, the language of the Commission and its representatives resembles that of a marketing campaign, mystifying the product while stressing its value and functional benefits. On the other hand, there is also a sense of urgency to capitalise on what is seen as a great chance and opportunity. In order not to miss out, and to secure Europe's digital future, quick and decisive action is said to be required. Failing to act is considered a major concern (see Kroes, 2013a), threatening to stifle innovation and the development of a flourishing EU data market (see EC, 2014d). Accordingly, the main task would be to tackle any obstacles, create the right environment, and "turn this asset into gold" (Kroes, 2014b: 3).

What emerges most clearly from the documents is the conceptualisation of Big Data as an economic imperative, the key to unlocking a bright and prosperous future. Big Data's

status is that of a raw material to be mined and exploited, a digital lubricant of growth and progress. While this master narrative already provides a rough account of the Commission's general position, a closer look reveals a somewhat more fine-grained pattern of claims and promises. More precisely, the EC's vision of Big Data appears to hinge upon three interrelated storylines: Big Data as the cornerstone of a thriving data-driven economy, Big Data as a way to transform and improve public services, and Big Data as a tool for evidence-informed policy and decision making. We shall briefly examine each of these in more detail.

Big Data as the cornerstone of a thriving data-driven economy: This vision holds that Big Data is bound to play an integral role in the development of a strong European data industry, with Big Data technology and services expected to grow at a compound annual growth rate of 40% (see EC, 2014d), saving manufacturers billions and boosting EU economic growth by an additional 1.9% by 2020 (see EC, 2016g). What is more, by increasing productivity and accelerating innovation, Big Data is expected to stimulate an opportunity-rich market environment, creating "hundreds of thousands of new jobs in the coming years" (EC, 2014b: 2), gradually replacing lower-skilled work with new and higher-quality occupations (see Oettinger, 2015c). From a global perspective, the availability and use of Big Data is said to be crucial for maintaining the EU's competitiveness (see ibid.); failure to harness this potential would mean becoming dependent on solutions from abroad (see Kroes, 2013a). The creation of a thriving data market is thus presented as a simple binary choice: "We can either be at the table – or on the menu" (Kroes, 2013b: 3). And since the Commission is adamant to position the European Union (EU) as a "digital world player" (Oettinger, 2015d) that takes "the global lead in data" (Kroes, 2013c: 3), its strong support of Big Data as one of "Europe's key economic assets" (EC, 2014c: 1) appears a foregone conclusion.

Big Data as a way to transform and improve public services: On the public sector side, the use of Big Data technologies is seen as way to increase both government efficiency and effectiveness, improving the quality of public services while reducing costs and administrative burden through new and optimised solutions (see EC, 2017c). Example applications include the development of interoperable healthcare platforms, better management of traffic flows and energy consumption, or the design and implementation of automated translation

systems (see EC, 2017d).[6] In this narrative, public sector information is framed as a goldmine that must be unlocked (see Kroes, 2013a), enabling a "smarter use of public money" (Kroes, 2014b: 1) together with more "personalized, user-friendly and innovative services" (EC, 2017c). In order to realise this vision, public data silos would have to be opened up, facilitating information sharing not only inside and between government agencies (improving efficiency) or between these agencies and ordinary citizens (improving transparency), but also between the public and the private sector. In the eyes of the Commission, the ability to reap Big Data benefits is thus tied to strong public-private partnerships (see EC, 2014b) and the re-use of public sector information (see EC, 2014a).

Big Data as a tool for evidence-informed policy and decision making: With regard to public policy, there is the expectation that Big Data will enable "policymakers to make informed and evidence-based decisions" (EC, 2017d), tackling societal challenges such as climate change (Vestager, 2016), unemployment (Thyssen, 2016), or migration (EC, 2016c) with high precision and accuracy. In essence, Big Data technologies are supposed to provide timely, actionable insights, allowing administrations to make the 'right' choices more quickly. Techniques such as performance monitoring, opinion mining, or policy modelling are considered promising avenues to more targeted, data-driven policy designs (see EC, 2015d). Rather than merely offering advice, however, Big Data technologies are meant to generate solutions (see EC, 2014d), testing strategies and simulating outcomes at a systemic level (see EC, 2016a). The need to predict and prevent figures prominently in this imaginary, marking a shift from reactive to more proactive modes of governance. From nowcasting[7] income distribution to forecasting crop yields, from "predicting the traffic to predicting the economy" (Kroes, 2013a: 2), there is hardly an area of public policy that is expected to remain untouched by advances in Big Data analytics.

Although clearly desired, according to the Commission, none of these outcomes are guaranteed. Instead, their realisation is said to depend upon certain prerequisites, including (I) a coherent data ecosystem that counters the paralysing fragmentation of the European market (see Kroes, 2013a), (II) a firm commitment to open data, facilitating the re-use of

[6] For a more comprehensive overview of potential application areas, see "Part III: Usage and Exploitation of Big Data" in Cavanillas et al. (2016).

[7] In essence, the practice of nowcasting involves computational methods to "predict the present" (Khan, 2012) and provide near-term forecasts, for example regarding weather conditions, market movements, or influenza activity.

public sector information (see EC, 2016d), (III) significant investment in education and training to ensure Europe's workforce adapts to the new digital environment (see Oettinger, 2015b), and (IV) the establishment of public-private partnerships that unite all relevant players and strengthen every part of the Big Data value chain (see EC, 2017b). To "get it right" would imply safeguarding a better economic future; to "get it wrong" would mean losing out and risking European competitiveness on the international market (see Kroes, 2013b: 2). With data governance believed to be at a crossroads, there is thus strong emphasis that the Union "can't miss out on that kind of growth opportunity" (Kroes, 2013c: 2). To turn away and 'not do' Big Data would constitute both a failure and a mistake since technology development is considered "the only sustainable long term response to secure our digital future" (Oettinger, 2015d). In the Commission's narrative, Big Data is thus viewed as a destiny without viable alternatives, a programmatic answer to the assumed truism that "tomorrow's world will be digital" (Kroes, 2013b: 3).

As might be expected, data-related worries and concerns only play a minor role in this imaginary. Rather, there appears to be agreement that one should not "[stop] the wonderful things, simply to prevent the not-so-wonderful" (ibid.: 2). If, however, potential risks and pitfalls are being addressed, these are, on the one hand, mostly thought to be limited to issues of privacy and security and, on the other hand, assumed to be amenable to technological fixes – that is, solutions that do not challenge the deployment of a technology or service as such, but only the particularities of its design, leaving little room for serious intervention. Legal regulation is also considered an option, but only to the extent that safeguards "[do] not come at the expense of innovation" (ibid.: 4). From the Commission's perspective, laws should be pragmatic and proportionate, with rules set up to "maximize the value and minimise the cost of data" (Kroes, 2013d: 2). Tools and laws together could then "empower people" (Kroes, 2014b: 2), giving users both control and responsibility over their data and making consent a "cornerstone of data protection" (Kroes, 2013b: 3). Ultimately, measures such as these are expected to boost public trust and confidence in digital services whose benefits are believed to outweigh any potential harms, with detrimental effects deemed the exception, not the rule (see Kroes, 2012).

To sum up, while public debate about the impacts of Big Data is polarized and critical (see Cukier, 2014), the European Commission upholds a decidedly positive vision, promoting the widespread use of modern data analytics as an opportunity not to be missed. In a digital

future, Europe is seen as having the choice between leading or following, and the Commission is clear about its ambition to "master Big Data" and put the Union "at the forefront of the global data race" (EC, 2014b: 1). Consequently, primary concerns include ways to strengthen the sector, accelerate innovation, and profit from an unleashed data-driven economy where "data protection cannot mean data protectionism" (Kroes, 2013b: 4). What emerges is an imaginary of Big Data for industrial growth and public sector transformation, spurred by aspirations of increased efficiency, reduced costs, and digitally aggregated value generation (see EC, 2015c). With the continued fragmentation of the European market considered a threat to this vision, the Commission keenly stresses the need for an open and coherent data ecosystem, and "every lawmaker, every public body, every vested interest who wants to push back [needs to be convinced] there's a better way of doing things" (Kroes, 2014a: 3). What is initially presented as a choice thus quickly turns into a normative imperative and obligation. As former Commissioner Kroes maintained: "'Data' isn't a four-letter word. […] It's something Europe needs to embrace" (ibid.: 3).

4. Discussion: Data Governance in Times of Crisis

The Commission's predominant focus on economic growth and organisational efficiency should not come as a surprise. Formally established in 1951 as the High Authority of the European Coal and Steel Community (ECSC) – and later joined by (in 1957) and merged (in 1965) with the executives of the European Economic Community (EEC) and the European Atomic Energy Community (EAEC) – the European Commission, throughout its institutional history, has sought to retain its legitimacy through policies aimed at contributing towards prosperity, competitiveness, and effective EU leadership (see Nugent and Rhinard, 2015; Cini, 1996). In times of social and economic unrest, pressures to strengthen the market and battle unemployment while cutting costs can be particularly pronounced, and there is no doubt that the Union has recently seen its fair share of turbulence, both financial (see EC, 2009) and otherwise (see EC, 2016e). Against such a troubled background, the promises of Big Data may seem particularly alluring: With average growth rates said to outpace not only the general economy but also the already healthy numbers of the traditional ICT market (see EC, 2014d), the data sector presents itself as an opportunity "Europe cannot afford to miss" (ibid.: 2). Moreover, by promising to disrupt other industries and change "the way we do business"

(Oettinger, 2015c: 2) – with expected cost savings of as much as €426 billion and an additional GDP increase of €206 billion until 2020 (see EC, 2016g) – the proclaimed data revolution may well appear as a panacea to an ailing European economy. It thus stands to reason that the Union's fragile economic situation has provided fertile ground for the promises of Big Data advocates, the current climate of excitement arguably a combined product of significant budgetary pressures and vested corporate interests.

However, it is not just about financials. As outlined in the previous section, there is considerable interest in data technologies for evidence-informed policy and decision making,[8] stretching from policy modelling and simulation to monitoring and analysis to enforcement and compliance, with the intention of improving "the effectiveness, efficiency, and quality [of] decisions in the public sector" (see EC, 2015d). But while the traditional view of the Commission as a technocratic body in constant need of expertise and information certainly holds some explanatory power (see Metz, 2015), the recent rise of 'datatrust' (Rieder and Simon, 2016) appears to be of a somewhat different nature. As Porter (1995: 146) argues, whereas technocracy presupposes relatively secure elites who would insist that "cultivated judgment is required to solve social problems", attempts to mechanise decision making usually occur under conditions of intense scrutiny, when pressured authorities seek to shield themselves against outside challenges by exchanging expert judgement for seemingly impersonal technical routines. Rather than the high point of technocratic ambition, the Commission's push for Big Data solutions is thus better understood as a contested political and administrative body's effort to maintain authority and legitimacy amidst strong opposition by both national political forces and a substantial proportion of the Union's population. With people's trust remaining low,[9] *trusted* analytical services are seen as a potential remedy, offering actionable insights surrounded by what boyd and Crawford (2012: 663) have referred to as an "aura of truth, objectivity, and accuracy". The "promise of mechanical neutrality" (Gillespie, 2014: 181) with its "glow of veracity" (Daston and Galison,

[8] For some background on the "ascendancy of evidence", see Solesbury (2002); for a recent state-of-the art account of "data for policy", see Poel et al. (2015).

[9] According to the Standard Eurobarometer 86 survey of autumn 2016 (EC ,2016f), only 36 percent of Europeans "tend to trust" the European Union, with 54 percent saying that they "tend not to trust". Compared to pre-crisis levels, this is a low value. In spring 2007, for instance, trust in the Union was at 57 percent (ibid.: 14).

1992: 111) thus serves as a valuable tool and means to establish credibility and justify political action.

Finally, and on a related note, while the availability of ever more data and the ability to collect and analyse them contribute to governmental efforts to "make society legible" (Scott, 1998: 2), Big Data's focus on machine learning and prediction caters to – but also fuels – policymakers' "craving for certainty" (Nowotny, 2016: 1) and "*ex ante* assurances" (ibid.: 7). In light of the expressed need "for organizing and managing […] increased complexity" (Šefčovič, 2016), data-driven insights are seen as a way to battle volatility and regain control, supporting the governance of the present by turning *unmeasurable uncertainty* into *measurable risk* (see Knight, 1921). Whereas the former threatens to grind the wheels of the political machinery to a halt, the latter provides a mandate for action based on probabilistic forecasts. Public officials' interest in Big Data is thus driven by a high demand for "anticipatory intelligence" (EC, 2015a: 18), which is thought to "augment decision makers" (EC, 2016b: 68) and "reconfigur[e] the policy system in a way that makes it more apt to address long-term challenges" (EC, 2015a: 19). In addition, the promise of universal applicability, most famously voiced by Anderson (2008), positions Big Data as a practical chance to expand the statistical colonisation of the future to an increasing number of governmental tasks and areas. To a bureaucratic institution that faces uncertainty but chases forward-looking policies, the prospect of 'knowing the odds' with unprecedented speed and accuracy may just constitute an offer too good to refuse. As a result, to date, the Commission's hopes are firmly pinned on Big Data.

In sum, the roots of the European Commission's Big Data imaginary can arguably be traced back to a specific amalgamation of economic, institutional, and epistemic factors and challenges: a Union in search of growth and prosperity amidst a severe financial crisis; a supranational authority experiencing a loss of trust and legitimacy; an agenda-setting political machinery in need of data and tools for probabilistic risk assessment. While a series of technological and methodological advancements have made Big Data possible, it is these contextual specificities that have elevated the term into a widespread sociotechnical phenomenon. The Commission's narrative, however, is not without its problems. In fact, it incorporates a number of claims and assumptions that warrant further critical examination. Even though a detailed discussion of the entrenched fallacies is beyond the scope of this chapter, a few key points should briefly be addressed.

Datatrust: As Bowker (2005) points out, data are never raw, but always already cooked, a process that needs to be recognised and considered with care. In a similar vein, Rieder (2017: 103) argues that the employment of algorithmic techniques may best be thought of as a situated epistemic practice that generates "interested readings of empirical, 'datafied' reality". The insight that even highly formalised analytical methods produce interpretations from specific 'somewheres' rather than a "view from nowhere" (Nagel, 1986) gains significance in a world where numbers and their mathematical treatment are increasingly considered the unquestioned – and unquestionable – backbone of decision making. Such *blind trust*, however, bears dangers of subverting the very foundations of deliberative democracy as well as key principles of the rule of law: With respect to the former, exaggerated faith in the neutrality and fairness of Big Data processes may undermine democratic deliberation by (a) deflecting from the goal-driven interests that shape and inform the design and implementation of analytical systems, (b) discouraging investigations into how Big Data practices can cause harm and unfair discrimination, and (c) rendering debates about alternative forms of business and governance both misguided and absurd. Regarding the latter, the sheer complexity of algorithmic operations can make the connection between input data and output decision-making murky (see Burrell, 2016), yielding an inscrutable evidence base that jeopardises existing due process norms, and disrespecting individuals' rights to transparent and accountable adjudication (see Citron, 2014; 2008). Given that Big Data's ultimate promise is not only to predict but to pre-empt, marking a fundamental shift from reactive to more "aggressive", proactive measures (see Kerr and Earle, 2013), blind trust in the veracity and social equity of mechanical reasoning appears a negative and irresponsible direction.

Trusted Technologies: Also related to the issue of trust, but approaching the topic from a different angle, are the Commission's attempts to strengthen individuals' trust and confidence in digital solutions, because "[w]ithout people's trust, a functioning [Digital Single Market] based on data will not work" (Ansip, 2015b). In light of a perceived "lack of trust in online services" (Ansip, 2015a), the Commission thus aims to reduce fears, mitigate concerns, and create "trusted" environments, encouraging citizens to "embrace the digital revolution" (Bieńkowska, 2016). What is problematic here is the rhetorical emphasis on *trust* rather than *trustworthiness*, on *trusted* rather than *trustworthy* computational systems. As Simon (2013)

stresses, "given the dangers of misplaced trust, […] we should not simply trust, but trust those who are worthy of our trust." From a governance perspective, attempts to "build", "boost", "secure", "restore", or "maintain" trust should always be accompanied – and ultimately surpassed – by efforts to create trustworthy frameworks and environments that contribute to citizens' social, economic, and legal protection. Put differently, if the political goal is to create an *atmosphere of trust* that "translate[s] into economic growth" and turns the Union into an "economic powerhouse" (Jourová, 2015: 5; see also Fukuyama, 1995), the democratically preferable and presumably more sustainable way would not be to directly target people's confidence level – e.g., through PR campaigns, shiny certificates, or repeated statements of intent – but to support and invest in systems that genuinely deserve such confidence. By strengthening citizens' digital rights, introducing new compliance obligations, and increasing monetary fines for violations, the EU's upcoming General Data Protection Regulation (GDPR), which will take effect in May 2018, arguably presents an important step in this direction. Yet it will be the actual implementation and enforcement of the Regulation by EU member states that will decide whether the longing for trust will eventually lead to the development and deployment of truly trustworthy digital solutions (see Davies, 2016).

Citizen Empowerment: One crucial way the Commission hopes to safeguard privacy and build trust in the digital economy is by giving people control over their data, conceptualising individuals as self-managing, entrepreneurial subjects who take charge and become makers of their own digital destiny. As Commissioner Kroes declared at the European Data Forum, "[w]hat you need is to empower people, give them control, give them a fair share of that value. Give them rights over their data – and responsibilities too, and the tools to exercise them" (Kroes, 2014b: 2). While such calls for the autonomous, self-reliant *homo datanomicus* may appeal to the ears of free-market liberals – and there are certainly good arguments for having users take part and profit from the monetisation of their personal data (e.g., see Rubinstein, 2013) – the idea that individuals should act as tech-savvy guardians of their own privacy has several caveats: First, as critique of the informed consent model has highlighted time and again, users tend not to read terms of services and privacy policies, thus calling into question the image of the attentive, dutifully engaged data citizen authorities seek to project (see Monteleone, 2015). Second, even if users were to devote time and effort into data privacy self-management, chances are that long and rhetorically cloudy customer

policies together with secretive corporate practices and opaque machine operations would render such attempts an almost impossible task, even for the digitally skilled (see Obar, 2015). Considering that many EU citizens still have no or very low digital competences (see Eurostat, 2013), the Commission's call for data privacy self-governance seems both unrealistic and, to a degree, insincere. Last but not least, the attribution of responsibility to the data subject involves a democratically dubious shifting of the burden to the individual who is (over)tasked with oversight in a context of high complexity and weak regulation. While the lengthy GDPR negotiations have shown how difficult it can be to strike a balance between industry and privacy interests,[10] a general redistribution of work and responsibility under the banner of citizen empowerment may prove incompatible with EU law, where the protection of personal data is recognised as a fundamental right (see European Communities, 2000).

Open Data: A final concern pertains to the Commission's strong support of open data, that is, "data made freely available for re-use to everyone for both commercial and non-commercial purposes" (EC, 2014d: 5). While for the Commission, the free flow of data "across sectors, languages, and borders" (ibid.: 6) constitutes a critical component in the establishment of a thriving, innovative digital economy, researchers have pointed to the dangers and limitations of this *sharing imperative*, arguing that the open distribution of public sector information may conflict with individuals' right to data privacy, underestimating the increasingly blurry lines between personal and non-personal data in the age of Big Data analytics (Kulk and van Loenen, 2012). As studies and real-world examples have shown, powerful computational models can now be used infer highly sensitive personal information from seemingly innocuous data (see Kosinski et al., 2013; Duhigg, 2012), and even when data have been anonymised, re-identification techniques may allow to link these data to specific individuals, undermining prior protection efforts (see Narayanan et al., 2008; Sweeney, 2000). Given this, but also in light of concerns regarding increased data marketisation (see Bates, 2012) and the establishment of open access research repositories (see Mauthner and Parry, 2013), the Commission's push for liberal data-sharing policies in the name of transparency, efficiency, and growth should be viewed critically and with caution. Instead of an "open by

[10] For a politically charged glimpse into the parliamentary negotiations around the EU data protection reform, see David Bernet's documentary *Democracy* (2015).

default" principle, as set out in the G8's Open Data Charter (see Group of Eight, 2013) and repeatedly referred to in Commission documents (e.g., see EC, 2014d), questions as to which data should be shared with whom, under what circumstances, and to what ends, ought to the be properly discussed and addressed. In the UK, public resistance against the government's care.data program has recently demonstrated that the distribution of personal-level information – even if anonymised – remains a highly controversial and divisive social issue (see Presser et al., 2015).

Though by no means comprehensive, the list above should provide a sense that the Commission's Big Data imaginary builds on a number of (epistemic) assumptions and (ideological) framings that, from an academic but also a purely democratic point of view, warrant critical scrutiny, if not outright rejection. To be sure, the problem is not so much that the Commission aims to create a strong and innovative digital market – after all, the fostering of trade and commerce has always been one of the institution's main responsibilities. Rather, it is that in this process, economic imperatives clearly outweigh wider social, ethical, and legal considerations, at times even to the detriment of fundamental rights and values. If the Commission truly wishes to (re)build trust as the basis of not only a thriving but a fair and sustainable digital economy, it should reconsider key elements of its Big Data narrative, sometimes even to the detriment of respective industry interests.

5. Conclusion

This article has sought to outline the European Commission's Big Data imaginary in some detail, reporting on the general thrust of the narrative, the claims and promises made, but also pointing to some of the major political challenges that have undoubtedly shaped the vision and its underlying values. What emerged from the assessed material was a strong free-market agenda in which trade barriers ought to be removed, regulatory burden ought to be reduced, and legislation ought to be tuned to industry requirements. In this normative framework, economic needs take precedence over questions of social desirability, and Big Data is mainly celebrated as the key to global competitiveness and a prosperous digital future.

But the strong economic imperative comes from a position of political weakness: Multiple crises and a deep sense of uncertainty have shaken the Union to its core, and the

Commission has operated under conditions of high pressure and distrust in what has been appropriately labelled "a challenging decade" (see EU, 2017). Against this background, the hype around Big Data technologies, which are expected to reduce costs, increase efficiency, generate value, and improve decision making, seems almost understandable. At the same time, however, the Commission's push for Big Data solutions involves certain concepts and convictions – from *blind trust* to *open by default* – that threaten to undermine basic democratic rights and principles. As an EU institution bound to uphold the Union's values and contribute to the protection of its citizens (see EU, 2007), the Commission should re-evaluate both the quality and societal plausibility of its Big Data discourse, pondering, for instance, whether the current narrative lives up to the standards of responsible research and innovation (RRI) laid out in the Horizon 2020 Framework Program.

In order to situate the findings of this study, it is important to note that the Commission's vision of and for Big Data is only one part of European-level data and ICT governance. The EU's new General Data Protection Regulation, for example, does not explicitly refer to Big Data (see Schneider, 2018; EU, 2016b), and neither does the equally new Directive, which lays down principles for the processing of personal data in the criminal justice system (see EU, 2016a). What does this imply? In essence, it indicates that in current EU policy-making, the Big Data discourse runs separate from parallel discussions about data regulation, a peculiar *state of detachment* that allows for the continued use of Big Data as a generic marketing term, unburdened by the challenges of creating fair and sustainable data policies. On a positive note, this means that there is in fact political awareness of the problems and issues associated with rapidly progressing computerization, even if such concerns are treated outside the Big Data rhetoric. In turn and somewhat less encouraging, however, this also entails that Big Data as an influential, value-laden catchphrase may remain politically unchallenged, fully co-opted by vendor interests and monetisation imperatives. Thus, while the Commission's Big Data narrative is heavily biased and skewed, its structural divorce from wider debates about the potential ramifications of large-scale datafication may allow this imbalance to persist.

From an academic and research perspective, the task is manifold: A first step would be to develop a better understanding of the Big Data imaginary itself, its origins, values, and inscribed politics. The present article has sought to contribute to this effort. A second step should then consider ways of improving the quality of the political discourse so that visions

of an EU-wide Digital Single Market do not merely reflect industry and administrative needs, but also remain open to questions of social and cultural desirability, including the option to halt or reject particular technological applications. The call here is for critical participation rather than distant antagonism, for constructive reform rather than dogmatic opposition. Last but not least, there is a rising need for cross-disciplinary efforts: scholarship at the intersection of IT, on the one hand, and ethics/STS/law, on the other, should be fostered to harness knowledge and expertise for sustainable development. European academia will have to rise to this challenge as high-quality interdisciplinary programs remain rare. Yet with policymakers positively enthralled by the promises of Big Data advocates, further research on potential pitfalls and consequences as well as on socially responsible ways forward seems absolutely paramount.

References

Anderson, B. (1983). *Imagined Communities: Reflections on the Origins and Spread of Nationalism*. London: Verso.

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. [online] Available at: https://www.wired.com/2008/06/pb-theory/ [Accessed 30 May 2017].

Ansip, A. (2015a). *Untitled Speech at the Digital Single Market Conference in Warsaw, Poland*. European Commission. Speech. [online] Available at: https://ec.europa.eu/commission/commissioners/node/190797_fr [Accessed 30 May 2017].

Ansip, A. (2015b). *Untitled Speech at the 3rd Annual Transatlantic Digital Economy Conference (AmCham EU)*. European Commission. Speech. [online] Available at: https://ec.europa.eu/commission/commissioners/node/404234_et [Accessed 30 May 2017].

Appadurai, A. (1996). *Modernity at Large*. Minneapolis: University of Minnesota Press.

Bates, J. (2012). "This Is What Modern Deregulation Looks Like": Co-optation and Contestation in the Shaping of the UK's Open Government Data Initiative. *The Journal of Community Informatics*, 8(2).

Bieńkowska, E. (2016). *Digital Transformation of the European Industries*. European Commission. Speech. [online] Available at: https://ec.europa.eu/commission/commissioners/2014-2019/bienkowska/announcements/digital-transformation-european-industries_en [Accessed 30 May 2017].

Bowker, G. C. (2005). *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.

boyd, d. and Crawford, K. (2012). Critical Questions for Big Data. Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication, & Society*, 15(5), pp. 662-679.

Burrell, J. (2016). How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society*, 3(1), pp. 1-12.

Burri, R. V. (2015). Imaginaries of Science and Society: Framing Nanotechnology Governance in Germany and the United States. In: S. Jasanoff and S.-H. Kim, eds., *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: The University of Chicago Press, pp. 233-253.

Buttarelli, G. (2016). The EU GDPR as a Clarion Call for a New Global Digital Gold Standard. *International Data Privacy Law*, 6(2), pp. 77–78.

Cavanillas, J. M., Curry, E. and Wahlster, W., eds., (2016). *New Horizons for a Data-Driven Economy. A Roadmap for Usage and Exploitation of Big Data in Europe*. Switzerland: Springer International Publishing.

Cini, M. (1996). *The European Commission. Leadership, Organisation and Culture in the EU Administration*. Manchester: Manchester University Press.

Citron, D. K. (2014). Big Data Should be Regulated by 'Technological Due Process'. *The New York Times*. [online] Available at: http://www.nytimes.com/roomfordebate/2014/08/06/is-big-data-spreading-inequality/big-data-should-be-regulated-by-technological-due-process [Accessed 30 April 2017].

Citron, D. K. (2008). Technological Due Process. *Washington University Law Review*, 85(6), pp. 1249-1313.

Columbus, L. (2015). 56% Of Enterprises Will Increase Their Investment in Big Data Over The Next Three Years. *Forbes*. [online] Available at: http://www.forbes.com/sites/louiscolumbus/2015/03/22/56-of-enterprises-will-increase-their-investment-in-big-data-over-the-next-three-years/#178b120388b1 [Accessed 7 May 2016].

Cukier, K. N. (2014). The Backlash Against Big Data. *The Economist*. [online] Available at: http://www.economist.com/blogs/economist-explains/2014/04/economist-explains-10 [Accessed 30 May 2017].

Daston, L. and Galison, P. (1992). The Image of Objectivity. *Representations*, 40, pp. 81-128.

Davies, S. (2016). The Data Protection Regulation: A Triumph of Pragmatism over Principle? *European Data Protection Law Review*, 2(3), pp. 290-296.

Democracy: Im Rausch der Daten. (2015). [film] Germany: David Bernet.

Duhigg, C. (2012). How Companies Learn Your Secrets. *The New York Times Magazine*. [online] Available at: http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html [Accessed 30 May 2017].

EC (European Commission) (2017a). *Big Data*. Website. [online] Available at: https://ec.europa.eu/digital-single-market/en/big-data [Accessed 30 May 2017].

EC (European Commission) (2017b). *Big Data Value Public-Private Partnership*. Website. [online] Available at: https://ec.europa.eu/digital-single-market/en/big-data-value-public-private-partnership [Accessed 30 May 2017].

EC (European Commission) (2017c). *Open Government*. Website. [online] Available at: https://ec.europa.eu/digital-single-market/en/open-government [Accessed 30 May 2017].

EC (European Commission) (2017d). *What Can Big Data Do for You?* Website. [online] Available at: https://ec.europa.eu/digital-single-market/en/what-big-data-can-do-you [Accessed 30 May 2017].

EC (European Commission) (2016a). *Big Data Supporting Public Health Policies*. Horizon 2020 Call. [online] Available at: http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/sc1-pm-18-2016.html [Accessed 30 May 2017].

EC (European Commission) (2016b). *Digital Futures – Final Report. A Journey into 2050 Visions and Policy Challenges*. [pdf] Available at:

https://ec.europa.eu/futurium/en/system/files/ged/futurium_scientific_report_v10revcl_v2_0.pdf [Accessed 30 May 2017].
EC (European Commission) (2016c). *Migration Data Catalogue and Dynamic Data Hub – EU Migration Data at a Glance*. Website. [online] Available at: https://ec.europa.eu/jrc/en/news/eu-migration-data-glance [Accessed 30 May 2017].
EC (European Commission) (2016d). *Open Data*. Website. [online] Available at: https://ec.europa.eu/digital-single-market/en/open-data [Accessed 30 May 2017].
EC (European Commission) (2016e). *Refugee Crisis in Europe*. Website. [online] Available at: http://ec.europa.eu/echo/refugee-crisis_en [Accessed 30 May 2017].
EC (European Commission) (2016f). *Standard Eurobarometer 86. Autumn 2016*. [online] Available at: https://ec.europa.eu/COMMFrontOffice/publicopinion/index.cfm/ResultDoc/download/DocumentKy/76422 [Accessed 30 May 2017].
EC (European Commission) (2016g). *The EU Data Protection Reform and Big Data Factsheet*. Website. [online] Available at: http://ec.europa.eu/justice/data-protection/files/data-protection-big-data_factsheet_web_en.pdf [Accessed 30 May 2017].
EC (European Commission) (2015a). *Tool #4: Evidence-Based Better Regulation*. [pdf] Available at: https://ec.europa.eu/info/sites/info/files/file_import/better-regulation-toolbox-4_en_0.pdf [Accessed 30 May 2017].
EC (European Commission) (2015b). *Digital Single Market Strategy: European Commission Agrees Areas for Action*. Press Release. [pdf] Available at: http://europa.eu/rapid/press-release_IP-15-4653_en.pdf [Accessed 30 May 2017].
EC (European Commission) (2015c). *Making Big Data work for Europe*. Website. [online] Available at: https://ec.europa.eu/digital-single-market/en/making-big-data-work-europe [Accessed 30 May 2017].
EC (European Commission) (2015d). *Policy-Development in the Age of Big Data: Data-Driven Policy-Making, Policy-Modelling, and Policy-Implementation*. Horizon 2020 Call. [online] Available at: https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/co-creation-06-2017.html [Accessed 30 May 2017].
EC (European Commission) (2014a). *Commission Encourages Re-Use of Public Sector Data*. Press Release. [online] Available at: http://europa.eu/rapid/press-release_IP-14-840_en.htm [Accessed 30 May 2017].
EC (European Commission) (2014b). *European Commission and Data Industry Launch €2.5 billion Partnership to Master Big Data*. Press Release. [online] Available at: http://europa.eu/rapid/press-release_IP-14-1129_en.pdf [Accessed 30 May 2017].
EC (European Commission) (2014c). *Frequently Asked Questions: Public-Private Partnership (PPP) for Big Data*. Memo. [pdf] Available at: http://europa.eu/rapid/press-release_MEMO-14-583_en.pdf [Accessed 30 May 2017].
EC (European Commission) (2014d). *Towards a Thriving Data-Driven Economy*. Communication. [pdf] Available at:

http://ec.europa.eu/newsroom/dae/document.cfm?action=display&doc_id=6210 [Accessed 30 May 2017].

EC (European Commission) (2009). *Economic Crisis in Europe: Causes, Consequences and Responses*. European Economy Series 7/2009. [pdf] Available at: http://ec.europa.eu/economy_finance/publications/pages/publication15887_en.pdf [Accessed 30 May 2017].

EU (European Union) (2017). *The History of the European Union*. Website. [online] Available at: https://europa.eu/european-union/about-eu/history_en [Accessed 30 May 2017].

EU (European Union) (2016a). *Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data by Competent Authorities for the Purposes of the Prevention, Investigation, Detection or Prosecution of Criminal Offences or the Execution of Criminal Penalties, and on the Free Movement of Such Data, and Repealing Council Framework Decision 2008/977/JHA*. Official Journal of the European Union, L 119/89. [online] Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016L0680&from=EN [Accessed 30 May 2017].

EU (European Union) (2016b). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)*. Official Journal of the European Union, L 119/1. [online] Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN [Accessed 30 May 2017].

EU (European Union) (2007). *Treaty of Lisbon Amending the Treaty on European Union and the Treaty Establishing the European Community*. Official Journal of the European Union, C 306/1. [online] Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12007L/TXT&from=en [Accessed 30 May 2017].

European Communities (2000). *Charter of Fundamental Rights of the European Union*. [pdf] Available at: http://www.europarl.europa.eu/charter/pdf/text_en.pdf [Accessed 30 May 2017].

Eurostat (2013). *Quality of Life in Europe - Facts and Views – Education*. Website. [online] Available at: http://ec.europa.eu/eurostat/statistics-explained/index.php?title=Quality_of_life_in_Europe_-_facts_and_views_-_education&oldid=238363 [Accessed 30 May 2017].

Felt, U. (2015). Keeping Technologies Out: Sociotechnical Imaginaries and the Formation of Austria's Technopolitical Identity. In: S. Jasanoff and S.-H. Kim, eds., *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: The University of Chicago Press, pp. 103-125.

Franks, B. (2012). Will the Big Data Bubble Burst in 2013? *International Institute for Analytics*. [online] Available at: http://iianalytics.com/research/will-the-big-data-bubble-burst-in-2013 [Accessed on 7 May 2016].

Fukuyama, F. (1995). *Trust. The Social Virtues and the Creation of Prosperity*. New York: Free Press Paperbacks.

Gartner (2013). *Gartner's 2013 Hype Cycle for Emerging Technologies Maps Out Evolving Relationship Between Humans and Machines*. [online] Available at: http://www.gartner.com/newsroom/id/2575515 [Accessed 7 May 2016].

Gillespie, T. (2014). The Relevance of Algorithms. In: T. Gillespie, P. J. Boczkowski and K. A. Foot, eds., *Media Technologies: Essays on Communication, Materiality, and Societ*y. Cambridge, MA: MIT Press, pp. 167-193.

Group of Eight (2013). *G8 Open Data Charter*. [online] Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/207772/Open_Data_Charter.pdf [Accessed 30 May 2017].

Hecht, G. (2009 [1998]). *The Radiance of France: Nuclear Power and National Identity after World War II*. Cambridge, MA: The MIT Press.

IBM (2012). *Business Value Accelerator for Big Data: Identify Opportunities for Enhanced Business Value from Big Data*. [online] Available at: http://www.ibm.com/services/de/gbs/pdf/bao-bigdata-GBS03157-USEN-00.pdf [Accessed on 7 May 2016].

IDC, EY, Tech4i2 and Europe Unlimited (2013). *Business Opportunities: Big Data*. [online] Available at: https://ec.europa.eu/growth/tools-databases/dem/sites/default/files/page-files/big_data_v1.1.pdf [Accessed on 7 May 2016].

Jasanoff, S. (2015). Future Imperfect: Science, Technology, and the Imaginations of Modernity. In: S. Jasanoff and S.-H. Kim, eds., *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: The University of Chicago Press, pp. 1-33.

Jasanoff, S. and Kim, S.-H., eds., (2015). *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: The University of Chicago Press.

Jasanoff, S. and Kim, S.-H. (2009). Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea. *Minerva*, 47: pp. 119-146.

Jourová, V. (2015). *EU Consumer Policy: Today's Challenges, Tomorrow's Solutions.* European Commission. Speech. [pdf] Available at: http://ec.europa.eu/justice/events/european-consumer-summit/2015/files/speech_commissiner_jourova__consumer_summit_2015_en.pdf [Accessed 30 May 2017].

Kerr, I. and Earle, J. (2013). Prediction, Preemption, Presumption. *Stanford Law Review Online*, 66, pp. 65-72.

Khan, I. (2012). Nowcasting: Big Data Predicts the Present. *ITworld*. [online] Available at: https://www.itworld.com/article/2719343/it-management/nowcasting--big-data-predicts-the-present.html [Accessed 30 May 2017].

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Boston, NY: Houghton Mifflin.

Kosinski, M., Stillwell, D. and Graepel, T. (2013). Private Traits and Attributes Are Predictable from Digital Records of Human Behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), pp. 5802-5805.

Kroes, N. (2014a). *Embracing the Open Opportunity*. European Commission. Speech/14/556. [pdf] Available at: http://europa.eu/rapid/press-release_SPEECH-14-556_en.pdf [Accessed 30 May 2017].

Kroes N. (2014b). *The Data Gold Rush*. European Commission. Speech/14/229. [pdf] Available at: http://europa.eu/rapid/press-release_SPEECH-14-229_en.pdf [Accessed 30 May 2017].

Kroes, N. (2013a). *Big Data for Europe*. European Commission. Speech/13/893. [pdf] Available at: http://europa.eu/rapid/press-release_SPEECH-13-893_en.pdf [Accessed 30 May 2017].

Kroes, N. (2013b). *Data Isn't a Four-Letter Word*. European Commission. Speech/13/1059. [pdf] Available at: http://europa.eu/rapid/press-release_SPEECH-13-1059_en.pdf [Accessed 30 May 2017].

Kroes, N. (2013c). *The Big Data Revolution*. European Commission. Speech/13/261. [pdf] Available at: http://europa.eu/rapid/press-release_SPEECH-13-261_en.pdf [Accessed 30 May 2017].

Kroes, N. (2013d). *The Economic and Social Benefits of Big Data*. European Commission. Speech/13/450. [pdf] Available at: http://europa.eu/rapid/press-release_SPEECH-13-450_en.pdf [Accessed 30 May 2017].

Kroes, N. (2012). *Opening Science through E-Infrastructures*. European Commission. Speech/12/258. [pdf] Available at: http://europa.eu/rapid/press-release_SPEECH-12-258_en.pdf [Accessed 30 May 2017].

Kulk, S. and van Loenen, B. (2012). Brave New Open Data World? *International Journal of Spatial Data Infrastructures Research*, 7, pp. 196-206.

Laney, D. (2012). Deja VVVu: Others Claiming Gartner's Construct for Big Data. *Gartner Blog Network*. [online] Available at: http://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/ [Accessed 7 May 2016].

Mashey, J. R. (1999). Big Data and the Next Wave of InfraStress. *1999 USENIX Annual Technical Conference*. Monterey, California. [online] Available at: http://static.usenix.org/event/usenix99/invited_talks/mashey.pdf [Accessed 30 May 2017].

Mauthner, N. S. and Parry, O. (2013). Open Access Digital Data Sharing: Principles, Policies, and Practices. *Social Epistemology: A Journal of Knowledge, Culture and Policy*, 27(1), pp. 47-67.

Mayer-Schönberger, V. and Cukier, K. (2013). *Big Data. A Revolution That Will Transform How We Live, Work, And Think*. New York: Houghton Mifflin Harcourt.

Metz, J. (2015). *The European Commission, Expert Groups, and the Policy Process. Demystifying Technocratic Governance*. European Administrative Governance Series. Basingstoke: Palgrave Macmillan.

Monteleone, S. (2015). Addressing the 'Failure' of Informed Consent in Online Data Protection: Learning the Lessons from Behaviour-Aware Regulation. *Syracuse Journal of International Law and Commerce*, 43(1), pp. 69-119.

Nagel, T. (1986). *The View from Nowhere*. New York: Oxford University Press.

Narayanan, A. and Shmatikov, V. (2008). Robust De-Anonymization of Large Sparse Datasets. *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, pp. 111-125.

Nowotny, H. (2016). *The Cunning of Uncertainty*. Cambridge: Polity Press.

Nugent, N. and Rhinard, M. (2015 [2001]) *The European Commission*. London: Palgrave.

Obar, J. A. (2015). Big Data and The Phantom Public: Walter Lippmann and the Fallacy of Data Privacy Self-Management. *Big Data & Society*, 2(2), pp. 1-16.

Oettinger, G. (2015a). *A Digital Single Market: The Key to Europe's Industrial Leadership in the Digital Economy*. European Commission. Speech. [online] Available at: https://ec.europa.eu/commission/commissioners/2014-2019/oettinger/announcements/speech-ict2015-digital-single-market-key-europes-industrial-leadership-digital-economy_en [Accessed 30 May 2017].

Oettinger, G. (2015b) *CeBit & IT in Companies*. European Commission. Blog Post. [online] Available at: https://ec.europa.eu/commission/commissioners/2014-2019/oettinger/blog/cebit-it-companies_en [Accessed 30 May 2017].

Oettinger, G. (2015c). *Europe's Future is Digital*. European Commission. Speech. [pdf] Available at: http://europa.eu/rapid/press-release_SPEECH-15-4772_en.pdf [Accessed 30 May 2017].

Oettinger, G. (2015d). *Innovation to Secure Digital Future*. European Commission. Blog Post. [online] Available at: https://ec.europa.eu/commission/commissioners/2014-2019/oettinger/blog/innovation-secure-digital-future_en [Accessed 30 May 2017].

OSTP (Office of Science and Technology Policy) (2012). *Obama Administration Unveils 'Big Data' Initiative: Announces $200 Million in New R&D Investments*. Press Release. [online] Available at: https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf [Accessed 7 May 2016].

Poel, M., Schroeder, R., Treperman, J., Rubinstein, M., Meyer, E., Mahieu, B., Scholten, Chiel and Svetachova, M. (2015). *Data for Policy: A Study of Big Data and Other Innovative Data-Driven Approaches for Evidence-Informed Policymaking*. Draft Report about the State-Of-The-Art: Invitation for Reflection. [online] Available at: https://ofti.org/wp-content/uploads/2015/05/dataforpolicy.pdf [Accessed 30 May 2017].

Porter, T. M. (1995). *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Presser, L., Hruskova, M., Rowbottom, H. and Kancir, J. (2015). Care.data and Access to UK Health Records: Patient Privacy and Public Trust. *Technology Science*. [online] Available at: https://techscience.org/a/2015081103/ [Accessed 30 May 2017].

Rieder, B. (2017). Scrutinizing the Algorithmic Technique: The Bayes Classifier as Interested Reading of Reality. *Information, Communication & Society*, 20(1), pp. 100-117.

Rieder, G. and Simon, J. (2016). Datatrust: Or, the Political Quest for Numerical Evidence and the Epistemologies of Big Data. *Big Data & Soci*ety, 3(1), pp. 1-6.

Rubinstein, I. S. (2013). Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law*, 3(2), pp. 74-87.

Schneider, I. (2018). Bringing the State Back in: Big Data-based Capitalism, Disruption, and Novel Regulatory Approaches in Europe. In: A.R. Sætnan, I. Schneider and N. Green, eds., *The Politics and Policies of Big Data: Big Data, Big Brother?* New York; London: Routledge, pp. 129-175.

Scott, J. C. (1998). *Seeing Like a State. How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, CT: Yale University Press.

Šefčovič, M. (2016). Untitled Speech at the European Political Strategy Centre (EPSC) Seminar on "The Energy Union and Climate Change Policy – A Transformative Agenda for the Modernisation of the Economy". [online] Available at: https://ec.europa.eu/commission/commissioners/2014-2019/sefcovic/announcements/speech-european-political-strategy-centre-epsc-seminar-energy-union-and-climate-change-policy_en [Accessed 30 May 2017].

Simon, Judith (2013). Trust. In: P. Duncan, ed., *Oxford Bibliographies in Philosophy*. New York: Oxford University Press.

Solesbury, W. (2002). The Ascendancy of Evidence. *Planning Theory & Practice,* 3(1), pp. 90-96.

STS-RPSI (STS Research Platform Sociotechnical Imaginaries) (2017). *Methodological Pointers*. [online] Available at: http://sts.hks.harvard.edu/research/platforms/imaginaries/ii.methods/methodological-pointers/ [Accessed 7 May 2017].

Sweeney, L. (2000). *Simple Demographics Often Identify People Uniquely*. Carnegie Mellon University, Data Privacy Working Paper 3. [online] Available at: https://dataprivacylab.org/projects/identifiability/paper1.pdf [Accessed 30 May 2017].

Taylor, C. (2004). *Modern Social Imaginaries*. Durham: Duke University Press.

Thyssen, M. (2016). *Speaking Points on Long Term Unemployment and Four Dimensions of Work Quality at OECD Employment and Labour Ministerial Meeting*. European Commission. Speech. [online] Available at: https://ec.europa.eu/commission/commissioners/2014-2019/thyssen/announcements/speaking-points-commissioner-marianne-thyssen-long-term-unemployment-and-four-dimensions-work_en [Accessed 30 May 2017].

Vestager, M. (2016). *Big Data and Competition*. European Commission. Speech. [online] Available at: https://ec.europa.eu/commission/commissioners/2014-2019/vestager/announcements/big-data-and-competition_en [Accessed 30 May 2017].

Ward, J. S. and Barker, A. (2013). Undefined by Data: A Survey of Big Data Definitions. *arXiv:1309.5821*. [online] Available at: http://arxiv.org/abs/1309.5821 [Accessed 7 May 2016].

Woody, B. (2012). Big Data is Just a Fad. *Simple Talk*. [online] Available at: https://www.simple-talk.com/sql/database-administration/big-data-is-just-a-fad/ [Accessed on 7 May 2016].

Paper IV

Big Data and Technology Assessment: Research Topic or Competitor?

by Gernot Rieder and Judith Simon

Presented Version:

Accepted, peer-reviewed manuscript corrected for typographical errors and citation style

Published Version:

Rieder, G. and Simon, J. (2017). Big Data and Technology Assessment: Research Topic or Competitor? *Journal of Responsible Innovation*, 4(2), pp. 234-253.

Big Data and Technology Assessment: Research Topic or Competitor?

by Gernot Rieder and Judith Simon

Abstract

With its promise to transform how we live, work, and think, Big Data has captured the imaginations of governments, businesses, and academia. However, the grand claims of Big Data advocates have been accompanied with concerns about potential detrimental implications for civil rights and liberties, leading to a climate of clash and mutual distrust between different stakeholders. Throughout the years, the interdisciplinary field of technology assessment (TA) has gained considerable experience in studying socio-technical controversies and as such is exceptionally well equipped to assess the premises and implications of Big Data practices. However, the relationship between Big Data as a socio-technical phenomenon and TA as a discipline assessing such phenomena is a peculiar one: Big Data may be the first topic TA deals with that is not only an object of inquiry, but also a major competitor, rivaling TA in several of its core functions, including the assessment of public views and visions, means and methods for exploring the future, and the provision of actionable knowledge and advice for political decision making. Our paper explores this dual relationship between Big Data and TA before concluding with some considerations on how TA might contribute to more responsible data-based research and innovation.

Keywords

Big Data, technology assessment, responsible research and innovation, interdisciplinarity

1. Introduction

Around the globe, the notion of Big Data[1] has captured the imaginations of governments, businesses, and academics. Heralded as a key enabler of public sector innovation (see

[1] Although many definitions have been proposed (see Press, 2014), there is "a pronounced lack of consensus about the definition, scope, and character of what falls within the purview of Big Data" (Ekbia et al., 2015: 3). One of the most popular characterizations is Laney's (2012, 2001) notion of the

European Commission (EC), 2015b), a catalyst for economic growth and well-being (see Organisation for Economic Co-operation and Development, 2015), and the emblem of a new "data-intensive" scientific paradigm (see Hey et al., 2009), there is hardly a segment of modern society that is not expected to be touched and transformed by the ongoing "Big Data revolution" (Mayer-Schönberger and Cukier, 2013). While wrapped in a rhetoric of hype and hope, applications of Big Data are no longer science fiction: From crime and disaster prediction to online advertising, from precision medicine and disease tracking to industry 4.0, from smart cities and climate research to credit and insurance scoring, the use of computational means to uncover patterns and trends in ever larger haystacks of data has found widespread appeal. Significant investments are being made, underpinned by promissory narratives of efficiency and security, progress and prosperity. To "unlock the value" and "reap [the] benefits" (EC, 2016a) of Big Data appears to have emerged as a primary concern of both public and private entities, the incorporation of advanced analytics into virtually all areas of human life already considered a foregone conclusion.

But the rise and spread of Big Data solutionism (see Morozov, 2013b) has not remained unchallenged. Observers in the media and academia, but also from watchdog organizations and public interest groups, have called for open debate and critical reflection, pointing to unresolved issues related to privacy and surveillance, bias and discrimination. Put poignantly: What may have started as an advertising campaign for the new "testosterone of business computing" (Vance, 2010) soon turned into a heated argument about civil rights and liberties (see Upturn, 2014). Though large-scale outcry has been the exception rather than the rule,[2] public reactions to the Snowden revelations (see Lyon, 2014), NHS England's

"3Vs", which focuses on *measures of magnitude* and conceptualizes Big Data as growth in data volume, velocity, and variety. Other approaches have shifted the focus from data properties to new *analytical possibilities*, describing Big Data science as a "God's-eye view" (Pentland, 2012) that "lets us examine society in fine-grained detail" (Pentland, 2014: 177). In contrast to such technology-oriented perspectives, scholars from the social sciences and humanities have pointed to the *cultural dimension* of Big Data, arguing that the real novelty of Big Data lies in the growing significance and authority of quantified information in ever more areas of everyday life (see Leonelli, 2014). From this perspective, Big Data constitutes a complex socio-technical phenomenon that rests on an interplay of science, technology, ideology, and mythology (see Jurgenson, 2014; boyd and Crawford, 2012). It is this latter perspective that will guide our analysis.

[2] Citizens' passivity may have multiple causes. A survey by Turow, Hennessy, and Draper (2015) on consumer data collection in both digital and physical commerce, for instance, finds that people's provision of personal information is not the result of either consent, ignorance, or indifference, but

care.data initiative (see Presser et al., 2015), or Facebook's emotional contagion experiment (see boyd, 2016) illustrate a growing discomfort with current data practices. The result is a climate of clash, an atmosphere of intense distrust between different parties and stakeholders with conflicting interests, values, and diverging visions of the future.

The situation is unlikely to resolve itself quickly: As rapid growth in both data generation (see IDC, 2014) and the analytics market (see Statista, 2016) suggests, at least from a techno-economic perspective, the age of Big Data has only just begun. Processes of datafication and computerization are sure to continue, and it is safe to assume that the combination of declining hardware costs, rising processing power, and ever more sophisticated software solutions will increase the allure of Big Data's 'capture all' imperative. With even more organizations planning to jump on the data train (see Gartner, 2015), and many people experiencing a loss of control over their personal information (see EC, 2015c; PEW, 2014), further conflict seems inevitable.

Throughout the years, the interdisciplinary field of technology assessment (TA) has gained considerable experience in studying socio-technical controversies. Extensive research on issues ranging from nuclear power and waste management to genetically modified organisms, geo engineering, stem cell research, and nanotechnology has left the discipline with a broad set of methods and techniques to assess and evaluate the ethical, legal, and social implications of new and emerging technologies. Thereby, the identification of current and future challenges, the facilitation of multi-actor involvement, and the search for both desirable and sustainable solutions has often been of central concern to the field and its scholars.[3] In addition, TA as a concept and practice aims to contribute to the governance of science and technology by (a) adopting an intermediary role and fostering dialogue between policy makers, industry, and the public sphere (see Joss and Bellucci, 2002) and (b) providing actionable knowledge and advice for democratic decision making in cases where the stakes are high, facts are uncertain, and values are in dispute (see Klüver et al., 2015; Funtowicz and Ravetz, 1993). Big Data has only recently appeared on the TA agenda,[4] but if the "march of

rather a sense of resignation and powerlessness, a feeling that it is futile to even try to manage and control what companies can learn about them.

[3] For an overview of the TA landscape and its various strands, see van Est and Brom (2012) and Grunwald (2009).

[4] While explicit references to Big Data were rare in the program of the 1st European TA Conference in Prague in 2013 (see PACITA, 2013; Michalek et al., 2014), two years later, at the 2nd European TA

quantification" (Gary King, quoted in Lohr, 2012) continues, it will probably stay there for a while.

In this article, we seek to explore the relationship between Big Data and TA from two distinct perspectives:

First, we set out to discuss whether and in what ways TA can contribute to the current debate around Big Data. In essence, it is argued that the field's experience in bridging disciplinary boundaries, its proficiency in facilitating (upstream) public engagement, and its expertise in developing deliberative methods for thinking about possible future scenarios may indeed prove a valuable addition to Big Data discourse, providing not only insight into but potentially also a way out of the current climate of clash.

Second, instead of merely conceiving Big Data as a new research topic for TA, we shall consider their relationship as one marked by rivalry and competition. Despite significant epistemic and methodological differences, Big Data's key promise bears striking similarities to that of TA, namely the provision of actionable, future-oriented knowledge. Consequently, the nascent field of Big Data analytics – home to a growing number of software solutions marketed by major IT companies – may soon challenge TA in one of the discipline's core roles and functions: as a scientific advisor to political and bureaucratic decision making. Ultimately, this rivalry could lead to gradual displacement, especially if the computational approach appears to outperform its competitor both practically (e.g., cheaper, faster) and epistemologically (i.e., recommendations believed to be objective and based on numerical facts). The potential consequences of such a shift will be discussed at length.

In a concluding section, we wish to go beyond this scenario of competition and replacement and instead envision possibilities for cooperation and mutual learning between TA and Big Data analytics. Embedding our considerations within the context of RRI, we will inquire how responsibility in data-based research and innovation may be achieved and ponder how more reflexive, inclusive, and participatory modes of computational knowledge generation could actually be put into practice. In particular, we will stress the need for a multidisciplinary research approach, call attention to the politics of participation and the

Conference in Berlin, the term had become more common and a dedicated session sought to investigate the "Governance of Big Data and the Role of TA" (PACITA, 2015).

performativity of temporalities, and comment on the chances and pitfalls of collaborative knowledge production.

2. Big Data as a Research Topic for Technology Assessment

Though still relatively new as a topic of investigation, with significant growth in scholarly publications from 2012 onwards (see Singh et al., 2015; Youtie et al., 2017), several TA-related research initiatives examining the rise and impact of Big Data analytics have already been launched. European examples include the Germany-based ABIDA (Assessing Big Data) project,[5] the Norwegian Board of Technology's study on data-driven analysis and predictive policing (Teknologirådet, 2015), the UK Parliamentary Office of Science and Technology's exploration of Big Data uses across various policy areas,[6] or background documents by the European Commission's Unit for eHealth and Health Technology Assessment on Big Data in the medical sector (see EC, 2014). In the United States, the White House report *Big Data: Seizing Opportunities, Preserving Values* (White House, 2014) and the complementary report *Big Data and Privacy: A Technological Perspective* (President's Council of Advisors on Science and Technology, 2014) are examples of high-profile technology assessments meant to inform and steer federal S&T policy.[7]

While the initiatives listed above differ in scale and scope, they share the common goal of examining the potential impacts of Big Data from a decidedly *multidisciplinary perspective*. The German ABIDA project, for instance, includes five specialized working groups who are tasked with assessing the opportunities and challenges of Big Data from either an ethical, legal, sociological, economical, or political science point of view.[8] Such multidisciplinary, which has been an integral part of TA programs for decades,[9] can contribute

[5] See the ABIDA website: http://www.abida.de/en [Accessed 20 June. 2018].
[6] See POST's Big Data program website: http://www.parliament.uk/mps-lords-and-offices/offices/bicameral/post/work-programme/big-data/ [Accessed 20 June 2018].
[7] Another example is the recent report *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* (White House, 2016).
[8] For a concise overview of the ABIDA project, see: http://www.abida.de/en/content/abida-das-projekt [Accessed 20 June. 2018].
[9] To give but one example, when reporting on the status of the now abolished US Office of Technology Assessment (OTA) back in the early 1980s, Project Director and Senior Analyst Fred B. Wood writes that "OTA's multidisciplinary staff [...] of 80-90 professionals spans the spectrum of physical, life, and social sciences, engineering, law, and medicine." (Wood, 1982: 214)

to Big Data discourse in two important ways: On the one hand, by bringing together expertise and insights from different fields, TA may provide a synoptic overview of what is often scattered across various disciplinary boundaries. Presented in a concise manner, this collected information may then allow TA scholars to act as "knowledge brokers" (Meyer, 2010) between the scientific and the political realm, thus strengthening what to date remains a notoriously difficult relationship (see Wilsdon et al., 2015). On the other hand, the influx of knowledge and know-how from different disciplines may provide a better understanding of the actual significance of Big Data as a complex socio-technical phenomenon. In fact, as review articles such as Ekbia et al. (2015) demonstrate, ethical and legal reflections alone are insufficient to cover the range of conceptual and practical dilemmas surrounding Big Data. And if it is true that modern data analytics will not only "transform how we live, work, and think" (Mayer-Schönberger and Cukier, 2013), but also how we know (see Kitchin, 2014), judge (see Christin et al., 2015), and govern (see Rieder and Simon, 2016), a concerted scholarly effort seems, indeed, indispensable.

Multidisciplinarity, however, is usually only the first step; the establishment of *interdisciplinary dialogue and collaboration* being the next. Perceived as a chance to transcend research silos and facilitate "more radical interactions between different styles of knowledge" (Stirling, 2014), interdisciplinary assessments are expected to provide responses to problems that "are not solvable by an individual scientific discipline alone" (Decker, 2001). Such '"wicked" (Rittel and Weber, 1973) or "post-normal" (Funtowicz and Ravetz, 1993) problem situations have become more frequent in the current "age of uncertainty" (Nowotny et al., 2001), and Big Data is no exception: Tamed by neither existing law (see Barocas and Selbst, 2016) nor new regulatory approaches (see Rubinstein, 2013), the search for hidden patterns and trends in ever larger – and increasingly diverse – datasets poses a host of intricate ethical and epistemic, social and political, legal, technical, and commercial challenges that evade traditional problem-solving strategies. Though not a panacea, issue-focused interdisciplinary research, as included in many TA programs (e.g., see Decker and Grunwald, 2001), can help in finding options for political action, providing practical guidance for problems that do not fit into the functional differentiation of academic disciplines.[10] Yet successful collaboration can be difficult to achieve, requiring significant effort, competence,

[10] For more on TA's "problem-oriented" version of interdisciplinarity, see Schmidt (2008).

a certain openness, and time. Especially in Europe, where higher education and research continue to be dominated by scholarly compartmentalization,[11] TA might play a crucial role in facilitating such interactions, nourishing a culture of interdisciplinarity ready to support the governance of new and emerging technologies, including Big Data.

While multi- and interdisciplinarity are key constituents, TA projects, including several of the Big Data-oriented initiatives mentioned above, often seek to take an additional step: the implementation of *transdisciplinary engagement*, meaning the active involvement of actors (e.g., laypeople, specialists, interest groups) from diverse social and professional backgrounds, in an effort to broaden the scope, gain new perspectives, and make the borders between science, technology, and society more permeable.[12] Depending on the stage of a techno-scientific development, such participatory approaches may serve two main purposes.[13] On the one hand, in the early stages of development, when a technology is new and societal consequences are difficult to foresee, public engagement – for example, through scenario exercises (see Selin, 2011), group discussions (see Felt et al., 2014), or online deliberation tools (see Rommetveit et al., 2013) – can assist in generating anticipatory knowledge about possible future trajectories and their implications, determining both the plausibility and desirability of an emerging socio-technical arrangement. The focus here is on preparation rather than prediction, on cultivating a capacity to identify viable options and alternatives in times of growing complexity and uncertainty (see Barben et al., 2008). On the other hand, however, when a controversial technology has already become entrenched and ex ante preparation is no longer possible, the involvement of heterogeneous groups of actors can provide a better understanding of the nature of the conflict, that is, the main concerns, the values at stake, the positions taken, the interests involved. Moreover, the inclusion of

[11] Regarding such compartmentalization in higher education, see Newell (2010); regarding research, see Pan, Boucherie, and Hanafi (2015).

[12] For a deeper, historically grounded discussion of such participatory technology assessment (pTA), see Joss and Bellucci (2002).

[13] In fact, reasons for public participation are manifold (see Wesselink et al., 2011). The argument proposed in this paragraph refers to David Collingridge's well-known "dilemma of control". Collingridge (1980: 11) states: "The social consequences of a technology cannot be predicted early in the life of the technology. By the time undesirable consequences are discovered, however, the technology is often so much part of the whole economics and social fabric that its control is extremely difficult." Like Collingridge, TA searches for ways and means to deal with and, both in theory and practice, overcome this quandary. For an insightful analytical discussion of the dilemma, its assumptions and relationship to TA, see Liebert and Schmidt (2010).

affected publics can reveal information about problems and issues that may otherwise be overlooked (see Cotton, 2014), offering a clearer picture not only of possible, but of actual harms and risks. In sum, participatory engagement can be considered a vital element for a more "anticipatory" (Guston, 2014) and "reflexive" (Braun et al., 2010) governance of science and technology, opening both existing conditions and future prospects to broader scrutiny and critical debate. In the case of Big Data, where impacts are already tangible (see O'Neil, 2016) but even bigger changes are on the way (see Davenport, 2015), such expanded modes of inquiry and reflection could prove essential for more sustainable, socially robust development.

All in all, we believe that TA's experience in multi-, inter-, and transdisciplinary research, its practical expertise in consultation, deliberation, and advice at the science-society-policy interface, can help to address and successfully deal with the manifold challenges posed by Big Data. In turn, the ongoing controversy about Big Data provides an opportunity for TA to prove itself as a theory and practice, demonstrating its value and relevance for the democratic governance of techno-scientific innovation and change.

Modern data analytics, however, are not only destined to become an important research topic for TA, they may also emerge as a serious competitor to TA, rivaling the field in several of its core competencies, including the assessment of public views and visions (i), means and methods for exploring the future (ii), and the provision of actionable knowledge and advice for political decision making (iii). Following, we shall elaborate on both the programmatic and epistemological similarities and differences between Big Data and TA, highlighting conceptual incommensurabilities as well as the potential for methodological complementarity and research synergy.

3. Big Data as a Competition for Technology Assessment

3.1. The assessment of public views and visions

While TA was initially conceived as a "rational-scientific tool" (Thompson Klein, 2001: 37) that would provide policy makers with "competent, unbiased information" concerning "probable

impacts of technology",[14] the field's focus has since shifted from mere risk-based assessments to greater consideration of public acceptance (see Assefa and Frostell, 2007) and social desirability (see Bennett and Sarewitz, 2006). Supported by a broad variety of survey and engagement methods,[15] numerous research projects have sought to investigate people's values and beliefs, but also their hopes and concerns regarding specific techno-scientific developments. However, there are certain problems: Quantitative survey research, for instance, has been criticized for relying on narrow 'tick-box' questionnaires that fail to account for the plurality and complexity of laypeople's thinking (see Macnaghten et al., 2010) and for the particular 'versions of reality' such surveys enact (see Law, 2009). Qualitative engagement exercises, in comparison, have come under fire for being slow and time-consuming,[16] for granting too much authority to the new "experts of community" (Rose, 1999) and their "technologies of participation" (Chilvers and Kearnes, 2016), and for issues of legitimacy and representativeness (see Lafont, 2015).

In view of such criticism, the growing interest in digital methods for controversy mapping, opinion mining, and sentiment analysis should not come as a surprise.[17] Paired with the epistemic promises of Big Data (see Rieder and Simon, 2017; Kitchin, 2014) such computational techniques may indeed seem like an offer too good to refuse: Advertised as fast and cheap, Big Data tools promise real-time analysis, claiming to provide guidance and orientation at a bargain price. Furthermore, by gleaning data from online sources – for example, from social networking sites such as Facebook or Twitter – computational methods are said to bridge the qualitative-quantitative divide, capturing an entire conflict or debate in

[14] Quoted from the U.S. Congress Technology Assessment Act of 1972, Public Law 92-484, § 2(d) and § 3(c), which created the now defunct Office of Technology Assessment (OTA), see: https://www.gpo.gov/fdsys/pkg/STATUTE-86/pdf/STATUTE-86-Pg797.pdf [Accessed 20 June 2018].

[15] For a selective overview of public engagement methods, see Parliamentary Office of Science and Technology (2001); for a review and critical discussion of large-scale survey research – and its paradigms – see Bauer (2008).

[16] For a list of participatory methods, including time and cost estimates, see Involve (2005) and the Participation Compass: http://participationcompass.org/article/index/method [Accessed 20 June 2018).

[17] As an indication of this interest in a European policy context, consider studies such as the commissioned report Big Data Analytics for Policy Making (EC, 2016b), events such as the EurActiv stakeholder workshop Big Data & Policy Making, see http://www.euractiv.com/section/digital/video/big-data-and-policy-making/, or research initiatives such as the Framework Programme 7 projects SENSEI, see http://www.sensei-conversation.eu/, and EuroSentiment, see https://web.archive.org/web/20160629170848/http://eurosentiment.eu [Accessed 20 June 2018].

full detail. Finally, presented as a disinterested reading of reality, the Big Data marketing narrative feeds into the ideal of "mechanical objectivity" (Daston and Galison, 2010), thus seemingly solving the problem of individual or institutional research bias by purely technical means. While scholars from different disciplinary backgrounds have pointed out the practical limitations and conceptual flaws of this heralded methodological revolution (e.g., see Crawford, 2013; Mustafaraj et al., 2011), the ability to measure public attitudes directly and without delay presents a compelling prospect for a political system facing issues of trust and uncertainty. And even though TA's role in facilitating public participation in science and technology governance goes well beyond the collection of views and opinions, the systematic monitoring of user-generated online data for feedback gathering and trend analysis may soon give traditional engagement methods a run for their money. In a society where more and more of people's interactions have migrated to the Web, the opportunities for such research proliferate – and policy makers are taking note (see Grubmüller et al., 2013). The question of how TA should respond to this challenge will thus be crucial to the field's future development.

3.2. Means and methods for exploring the future

Technology assessment's relationship to the future is the next potential site of competition. As mentioned above, the orientation towards the future is a central element of TA and has been a guiding issue from its very beginning. Yet, following some disillusionment with positivistic and deterministic "prognosticism" (Grunwald, 2009), there has been a shift from notions of early warning and control to those of shaping and designing (see Grunwald, 2014). The future orientation of TA becomes obvious both in methodology and conceptual work: On the one hand, empirical studies on the societal impacts of technology make use of an abundance of foresight methods such as Delphi surveys, roadmapping exercises, or scenario development (see Porter, 2010). On the other hand, conceptual work on technology futures has blossomed in recent years, and there have been numerous attempts – both from within and beyond TA – to conceptually grasp the dynamic and performative relationships between past, present, and future (see Esposito, 2007; Brown and Michael, 2003). Two prominent non-deterministic approaches at the core of the TA movement are the concept of "anticipatory governance" (Guston, 2014) and the notion of "technology futures" (Grunwald, 2012). While the former advocates broad-based capacity building to manage emerging technologies as

long as such management is still possible (see Guston, 2008), the latter stresses the value of technology futures as a common point of reference between developers, political actors, and the wider public, thus emphasizing their contribution to a sustainable co-evolution of technology and society by stimulating critical reflection and debate (see Grunwald, 2012). What unites the two approaches – and TA foresight practices in general – is an understanding of the future as open and malleable, as something that can be steered and shaped, not "determined by natural necessities, but contingent and influenced by human action" (Voß et al., 2006: 166). Avoiding any 'crystal ball ambitions', contemporary TA conceptualizes foresight and anticipation as a fundamentally democratic practice, an inclusive societal learning process that is meant to reduce the costs of learning by trial and error.

The reduction of costs through future-oriented analysis is also one of the main selling points of Big Data. The methods employed, however, differ considerably: Instead of deliberative foresight, Big Data specializes in predictive forecasting; instead of negotiating plausible future scenarios, Big Data technologies estimate probable future trajectories. In essence, historical and near-time data are used to identify patterns and trends, marking an epistemological shift from futures as socially created (see Adam, 2005) to the future as an object of machine calculation. The spirit of positivism thus returns (see Jurgenson, 2014), and it appears more powerful than ever, fueled by a massive increase in data availability and advanced tools and techniques to process and leverage them. The real challenge for TA, however, is arguably not so much the upcoming field of data science,[18] but, once again, the Big Data imaginary, which promises almost universal applicability (see Anderson 2008) and the restoration of certainty in uncertain times (see Hardy, 2013). By rendering the future knowable and its outcome optimizable,[19] Big Data "revitalize[s] the promise of prediction across social, political, and economic worlds" (Aradau and Blanke, 2016: 2), becoming both product and enabler of a new "regime of futurity" (Ekbia et al., 2015: 1539), in which slower and less accurate methodologies are considered obsolete. In a society that increasingly thinks and lives towards the future, that is marked by a constant need to conquer and colonize the 'not yet' (see Adams et al., 2009), technology assessment's deliberative modes of future

[18] Data scientists are usually well aware of the various limitations of their craft. For a balanced account of prediction in the era of Big Data, see Silver (2012).

[19] Consider, for instance, IBM's advertising slogan for their predictive analytics products, which prompts customers to "optimize the future with better decisions today". See http://www.ibm.com/analytics/us/en/technology/predictive-analytics/ [Accessed 20 June 2018].

engagement may soon find themselves outgunned and outpaced by the grand claims of the algorithmic forecasting industry. Once again, an open discussion of how TA as a future-oriented discipline should and could respond to this challenge seems paramount.

3.3. The provision of actionable knowledge and advice

While "anticipating future developments and their impacts" is a key objective of TA, the field also seeks to "accommodate such insights in decision making and its implementation" (Rip, 2012: 31). The focus on providing actionable knowledge for political decision making has been a major concern of TA since its formal inception in the 1970s. Back then, the now defunct Office of Technology Assessment (OTA) was commissioned to advise the US Congress in matters of science and technology. While the executive branch of the US government could rely on an extensive apparatus of departments and agencies, Congress as the legislative branch was lacking such resources. Thus, a crucial function of OTA was to re-establish the knowledge/power balance between the government's legislative and executive branches (see Sadowski, 2015; Bimber, 1996). The focus on actionable knowledge, however, becomes apparent not only in the specific case of parliamentary TA, which aims to "strengthen representative democracy by timely informing MPs about the potential social impacts of technological change" (van Est and Brom, 2012: 306), but also when participatory TA is used as a means to mediate between the interests of different stakeholders, for instance in the context of selecting sites for nuclear waste disposal (see Hocke and Renn, 2009). The distinction between consultation, on the one hand, and decision making, on the other, is crucial for the disciplinary self-understanding of TA. In order to remain trustworthy in its advisory function, TA aims at providing independent, high-quality knowledge about techno-scientific developments and their potential social, ethical, and legal implications. It does not, however, actively participate in the decision-making process.[20] Moreover, TA as a discipline is well aware that the impact of its advice varies greatly and is hard to predict. While some reports may directly influence parliamentary decisions, others may get tucked away in filling cabinets, never to be read again. This relative openness and uncertainty should not be seen

[20] For an overview of the different practices and institutions of parliamentary TA in Europe, see Nentwich (2016).

as a failure of the discipline, but as a testimony to its facilitative and supportive rather than deciding societal function.

In the case of Big Data, the distinction between consultation and decision making is far less obvious: While Big Data technologies are said to provide insight and guidance for human decision making, they are increasingly used to generate decision recommendations or even take action on their own (see Citron and Pasquale, 2014). What can thus be observed is a gradual shift from description (i.e., data reporting) and prediction (i.e., identifying trends) to prescription and automation (see Davenport, 2015). Whereas prescriptive analytics are meant to suggest actions and "tell you what to do" (Davenport, 2013), the move towards automation shifts the power – and burden – of decision making from the human actor to ever smarter programs and machines. In the latter case, algorithmic systems do not merely participate in decision processes, but perform certain actions with no or minimal human intervention. And even though a fully automated state may still be a distant utopian or dystopian vision (see Forster, 1909), there is clear indication that the demand for such solutions in the public sector is growing (see Hartzog et al., 2015). In such a context, a major task for TA will be to critically assess and question the "prominence and status acquired by data as a commodity and recognized output" (Leonelli, 2014: 1) as well as to challenge the "widespread belief that large data sets offer a higher form of intelligence […], with an aura of truth, objectivity, and accuracy" (boyd and Crawford, 2012: 663). In addition, however, the field will also have to develop strategies to maintain its relevance in a crisis-ridden political environment that longs for seemingly clean, unambiguous knowledge and advice. To be blunt, the recent push for "data for policy" (EC, 2016c) and "evidence-informed decision making" (EC, 2015a) does not aim to raise the budget for traditional public engagement exercises, but encourages the development of computational solutions that may make such methodologies appear increasingly redundant. Going forward, fast-paced innovation and the ongoing datafication of society will make this an even more pressing matter of concern.

4. Discussion: Towards Responsible Data-Based Research and Innovation

The relationship between Big Data as a complex socio-technical phenomenon and TA as a discipline assessing such phenomena is a peculiar one: Big Data may be the first topic TA deals with that is not only an object of inquiry, but also a major competitor, rivaling TA in several

of its core functions. Having outlined a narrative of competition in the previous section, we now want to conclude by sketching an alternative way forward, one that considers the relationship between Big Data and TA with respect to the concept of RRI. In essence, we believe that TA's focus on multi-, inter-, and transdisciplinary research, its reflexive orientation towards the future, and its practical expertise in providing policy advice, can help to address and successfully deal with the manifold challenges posed by Big Data. In doing so, TA-based analysis may provide valuable insight and support for the alignment of Big Data governance with the aims and goals of RRI, a widespread policy agenda the European Commission broadly defines as "an approach that anticipates and assesses potential implications and societal expectations […], with the aim to foster the design of inclusive and sustainable research and innovation" (EC, n.d. b). While we do not wish to engage in a detailed discussion of RRI as a concept and funding strategy (see, however, Simon, 2017; 2015), we do seek to highlight a number of central issues and concerns that may require special attention when moving Big Data under the RRI umbrella.

4.1. Multidisciplinarity beyond ELSIfication

As the range of topics and issues covered in journals such as *Big Data & Society* indicates, the scope and complexity of the Big Data phenomenon extends well beyond the purview of any single academic discipline. Meaningful assessments of societal impacts will thus require the collaboration of researchers from different fields, not only contributing their domain-specific knowledge, but also engaging in cross-disciplinary investigations, considering complex socio-technical entanglements from various angles and perspectives. As argued, TA as an analytic practice is well equipped for such a task, but the selection of the relevant scientific disciplines is both crucial and tricky and should be made with care (see Decker, 2004). While the choice ultimately depends on the question to be answered and the problem to be solved, in the case of Big Data, the scope of traditional ELSI research often will not suffice. In particular political (see Morozov, 2013a), economic (see Newman, 2015), and epistemic (see Rieder and Simon, 2017) premises and implications should be taken into account, and some technical expertise may prove necessary when dealing with matters related to advanced computational methods such as data mining or machine learning (e.g., see Barocas and Selbst, 2016; Burrell, 2016). Thus, in order to truly grasp the impacts and consequences of Big Data and path the way for

more responsible data-based research and innovation, finding the 'right' research partners and establishing a high-quality exchange relationship will be key.

4.2. Public Engagement and the Politics of Participation

The idea(l) of RRI emphasizes the need for "deepening the relationship between science and society" (European Parliament and Council, 2013) by "includ[ing] multi-actor and public engagement in research and innovation" (EC, n.d. b), fostering "dialogues between researchers, policy makers, industry and civil society organizations, NGOs, and citizens" (EC, n.d. a). While the general aim of "bringing on board the widest possible diversity of actors" (EC, n.d. a) may be democratically laudable, and TA certainly has a lot to offer in this regard (see Section 2 above), the specific modes and modalities of engagement remain a major issue of concern. Despite a rhetoric of openness and inclusion, consultation exercises are frequently designed as one-way, top-down public education approaches where participants are taught about scientific facts, expert knowledge is given primacy over lay expertise, and a strong commitment to consensus stifles deliberative disagreement. Given Big Data's now well-documented potential for causing harm (see O'Neil, 2016) and people's growing discomfort with the widespread application and impact of data analytics in numerous areas of life (see Pew, 2015), such strategies of appeasement seem both futile and utterly misplaced. Instead, controversies should be embraced as sites of social learning where citizens can share and discuss their experiences with specific services and applications (see Rip, 1986). Moreover, measures should be taken that the outcomes and findings of deliberative engagements can actually affect the regulation of new and emerging technologies, which means that even stopping certain developments – for example, through temporary or permanent moratoria – must be considered a real option. Otherwise, public consultation and stakeholder involvement risks becoming a farce and may rightly be accused as a means to silence critical voices and fabricate consent.

4.3. Performative Temporalities and Contestable Futures

As outlined in Section 3, TA and Big Data practices share a common interest in the future. While Big Data is mainly discussed as a forecasting technology that may soon "predict our

every move" (Hassani and Silva, 2015), TA is rooted in the broader and less determinant but equally forward looking tradition of foresight (see Harper, 2013), focusing on different stakeholder's visions, expectations, and fears rather than statistical patterns. From a TA perspective, the future cannot be discovered, but has to be created and constructed – for instance through scenario building – resulting in a plurality of possible futures and paths that are open to public scrutiny and deliberation. In addition, TA emphasizes the interconnectedness between past, present, and future, acknowledging that we can only think about futures according to our present-day's knowledge and that the ways how futures are constructed are decisive for their content (see Grunwald, 2010). Such a reflexive stance may prove valuable when assessing the performativity and politics of Big Data forecasts, which are marked by a shift from prediction to prescription, no longer limited to the confines of prognosis, but actively telling people "what they should be doing next" (Eric Schmidt, quoted in Jenkins, 2010). Ultimately, Big Data's predictive power may enable a "new philosophy of preemption" (Kerr and Earle, 2013), which forestalls (human) action based on algorithmic estimates. If we allow such systems of digital regulation to proliferate, we need to be very sure about their epistemological premises as well as their potential ethical, social, and political implications.

4.4. Actionable, Situated, and Inclusive Knowledge

Given the prevalent "trust in numbers" (Porter, 1995) within political and administrative circles, on the one hand, and Big Data proponents' claims of predictive superiority and analytical neutrality (see Anderson 2008), on the other, TA may soon face a new competitor in providing guidance and support for public policy. But policy makers should be aware that the very kind of knowledge they receive may differ considerably between the two approaches: While TA focuses on collective problem solving, discussions and critical deliberation (see Abelson et al., 2003), Big Data methods tend to look at networks and opinions from a distance (see Moretti, 2013). Two-way interactions, mutual learning, and external subject-matter expertise are key elements of TA, but much less so in data mining and computational analytics.

However, rather than substituting one for the other, we believe that both fields can learn and benefit from one another. For instance, whereas Big Data methods could expand the breadth

of traditional scoping exercises (see Gandomi and Haider, 2015), facilitate the tracking of trends (Nguyen et al., 2016) and public sentiment (see Cambria et al., 2014), and help map the dynamics of controversies over time (see Lansdall-Welfare, 2014), TA could use its methodological know-how, its reflexive capacities, and its experience in policy advice to make Big Data RRI-ready.[21] In the end, diligently supervised mixed-methods approaches could contribute to the theoretical and methodological development of both fields as well as to the general advancement of responsible data-based research and innovation.

In this concluding section, we have argued that Big Data practices may clearly benefit from the insights and experience of TA and have pointed towards certain issues and concerns that may prove crucial when seeking to align Big Data research with central RRI tenets. While our analysis had a critical edge and sought to debunk certain exaggerated hopes and claims mainly voiced by industry stakeholders, we do wish to acknowledge the great social, economic, and academic opportunities that Big Data and related methods provide. There is no doubt that Big Data tools can be used for the common good, but there are pitfalls along the way that must be thoroughly understood and addressed. Thus, while opportunities should be exploited, this needs to be done in a responsible, socially sustainable manner.

Throughout the paper, we have also argued that Big Data poses a considerable challenge to TA, rivaling the field in several core functions, including the assessment of public views and visions, means and methods for exploring possible future trajectories, and the provision of actionable knowledge and advice for political decision making. We believe that in order to stay in the game, TA will have to engage with the new methods and techniques offered by Big Data technologies. Such an engagement should include, but not be limited to, critical reflection. Instead, TA should consider forming coalitions of mutual learning, for instance by including data scientists into future project designs, thereby expanding its multidisciplinary expertise by yet another approach. What we propose is a third way between industry hype and Big Data doom and gloom, one that acknowledges the value of data science as a powerful epistemic practice, that is open to the opportunities granted by the proliferation of digital

[21] In this respect, TA could also learn from the digital methods community, which has employed Web-based tools to map controversies around, for example, global warming (see Weltevrede and Borra, 2016), biofuels (see Eklöf and Mager, 2013), or GM food (see Marres and Rogers, 2000), embracing the epistemic opportunities of online data mining while remaining attentive to potential limitations and the perils of competitive marketization (see Rieder and Sire, 2014).

(social) data, and that takes a proactive stance in developing tools and methods that function as best practice examples. If done right, modern data analytics could become an ally rather than a competitor to the field of technology assessment, potentially extending the scope, speed, and quality of discourse, dispute, and trend analysis. If disregarded and left to the proprietary discretion of commercial products, however, Big Data may not only grow to challenge the TA community, but could also pose a considerable threat to the core principles of Responsible Research and Innovation.

Acknowledgments

The authors wish to acknowledge the financial support of the Austrian Science Fund (P-23770).

References

Abelson, J., Forest, P.-G., Eyles, J., Smith, P., Martin, E. and Gauvin, F.-P. (2003). Deliberations about Deliberative Methods: Issues in the Design and Evaluation of Public Participation Processes. *Social Science & Medicine*, 57(2), pp. 239–251.

Adam, B. (2005). *Futures in the Making: Contemporary Practices and Sociological Challenges*. Paper presented at the ASA conference. [pdf] Available at: http://www.cardiff.ac.uk/socsi/futures/conf_ba_asa230905.pdf [Accessed 20 June 2018].

Adams, V., Murphy, M. and Clarke, A. E. (2009). Anticipation: Technoscience, Life, Affect, Temporality. *Subjectivity*, 28(1), pp. 246-265.

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. [online] Available at: https://www.wired.com/2008/06/pb-theory/ [Accessed 20 June 2018].

Aradau, C. and Blanke, T. (2016). Politics of Prediction: Security and the Time/Space of Governmentality in the Age of Big Data. *European Journal of Social Theory*, 20(3), pp. 373-391.

Assefa, G. and Frostell, B. M. (2007). Social Sustainability and Social Acceptance in Technology Assessment: A Case Study of Energy Technologies. *Technology in Society*, 29(1), pp 63-78.

Barben, D., Fisher, E., Selin, C. and Guston, D. H. (2008). Anticipatory Governance and Nanotechnology: Foresight, Engagement, and Integration. In: E. J. Hackett, O. Amsterdamska, M. Lynch and J. Wajcman, eds., *The Handbook of Science and Technology Studies*. Third Edition. Cambridge, MA: MIT Press, pp. 979-1000.

Barocas, S. and Selbst, A. D. (2016 [2014]). Big Data's Disparate Impact. *California Law Review*, 104, pp. 671-732.

Bauer, M. W. (2008). Survey Research and the Public Understanding of Science. In: M. Bucchi and B. Trench, eds., *Handbook of Public Communication of Science and Technology*. London: Routledge, pp. 111-130.

Bennett, I. and Sarewitz, D. (2006) Too Little, Too Late? Research Policies on the Societal Implications of Nanotechnology in the United States. *Science as Culture*, 15(4), pp. 309-325.

Bimber, B. (1996). *The Politics of Expertise in Congress. The Rise and Fall of the Office of Technology Assessment*. New York: State University of New York Press.

boyd, danah (2016). Untangling Research and Practice: What Facebook's "Emotional Contagion" Study Teaches Us. *Research Ethics*, 12(1), pp. 4-13.

boyd, d. and Crawford, K. (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, 15(5), pp. 662-679.

Braun, K., Moore, A., Herrmann, S. L., Könninger, S. (2010). Science Governance and the Politics of Proper Talk: Governmental Bioethics as a New Technology of Reflexive Government. *Economy and Society*, 39(4), pp. 510-533.

Brown, N. and Michael, M. (2003). A Sociology of Expectations: Retrospecting Prospects and Prospecting Retrospects. *Technology Analysis and Strategic Management*, 15(1), pp. 3-18.

Burrell, J. (2016). How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society*, 3(1), pp. 1-12.

Cambria, E., Rajagopal, D., Olsher, D. and Das, D. (2014). Big Social Data Analysis. In: R. Akerkar, ed., *Big Data Computing*. Boca Raton, FL: CRC Press, pp. 401-414.

Chilvers, J. and Kearnes, M. (2016). Participation in the Making. Rethinking Public Engagement in Co-Productionist Terms. In: J. Chilvers and M. Kearnes, eds., *Remaking Participation. Science, Environment and Emergent Publics*. Abingdon: Routledge, pp. 31-63.

Christin, A., Rosenblat, A. and boyd, d. (2015). *Courts and Predictive Algorithms*. Workshop Primer. [pdf] Available at: http://www.datacivilrights.org/pubs/2015-1027/Courts_and_Predictive_Algorithms.pdf [Accessed 20 June 2018].

Citron, D. K. and Pasquale, F. (2014). The Scored Society: Due Process for Automated Predictions. *Washington Law Review*, 89(1), pp. 1-33.

Collingridge, D. (1980). *The Social Control of Technology*. London: Frances Pinter.

Cotton, M. (2014). *Ethics and Technology Assessment: A Participatory Approach*. Heidelberg; New York: Springer.

Crawford, K. (2013). The Hidden Biases in Big Data. *Harvard Business Review*. [online] Available at: https://hbr.org/2013/04/the-hidden-biases-in-big-data [Accessed 20 June 2018].

Daston, L. and Galison, P. (2010 [2007]). *Objectivity*. New York: Zone Books.

Davenport, T. H. (2015). The Rise of Automated Analytics. *The Wall Street Journal*. [online] Available at: http://blogs.wsj.com/cio/2015/01/14/the-rise-of-automated-analytics/ [Accessed 20 June 2018].

Davenport, T. H. (2013). Analytics That Tell You What to Do. *The Wall Street Journal*. [online] Available at: http://blogs.wsj.com/cio/2013/04/03/analytics-that-tell-you-what-to-do/ [Accessed 20 June 2018].

Decker, M. (2001). Preface. In: M. Decker, ed., *Interdisciplinarity in Technology Assessment. Implementation and its Chances and Limits*. Wissenschaftsethik und Technikfolgenbeurteilung Band 11. Berlin: Springer, IX-X.

Decker, M. (2004). The Role of Ethics in Interdisciplinary Technology Assessment. *Poiesis & Praxis*, 2, pp. 139-156.

Decker, M. and Grunwald, A. (2001). Rational Technology Assessment as Interdisciplinary Research. In: M. Decker, ed., *Interdisciplinarity in Technology Assessment. Implementation and its Chances and Limits*. Wissenschaftsethik und Technikfolgenbeurteilung Band 11. Berlin: Springer, pp. 33-60.

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V. R., Tsou, A., Weingart, S. and Sugimoto, C. R. (2015). Big Data, Bigger Dilemmas: A Critical Review. *Journal of the Association for Information and Technology*, 66(8), pp. 1523-1545.

Eklöf, J. and Mager, A. (2013). Technoscientific Promotion and Biofuel Policy: How the Press and Search Engines Stage the Biofuel Controversy. *Media, Culture & Society*, 35(4), pp. 454-471.

Esposito, E. (2007). *Die Fiktion der Wahrscheinlichen Realität*. Frankfurt am Main: Suhrkamp.

van Est, R. and Brom, F. (2012). Technology Assessment as an Analytic and Democratic Practice. In: R. Chadwick, D. Callahan, P. Singer, eds., *Encyclopedia of Applied Ethics*. 2nd Edition. San Diego: Academic Press, pp. 306-320.

European Commission (n.d. a). *Public Engagement in Responsible Research and Innovation*. [online] Available at: https://ec.europa.eu/programmes/horizon2020/node/766 [Accessed 20 June 2018].

European Commission (n.d. b). *Responsible Research & Innovation*. [online] Available at: https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation [Accessed 20 June 2018].

European Commission (2016a). *European Cloud Initiative to Give Europe a Global Lead in the Data-Driven Economy*. Press Release. [online] Available at: http://europa.eu/rapid/press-release_IP-16-1408_en.htm [Accessed 20 June 2018].

European Commission (2016b). *Big Data Analytics for Policy Making*. Report. [online] Available at: https://joinup.ec.europa.eu/sites/default/files/dg_digit_study_big_data_analytics_for_policy_making.pdf [Accessed 20 June 2018].

European Commission (2016c). *Data for Policy: An Increasingly Global Affair*. Conference Announcement. [online] Available at: https://ec.europa.eu/digital-single-market/en/news/data-policy-increasingly-global-affair [Accessed 20 June 2018].

European Commission (2015a). *Data for Policy: When the Haystack is Made of Needles*. Call for Contributions. [online] Available at: https://ec.europa.eu/digital-single-market/news/data-policy-when-haystack-made-needles-call-contributions [Accessed 20 June 2018].

European Commission (2015b). *ICT-Enabled Public Sector Innovation in Horizon 2020*. [online] Available at: https://ec.europa.eu/digital-single-market/en/ict-enabled-public-sector-innovation-horizon-2020 [Accessed 20 June 2018].

European Commission (2015c). *Special Eurobarometer 431 – Data Protection*. Report. [pdf] Available at: http://ec.europa.eu/public_opinion/archives/ebs/ebs_431_en.pdf [Accessed 20 June 2018].

European Commission (2014). *The Use of Big Data in Public Health Policy and Research.* [pdf] Available at: http://ec.europa.eu/health/ehealth/docs/ev_20141118_co07b_en.pdf [Accessed 20 June 2018].

European Parliament and Council (2013). *Regulation (EU) No 1291/2013 of the European Parliament and of the Council of 11 December 2013 establishing Horizon 2020 – the Framework Programme for Research and Innovation (2014-2020) and repealing Decision No 1982/2006/EC Text with EEA relevance*. [online] Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013R1291&from=EN [Accessed 20 June 2018].

Felt, U., Schumann, S., Schwarz, C. and Strassnig, M. (2014). Technology of Imagination: A Card-Based Public Engagement Method for Debating Emerging Technologies. *Qualitative Research*, 14(2), pp. 233-251.

Forster, E. M. (1909). *The Machine Stops*. [online] Available at: https://en.wikisource.org/wiki/The_Machine_Stops [Accessed 20 June 2018].

Funtowicz, S. O. and Ravetz, J. R. (1993). Science for the Post-Normal Age. *Futures*, 25(7), pp.739-755.

Gandomi, A. and Haider, M. (2014). Beyond the Hype: Big Data Concepts, Methods, and Analytics. *International Journal of Information Management*, 35(2), pp. 137-144.

Gartner (2015). *Gartner Survey Shows More Than 75 Percent of Companies Are Investing or Planning to Invest in Big Data in the Next Two Years*. [online] Available at: http://www.gartner.com/newsroom/id/3130817 [Accessed 20 June 2018].

Grubmüller, V., Götsch, K. and Krieger, B. (2013). Social Media Analytics for Future Oriented Policy Making. *European Journal of Futures Research*, 1(20), pp. 1-9.

Grunwald, A. (2014). Technology Assessment for Responsible Innovation. In: J. van den Hoven, N. Doorn, T. Swierstra, B.-J. Koops, H. Romijin, eds., *Responsible Innovation 1: Innovation Solutions for Global Issues*. Dordrecht: Springer, pp. 15-32.

Grunwald, A. (2012). *Technikzukünfte als Medium von Zukunftsdebatten und Technikgestaltung*. Karlsruhe: KIT Scientific Publishing.

Grunwald, A. (2010). Future. In: D. H. Guston, ed., *Encyclopedia of Nanoscience and Society. Volume 1.* Los Angeles: SAGE, pp. 265-267.

Grunwald, A. (2009). Technology Assessment: Concepts and Methods. In: A. Meijers, ed., *Philosophy of Technology and Engineering Sciences.* Handbook of the Philosophy of Science. Volume 9. Amsterdam: Elsevier, pp. 1103-1146.

Guston, D. H. (2014). Understanding 'Anticipatory Governance'. *Social Studies of Science*, 44(2), pp. 218-242.

Guston, D. H. (2008). Preface. In: E. Fisher, C. Selin and J. Wetmore, eds., *The Yearbook of Nanotechnology in Society. Volume 1: Presenting Futures*. New York: Springer, pp. V-VIII.

Hardy, Q. (2013). Why Big Data is not Truth. [online] *The New York Times*. Available at: http://bits.blogs.nytimes.com/2013/06/01/why-big-data-is-not-truth/ [Accessed 20 June 2018].

Hartzog, W., Conti, G., Nelson, J. and Shay, L. A. (2015). Inefficiently Automated Law Enforcement. *Michigan State Law Review*, 2015(5), pp. 1763-1796.

Harper, J. C. (2013). *Impact of Technology Foresight.* Nesta Working Paper 13/16. [pdf] Available at: https://media.nesta.org.uk/documents/1316_impact_of_technology_foresight_final_version.pdf [Accessed 20 June 2018].

Hassani, H. and Silva, E. S. (2015). Forecasting with Big Data: A Review. *Annals of Data Science*, 2(1), pp. 5-19.

Hey, T., Tansley, S. and Tolle, K., eds., (2009). *The Fourth Paradigm. Data-Intensive Scientific Discovery*. Redmond: Microsoft Research.

Hocke, P. and Renn, O. (2009). Concerned Public and the Paralysis of Decision-Making: Nuclear Waste Management Policy in Germany. *Journal of Risk Research*, 12(7-8), pp. 921-940.

International Data Research (IDC) (2014). *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*. [online] Available at: www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm [Accessed 20 June 2018].

Involve (2005). *People & Participation. How to Put Citizens at the Heart of Decision-Making*. [pdf] Available at: http://www.involve.org.uk/wp-content/uploads/2011/03/People-and-Participation.pdf [Accessed 20 June 2018].

Jenkins, H. W. (2010). Google and the Search for the Future. *The Wall Street Journa*l. [online] Available at: http://www.wsj.com/articles/SB10001424052748704901104575423294099527212 [Accessed 20 June 2018].

Joss, S. and Bellucci, S. (2002). *Participatory Technology Assessment: European Perspectives*. Centre for the Study of Democracy. Gateshead: Athenaeum Press.

Jurgenson, N. (2014). View from Nowhere. *The New Inquiry*. [online] Available at: https://thenewinquiry.com/essays/view-from-nowhere/ [Accessed 20 June 2018].

Kerr, I. and Earle, J. (2013). Prediction, Preemption, Presumption. How Big Data Threatens Big Picture Privacy. *Stanford Law Review Online*, 66, pp. 65-72.

Kitchin, R. (2014). Big Data, New Epistemologies and Paradigm Shifts. *Big Data & Society*, 1(1), pp. 1-12.

Klüver, L., Nielsen, R. Ø. and Jørgensen, M. L. (2015). *Policy-Oriented Technology Assessment Across Europe. Expanding Capacities*. Houndmills: Palgrave Macmillan.

Lafont, C. (2015). Deliberation, Participation, and Democratic Legitimacy: Should Deliberative Mini-Publics Shape Public Policy? *The Journal of Political Philosophy*, 23(1), pp. 40-63.

Laney, D. (2012). Deja VVVu: Others Claiming Gartner's Construct for Big Data. *Gartner Blog Network*. [online] http://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/ [Accessed 20 June 2018].

Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. *META Group. Application Delivery Strategies*. [pdf] Available at:

https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf [Accessed 20 June 2018].

Lansdall-Welfare, T., Sudhahar, S., Veltri, G. A. and Christianini, N. (2014). On the Coverage of Science in the Media: A Big Data Study on the Impact of the Fukushima Disaster. *Proceedings of the 2014 IEEE International Conference on Big Data*, pp. 60-66.

Law, J. (2009). Seeing Like a Survey. *Cultural Sociology*, 3(2), pp. 239-256.

Leonelli, S. (2014). What Difference Does Quantity Make? On the Epistemology of Big Data in Biology. *Big Data & Society*, 1(1), pp. 1-11.

Liebert, W. and Schmidt, J. C. (2010). Collingridge's Dilemma and Technoscience. An Attempt to Provide a Clarification from the Perspective of the Philosophy of Science. *Poiesis & Praxis*, 7, pp. 55-71.

Lohr, S. (2012). The Age of Big Data. *The New York Times*. [online] Available at: http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?_r=0 [Accessed 20 June 2018].

Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, Consequences, Critique. *Big Data and Society*, 1(2), pp. 1-13.

Macnaghten, P., Davies, S. and Kearnes, M. (2010). Narrative and Public Engagement: Some Findings from the DEEPEN Project. In: R. von Schomberg and S. Davies, eds., *Understanding Public Debate on Nanotechnologies. Options for Framing Public Policy*. Luxembourg: Publications Office of the European Union, pp. 13-29.

Marres, N. and Rogers, R. (2000). Depluralising the Web and Repluralising Public Debate: The Case of the GM Food Debate on the Web. In: R. Rogers, ed., *Preferred Placement: Knowledge Politics on the Web*. Maastricht: Jan van Eyck Editions, pp. 113-135.

Mayer-Schönberger, V. and Cukier, K. (2013). *Big Data. A Revolution That Will Transform How We Live, Work, And Think*. New York: Houghton Mifflin Harcourt.

Meyer, M. (2011). The Rise of the Knowledge Broker. *Science Communication*, 32(1), pp. 118-127.

Michalek, T., Hebáková, L., Hennen, L., Scherz, C., Nierling, L. and Hahn, J. (2014). Technology Assessment and Policy Areas of Great Transitions. *Proceedings from the PACITA 2013 Conference in Prague*. [pdf] Available at: http://www.pacitaproject.eu/wp-content/uploads/2014/11/Conference-Book.pdf [Accessed 20 June 2018].

Moretti, F. (2013). *Distant Reading*. London: Verso.

Morozov, E. (2013a). The Real Privacy Problem. *MIT Technology Review*. [online] Available at: https://www.technologyreview.com/s/520426/the-real-privacy-problem/ [Accessed 20 June 2018].

Morozov, E. (2013b). *To Save Everything, Click Here. The Folly of Technological Solutionism*. New York: Public Affairs.

Mustafaraj, E., Finn, S., Whitlock, C. and Metaxas, P. T. (2011). *Vocal Minority Versus Silent Majority: Discovering the Opinions of the Long Tail*. Paper presented at the 2011 IEEE

International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, pp. 103-110.

Nentwich, M. (2016). Parliamentary Technology Assessment Institutions and Practices. A Systematic Comparison of 15 Members of the EPTA Network. *ITA manu:script*, 16-02. [pdf] Available at: http://epub.oeaw.ac.at/ita/ita-manuscript/ita_16_02.pdf [Accessed 20 June 2018].

Newell, W. H. (2010). Undergraduate General Education. In: R. Frodeman, J. Thompson Klein, C. Mitchman, B. Holbrook and J. Britt, eds., *The Oxford Handbook of Interdisciplinarity*. New York: Oxford University Press, pp. 360-371.

Newman, N. (2015). *Data Justice: Taking on Big Data as an Economic Justice Issue*. Data Justice Report. [pdf] Available at: http://www.datajustice.org/sites/default/files/Data%20Justice-%20Taking%20on%20Big%20Data%20as%20an%20Economic%20Justice%20Issue.pdf [Accessed 20 June 2018].

Nguyen, K.-L., Byung-Joo, S. and Yoo, S. J. (2016). Hot Topic Detection and Technology Trend Tracking for Patents Utilizing Term Frequency and Proportional Document Frequency and Semantic Information. *Proceedings of the 2016 International Conference on Big Data and Smart Computing (BigComp)*, pp. 223-230.

Nowotny, H., Scott, P. and Gibbon, M. (2001). *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge: Polity Press.

O'Neil, C. (2016). *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Organisation for Economic Co-operation and Development (2015). *Data-Driven Innovation: Big Data for Growth and Well-Being*. Paris: OECD Publishing.

PACITA (2015). *The Next Horizon of Technology Assessment. 2nd European TA Conference.* Book of Abstracts. [pdf] Available at: http://berlinconference.pacitaproject.eu/wp-content/uploads/2015/02/Book-of-Abstracts-UPDATED3.pdf [Accessed 20 June 2018].

PACITA (2013). *Technology Assessment and Policy Areas of Great Transitions. 1st PACITA Project Conference.* Book of Abstracts. Available at: http://www.pacitaproject.eu/wp-content/uploads/2013/03/Book-of-abstracts-Prague.pdf [Accessed 20 June 2018].

Pan, L., Boucherie, S. and Hanafi, S. (2015). *Interdisciplinary Research: How Do 9 Nations Compare?* Elsevier Connect. [online] Available at: https://www.elsevier.com/connect/interdisciplinary-research-how-do-9-nations-compare [Accessed 20 June 2018].

Parliamentary Office of Science and Technology (2001). *Open Channels: Public Dialogue in Science and Technology*. Report No. 153. [pdf] Available at: http://www.parliament.uk/documents/post/pr153.pdf [Accessed 20 June 2018].

Pentland, A. (2014). *Social Physics. How Good Ideas Spread – The Lessons from a New Science*. New York: Penguin Press.

Pentland, A. (2012). Society's Nervous System: Building Effective Government, Energy, and Public Health Systems. *IEEE Computer*, 45(1), pp. 31-38.

Pew Research Center (2015). *Americans' Attitudes about Privacy, Security and Surveillance*. [pdf] Available at: http://www.pewinternet.org/files/2015/05/Privacy-and-Security-Attitudes-5.19.15_FINAL.pdf [Accessed 20 June 2018].

Pew Research Center (2014). *Public Perceptions of Privacy and Security in the Post-Snowden Era*. [pdf] Available at: http://www.pewinternet.org/files/2014/11/PI_PublicPerceptionsofPrivacy_111214.pdf [Accessed 20 June 2018].

Porter, A. L. (2010). Technology Foresight: Types and Methods. *International Journal of Foresight and Innovation Policy*, 6(1), pp. 36-45.

Porter, T. M. (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.

President's Council of Advisors on Science and Technology (2014). *Big Data and Privacy: A Technological Perspective.* [pdf] Available at: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf [Accessed 20 June 2018].

Press, G. (2014). 12 Big Data Definitions: What's Yours? *Forbes*. [online] Available at: https://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/#3bec4cba13ae [Accessed 20 June 2018].

Presser, L., Hruskova, M., Rowbottom, H. and Kancir, J. (2015). Care.data and Access to UK Health Records: Patient Privacy and Public Trust. *Technology Science*. [online] Available at: http://techscience.org/a/2015081103/ [Accessed 20 June 2018].

Rieder, B. and Sire, G. (2014) Conflicts of Interest and Incentives to Bias: A Microeconomic Critique of Google's Tangled Position on the Web. *New Media & Society*, 16(2), pp. 195-211.

Rieder, G. and Simon, J. (2017). Big Data: A New Empiricism and its Epistemic and Socio-Political Consequences. In: W. Pietsch, J. Wernecke and M. Ott, eds., *Berechenbarkeit der Welt? Philosophie und Wissenschaft im Zeitalter von Big Data*. Wiesbaden: Springer VS, pp. 85-105.

Rieder, G. and Simon, J. (2016). Datatrust: Or, the Political Quest for Numerical Evidence and the Epistemologies of Big Data. *Big Data & Soci*ety, 3(1), pp. 1-6.

Rip, A. (2012). Futures of Technology Assessment. In: M. Decker, A. Grunwald, Armin and M. Knapp, eds., *Der Systemblick auf Innovation. Technikfolgenabschätzung in der Technikgestaltung*. Berlin: edition sigma, pp. 29-39.

Rip, A. (1986). Controversies as Informal Technology Assessment. *Science Communication*, 8(2), pp. 349-371.

Rittel, H. W. J. and Webber, M. M. (1973). Dilemmas in a General Theory of Planning. *Policy Sciences*, 4, pp. 155-169.

Rommetveit, K., Gunnarsdóttir, K., Jepsen, K. S., Bertilsson, T. M., Verrax, F. and Strand, R. (2013). The Technolife Project: An Experimental Approach to New Ethical Frameworks

for Emerging Science and Technology. *International Journal of Sustainable Development*, 16(1/2), pp. 23-45.

Rose, N. (1999). *Power of Freedom. Reframing Political Thought*. Cambridge: Cambridge University Press.

Rubinstein, I. (2013). Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law*, 3(2), pp. 74-87.

Sadowski, J. (2015). Office of Technology Assessment: History, Implementation, and Participatory Critique. *Technology in Society*, 42, pp. 9-20.

Schmidt, J. C. (2008). Towards a Philosophy of Interdisciplinarity. An Attempt to Provide a Classification and Clarification. *Poiesis & Praxis*, 5(1), pp. 53-69.

Selin, C. (2011). Negotiating Plausibility: Intervening in the Future of Nanotechnology. *Science and Engineering Ethics*, 17, pp. 723-737.

Silver, N. (2012). *The Signal and the Noise. Why So Many Prediction Fail – But Some Don't*. New York: Penguin Press.

Simon, J. (2017). Value-Sensitive Design and Responsible Innovation. In: S. O. Hansson, ed., *Methods for the Ethics of Technology*. London: Rowman and Littlefield International, pp. 219-236.

Simon, J. (2015). Distributed Epistemic Responsibility in a Hyperconnected Era. In: L. Floridi, ed., *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Charm: Springer, pp. 145-159.

Singh, V. K., Banshal, S. K., Singhal, K. and Uddin, A. (2015). Scientometric Mapping of Research on 'Big Data'. *Scientometrics*, 105(2), pp. 727-741.

Statista (2016). *Forecast of Big Data Market Size, Based on Revenue, from 2011 to 2026 (in Billion U.S. Dollars)*. [online] Available at: http://www.statista.com/statistics/254266/global-big-data-market-forecast/ [Accessed 20 June 2018].

Stirling, A. (2014). Disciplinary Dilemma: Working Across Research Silos Is Harder than It Looks. *The Guardian*. [online] Available at: https://www.theguardian.com/science/political-science/2014/jun/11/science-policy-research-silos-interdisciplinarity [Accessed 20 June 2018].

Teknologirådet (2015). *Predictive Policing: Can Data Analysis Help the Police to Be in the Right Place at the Right Time?* [pdf] Available at: https://teknologiradet.no/wp-content/uploads/sites/19/2013/08/Predicitve-policing.pdf [Accessed 20 June 2018].

Thompson Klein, J. (2001). The Discourse of Transdisciplinarity: An Expanding Global Field. In: J. Thompson Klein, W. Grossenbacher-Mansuy, R. Häberli, B. Alain, R. W. Scholz and M. Welti, eds., *Transdisciplinarity: Joint Problem Solving among Science, Technology, and Society*, pp. 35-44. Basel: Spring Basel AG.

Turow, J., Hennessy, M. and Draper, N. (2015). *The Tradeoff Fallacy. How Marketers Are Misrepresenting American Consumers and Opening Them Up to Exploitation*. A Report from the Annenberg School for Communication. [pdf] Available at:

https://www.asc.upenn.edu/sites/default/files/TradeoffFallacy_1.pdf [Accessed 20 June 2018].

Upturn (2014). *Civil Rights, Big Data, and Our Algorithmic Future. A September 2014 Report on Social Justice and Technology*. [pdf] Available at: https://bigdata.fairness.io/wp-content/uploads/2015/04/2015-04-20-Civil-Rights-Big-Data-and-Our-Algorithmic-Future-v1.2.pdf [Accessed 20 June 2018].

Vance, A. (2010). EMC Takes a Big-Data Swipe at Oracle, H.P. and I.B.M. *The New York Times*. [online] Available at: http://bits.blogs.nytimes.com/2010/10/13/emc-takes-a-big-data-swipe-at-oracle-h-p-and-i-b-m/ [Accessed 20 June 2018].

Voß, J.-P., Truffer, B. and Konrad, K. (2006). Sustainability Foresight: Reflexive Governance in the Transformation of Utility Systems. In: J.-P. Voß, D. Bauknecht, R. Kemp, eds., *Reflexive Governance for Sustainable Development*. Cheltenham: Edward Elgar, pp. 162-188.

Weltevrede, E. and Borra, E. (2016). Platform Affordances and Data Practices: The Value of Dispute on Wikipedia. *Big Data & Society*, 3(1), pp. 1-16.

Wesselink, A., Paavola, J., Fritsch, O. and Renn, O. (2011). Rationales for Public Participation in Environmental Policy and Governance: Practitioners' Perspectives. *Environment and Planning A: Economy and Space,* 43(11), pp. 2688-2704.

White House (2016). *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights.* [pdf] Available at: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf [Accessed 20 June 2018].

White House (2014). *Big Data: Seizing Opportunities, Preserving Values.* Available at: https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [Accessed 20 June 2018].

Wilsdon, J., Doubleday, R. and Hynard, J. (2015). Future Directions for Scientific Advice in Europe. In: J. Wilsdon and R. Doubleday, eds., *Future Directions for Scientific Advice*. Cambridge: Centre for Science and Policy. [pdf] Available at: http://sro.sussex.ac.uk/54004/4/future-directions-for-scientific-advice-in-europe-v10.pdf [Accessed 20 June 2018].

Wood, F. B. (1982). The Status of Technology Assessment. A View from the Congressional Office of Technology Assessment. *Technological Forecasting and Social Change*, 22, pp. 211-222.

Youtie, J., Porter, A. L. and Huan, Y. (2017). Early Social Science Research about Big Data. *Science and Public Policy*, 44(1), pp. 65-74.

Paper V

Datafictions: Or, How Measurements and Predictive Analytics Rule Imagined Future Worlds

by Gernot Rieder and Thomas Völker

Presented Version:

Unpublished manuscript
(currently under review for publication)

Datafictions: Or How Measurements and Predictive Analytics Rule Imagined Future Worlds

by Gernot Rieder and Thomas Völker

Abstract

As the digital revolution continues and our lives become increasingly governed by smart technologies, there is a rising need for reflection and critical debate about where we are, where we are headed, and where we want to be. Against this background, the paper suggests that one way to foster such discussion is by engaging with the world of fiction, with imaginative stories that explore the spaces, places, and politics of alternative realities. Hence, after a concise discussion of the concept of speculative fiction, we introduce the notion of datafictions as an umbrella term for speculative stories that deal with the datafication of society in both imaginative and imaginable ways. We then outline and briefly discuss fifteen datafictions subdivided into five main categories: surveillance; social sorting; prediction; advertising and corporate power; hubris, breakdown, and the end of Big Data. In a concluding section, we argue for the increased use of speculative fiction in education, but also as a tool to examine how specific technologies are culturally imagined and what kind of futures are considered plausible given current implementations and trajectories.

Keywords

Big Data, science fiction, speculative fiction, algorithmic control, responsible research and innovation, anticipatory governance

1. Introduction

Our world is changing rapidly and profoundly. In ever more contexts, computers, sensors, and software are taking command. What can be observed is the gradual replacement of human decision making with algorithmic systems that guide and govern everyday life. What people buy, read, or watch, how doctors diagnose, the police operate, or judges judge, whether someone qualifies for a job, a loan, or an education – to an increasing degree, modern society

is permeated by 'smart' digital technologies that shape and control our collective and individual futures.

The changes society is undergoing are not easy to grasp. Throughout the past years, the computerization of everything has progressed at breakneck speed, effectively outpacing our ability to carefully reflect, debate, and legislate. More recently, and in part as a reaction to the excitement around Big Data (see Mayer-Schönberger and Cukier, 2013), one can, however, observe the development of a critical discourse, with a growing body of literature – both academic and otherwise – examining the pitfalls of a "scored society" (Citron and Pasquale, 2014) in which automated systems rank and rate individuals without pause or limit. Primary concerns range from issues around privacy and surveillance (see Lyon, 2014) over the potential negative effects of algorithmic bias and discrimination (see O'Neil, 2016) to questions around transparency and accountability (see Pasquale, 2015), all with a strong focus on the here and now, the problems and challenges of our early 21st-century digital situation. What tends to receive less attention, however, are possible longer-term developments and related cultural shifts – thoughtful inquiries into the *what if* rather than the *what is*. In a time marked by a constant state of anticipation, of "thinking and living toward the future" (Adams et al., 2009: 246), we concern ourselves remarkably little with the prospect of becoming an increasingly datafied, software-regulated society, dominated by the dictates of algorithmic gods.

One reason for the lack of critical engagement with all but the most immediate consequences of widespread computerization is that the contours of our digital future are still blurry and unclear. As indicated, the pace of ICT-led innovation is high, with new products and services launching every day (see Jaruzelski et al., 2016), multi-sided business models challenging established customer-vendor relationships (see Goodman, 2015), and legislation not only lagging behind, but also quite different in different parts of the world (see DLA Piper, 2017). Factors such as these add volatility and increase uncertainty, making future developments both hard to predict and difficult to discuss.[1] Nevertheless, as scholars of Science and Technology Studies (STS) have emphasized, the ability to engage in forward-looking deliberation is vital for a healthy democratic system that assumes responsibility and prepares for the potential long-term implications of wider socio-technical change (e.g., see

[1] For two classic versions of this argument, see Toffler (1970) and Collingridge (1980).

Michelson, 2016). Rather than seeking accurate predictions, however, such *anticipatory modes of governance* are said to require a broad societal capacity to "collectively imagine, critique, and thereby shape issues presented by emerging technologies before they become reified in particular ways" (Barben et al., 2008: 992-93). Consequently, a more proactive approach to engaging with our algorithmic future would necessitate the cultivation of a *critical imaginative gaze* that a) contextualizes new technologies within plausible future scenarios, b) illustrates the different ways in which these technologies could evolve, c) opens up technological trajectories to considerations of social desirability, and d) develops robust capacities to deal with – and thus govern – unforeseen consequences and events (see Karinen and Guston, 2010).

In an effort to contribute to a greater *future consciousness*, this article examines a particular kind of literature – speculative fiction – asking whether, or in what ways, the stories and narratives presented may encourage reflective engagement with, and a better understanding of, the possible detrimental effects of expanding algorithmic regulation.[2] While our review of works of fiction does not follow a strict analytical framework, our focus lies on power relations and their technology-provoked renegotiation. What functions do these technologies fulfil? What laws do they follow? What status do they acquire? What hierarchies do they impose? The paper is intended as a discussion starter rather than a definitive study or guide, arguing for the value of fiction as a tool for critical reflection and providing ideas for "what to read and watch before the robots take over" (Bergstein, 2017).

The article is structured as follows. First, Section 2 defines and discusses the notion of *speculative fiction*, providing some conceptual clarifications while comparing it to its sister term science fiction. Section 3 then introduces the notion of *datafictions* as an umbrella term for speculative stories that deal with the datafication of society in imaginative ways. In Section 4, we outline and briefly discuss fifteen datafictions, subdivided into five main categories: surveillance; social sorting; prediction; advertising and corporate power; hubris, breakdown,

[2] Of course, the application of advanced data analytics entails not only risks but also opportunities. This article neither denies this nor argues for a return to an analog age. Yet, given the rate and scale of digital transformation, it advocates for raised awareness that certain gains and benefits may come at a cost. To work towards a better understanding of the nature of these costs should not be interpreted as a sign of innovation-blocking technophobia, but as a way of preparing for and navigating an already highly technologized world – and, presumably, future (see Felt, 2015; Callon et al., 2009 [2001]).

and the end of Big Data. In a concluding segment, we argue for the use of speculative fiction in education, but also as a tool to uncover and examine collective socio-technical imaginations inscribed into the spaces and places of various brave new worlds.

2. On Speculative Fiction

The term "speculative fiction" has a long and contested history. Its origins are usually traced back to an 1889 issue of *Lippincott's Monthly Magazine*, where two then-recent publications – Edward Bellamy's novel *Looking Backward 2000-1887* and George Parsons Lathrop's short story *The New Poverty* – were described as "speculative fiction put in the future tense" (Egan, 1889: 597). The notion was further popularized by the American author Robert A. Heinlein, who stated that he would "prefer the term to science fiction", stressing its ability to better capture the field's concern with "sociology, psychology, [and] esoteric aspects of biology" (Heinlein, 1949: 49). Heinlein was careful, however, to differentiate speculative fiction from fantasy, arguing that the former would "rul[e] out the use of anything as material which violates established scientific fact, laws of nature […], i.e., it must be possible to the universe as we know it" (ibid.). In a similar vein, Margaret Atwood, acclaimed author of several dystopian novels, has shown reluctance to classify her work as science fiction, where "things happen that are not possible today", instead favoring speculative fiction as a label for prose where "nothing inconceivable takes place, and the projected trends […] are already in motion" (Atwood, 2005b: 92). As Atwood explains, "We've done it, or we're doing it, or we could start doing it tomorrow" (ibid.).

While Atwood's attempt to position speculative fiction as a genre separate from and opposite to science fiction has not gone unchallenged (see Le Guin, 2009), the distinction draws attention to the issue that much of what is commonly considered science fiction involves dubious plots and technologies that have no or very little connection to the world we live in. The problem is not so much a lack of realism, but a missing cognitive link between the imagined and the actual, a sense that in one way or another, what could happen there could also happen here (see Atwood, 2011). Such a link can be rather abstract or 'on the nose', but it is only through *jolts of recognition* – e.g., engendered by historical, cultural, or technical references and associations – that fictional tales acquire a deeper meaning and purpose, effectively generating "other worlds as a comment upon our own" (Miner, 1991: 150). As

Atwood (2005b: 158-159) emphasizes, "The fictional world so lovingly delineated by the writer may bear a more obvious or a less obvious relation to the world we actually live in, but bearing no relation to it at all is not an option." Ultimately, it is by drawing such connections that speculative fiction "take[s] part in the discourses of the contemporariness" (Kuźnicki, 2017: 18), raising questions "not only about what might happen, but also about what is happening" (Miner, 1991: 150).

Given this focus on plausible stories grounded, at least to some extent, in human history and experience, it should come as no surprise that the genre often addresses issues quite similar to those of real-life society: the unequal distribution of power, wealth, and privilege; the violation of civil rights and liberties under systems of oppression; racism, sexism, and other forms of social discrimination; environmental change and human interference with nature; the potential risks and harms of techno-scientific innovation, to name but a few. Yet, rather than directly dealing with the here and now, speculative fiction explores these issues in alternative realities, pushing them to extremes in "strange but imaginable" future contexts (see Onyett, 2016: 62). Such visions can then serve as "dire warnings […], dark shadows cast by the present into the future" (Atwood, 2005b: 94). They are, figuratively speaking, "what will happen if we don't pull up our socks" (ibid.). By showing what it might be like to live in such a universe, speculative novels may stimulate reflection and promote critical thinking, providing "rich and complex avenues for reading and rereading the world" (Thomas, 2013: 4). What is important here is the interaction between the narrated story and the reader. As de Smedt and de Cruz (2015) argue, speculative fiction's epistemic value lies in its ability to present fictional worlds in great detail, immersing the reader in an altered reality and allowing her to become emotionally invested in a concrete situation or event. Furthermore, by showing new technologies fully operational and graphically depicting what changes in social organization could mean for those affected, the genre can contribute to a better appraisal of the actual consequences such hypothetical situations may have, should they ever occur. It is for these qualities that speculative text has been lauded as a valuable resource for critical pedagogy (see Patch, 2014), public deliberation (see Milkoreit, 2017), but also technology assessment (see Miller, 2015), providing opportunities for developing a sense of "informed skepticism" while exploring the "social, political, and ethical implications of science [and technology] that are so powerfully revealed through story" (Svec & Winiski, 2013: 37).

The notion of speculative fiction as outlined above has guided our literature review in several important ways. On the one hand, it has prompted us to discard any fictional stories that bear no or hardly any resemblance to the world and universe as we currently know it. Thereby, the cut-off point was not whether the story features space travel and aliens or not – a distinction Atwood makes when differentiating between sci-fi proper and speculative fiction (see Atwood, 2011: 115) – but rather the degree to which the topics and issues addressed connect to our own history and culture. For instance, despite its space-age setting, it is the thematization of mathematical prediction and science as a religion that makes Isaac Asimov's (2010) *Foundation* trilogy such a striking illustration of today's Big Data-invigorated social physics movement (see Pentland, 2014; Strogatz, 2003). Thus, while lacking in sociological realism, the series offers a quintessential portrayal of an enduring positivist dream[3] – i.e., to uncover the invariant, 'natural' laws of collective human behavior (see Ball, 2001) – providing an imaginative take on a field and vision that has gained considerable traction in recent years. On the other hand, the flexibility of the genre has also allowed us to include fictions that are so close to home they can hardly be considered sci-fi. Examples for this are Wilson's (2015) novel *The Affinities*, Egger's (2013) tech thriller *The Circle*, or Liu's (2012) short story *The Perfect Match*. None of these titles include technologies that either have not already been realized or else could not conceivably be realized at any given moment. Instead, they ponder what it means to be human in ever more tech-regulated environments, creating unsettlingly familiar sceneries and then "pushing the envelope as far as it will go" (Atwood, 2005a). What can be observed in such cases is a progressing conflation of fiction and reality, sometimes to the point where the fiction – even if recent – appears to fall behind the actual pace of techno-cultural change. This does not, however, imply that fictional tales are becoming irrelevant as tools for critical assessment and reflection, but rather, as Atwood (ibid.) argues, that exploring the imagination must no longer be considered a pastime, but a necessity, because "increasingly, if we can imagine it, we'll be able to do it."

Subsequently, this paper provides insight into what can broadly be subsumed under the label of "datafictions". After some conceptual clarifications (Section 3), we present a selective overview of works of fiction that may contribute to a deeper engagement with the logics and politics of spreading algorithmic regulation (Section 4). While by no means

[3] Others have referred to it as the "Cartesian dream" (see Pereira and Funtowicz, 2015).

comprehensive or conclusive, this review hopes to serve as a starting point and guide for researchers, educators, and readers who wish to delve further into the subject matter.

3. Datafictions: Conceptual and Methodological Considerations

Whereas speculative fiction functions as a basic framing for the kind of literature we are interested in, the notion of "datafictions" further specifies our thematic interest. In short, we propose datafictions as an umbrella term for speculative stories that deal with the datafication of society in both imaginative and imaginable ways. While this definition is fairly broad, so is the range of topics and issues commonly addressed: questions of surveillance and control, the use of predictive analytics, the emergence of new reputation economies, the intricate politics of computer-curated information, the development of sophisticated AI systems, or society's growing dependence on technology and machines are but some of the subgenre's most prevalent themes. Many of these stories are bleak and dystopian, depicting the oppressing force of all-knowing totalitarian regimes, while others are more satirical and playful, commenting, for instance, on people's tendency to overshare personal information or to rank and rate anything that can possibly be rated and ranked. At their core, however, most stories are quite serious interrogations of techno-cultural dynamics and trends, inviting the reader not only to consider but to experience possible future scenarios through a protagonist's eyes and actions. Thereby, questions of power, autonomy, and identity may become much more tangible and 'real', effectively replacing abstract concepts with something that can feel uncomfortably close and personal. Narrative fiction may thus prove fertile ground for a more critical appreciation of data as a source of social order and control, their collection and analysis a key instrument of modern population management. What datafictions can then ultimately provide are vivid accounts of data power in action – the concrete practices, the laws and ethics, the immediate but also the long-term societal repercussions. Although imagined, these accounts may serve as a basis for further reflection, essentially acting as a diffracted window to our own culture and time.

Our search for relevant works of fiction included a reassessment of well-known genre classics; regular perusal of the culture and literature sections of *The New York Times*, *The Guardian*, *ArsTechnica*, and *The Verge*; reading through related discussions on various blogs and online forums; and personal conversations with scholars familiar with the subject matter.

Apart from fitting into the framework of speculative fiction outlined above, the main selection criterion was whether the problems and issues addressed in the story are pertinent to current debates about Big Data and algorithmic decision making. In particular, we sought to focus on literary and audio-visual texts – i.e., novels, short stories, films, and television shows – that examine the impact of technology on people's lives, the formative power these technologies can have, but also potential forms of resistance, if only to show their utter futility.[4] All in all, we present and briefly discuss fifteen datafictions, most of them of a decidedly dystopian, techno-nightmarish nature. As indicated, this overview is neither exhaustive nor definitive, but rather seeks to provide some initial orientation and impetus for further exploration.

There are two notable omissions in our selection. First, though extremely popular, we are excluding stories about artificial intelligence (AI) if AI is discussed in terms of machine consciousness and linked to ethical debates about robot rights and human-like emotions. Why so? On the one hand, while existing computational models have successfully captured a number of cognitive and behavioral correlates of conscious information processing, to date, "no existing approach to artificial consciousness has presented a compelling demonstration of phenomenal machine consciousness, or even clear evidence that phenomenal consciousness will eventually be possible" (Reggia, 2013: 112). Given this, many fictional treatments of AI – think of the HBO series *Westworld* (2016), the sci-fi thriller *Ex Machina* (2014), or the robot drama *A.I. Artificial Intelligence* (2001) – reside in the realm of fantasy rather than plausible speculation. On the other hand, we would argue that the focus on consciousness – regardless of whether rationally conceivable or not – diverts attention from other AI-related issues that are already evident, and arguably much more pressing. As one news headline poignantly put it, "Intelligent robots don't need consciousness to turn against us" (Del Prado, 2015). A second, related omission concerns fictions of an oncoming singularity depicted as a runaway, self-improving superintelligence that far exceeds the human mind and intellect. More often than not, these rogue systems then strive to either enslave – as seen in, for example, *I, Robot* (2004), *The Matrix* (1999), *Colossus: The Forbin Project (1970)* – or eradicate – e.g. , *Robopocalypse* (Wilson, 2014)*, Computer One* (Collins, 1993), *The Terminator* (1984) – humanity, no longer following the will of their increasingly obsolescent creators.

[4] In fact, many stories – after allowing for initial bursts of hope and optimism – eventually lead to the pessimistic conclusion that the reigning regime cannot be weakened or overthrown and that resistance, under Big Brother's watchful gaze, is indeed futile.

While such visions might make for an entertaining plot, AI experts have argued that they rest on a "profound misunderstanding of both the nature of intelligence and the behavior of recursively self-augmenting systems" (Chollet, 2017), stressing that the development of artificial general intelligence (AGI) remains a far-off pipe dream. A more realistic scenario, one depicted in many of the datafictions mentioned below, is the continued proliferation of comparatively narrow, specialized AI into virtually all corners of society. Importantly, however, these imagined AI systems are seldom rogue or beyond human control, but serve specific purposes driven by certain government or corporate interests and objectives. Neither conscious nor ultrasmart, the technologies showcased in the selected fictions are often surprisingly mundane, concurring with Atwood's aforementioned warning: "We've done it, or we're doing it, or we could start doing it tomorrow" (2005b: 92).

4. Datafictions: A Selective Overview

4.1. Surveillance

The rise of Big Data has been accompanied by renewed interest in a classic of dystopian fiction – George Orwell's (1987 [1949]) novel *Nineteen Eighty-Four*. In a world where Big Brother watches one's every move, private life has effectively come to an end. Through always-on telescreens and hidden microphones, the one-party government has turned society into an electronic panopticon: "[a]sleep or awake, working or eating, indoors or out of doors, in the bath or in bed – no escape" (ibid.: 29). While Orwell's vision also connects to recent debates around fake news and alternative facts (see Andrews, 2017), in the main, it provides a chilling depiction of life under constant surveillance, showcasing possible consequences such as the loss of individuality in a culture marked by distrust, emotional solitude, self-censorship and enforced conformity. What is more, the book offers a haunting illustration of the use of technological progress for the diminution of human liberty, sensitizing the reader to the interrelationship between (ICT) technology, surveillance, and social control. Consider the following excerpt indicating how intimate and fine-grained such tech-enabled surveillance in Orwell's post-World War II fiction has already become: "He […] pushed back his chair, so as to get as far away from the telescreen as possible. To keep your face expressionless was not difficult, and even your breathing could be controlled, with an effort: but you could not

control the beating of your heart, and the telescreen was quite delicate enough to pick it up" (Orwell, 1987: 82).

Moving from state to corporate surveillance, Dave Eggers' (2013) novel *The Circle* chronicles the rise of a data-hungry tech company and its rapid growth in influence and power. As a Google-Facebook-Twitter hybrid that has already devoured most of its competitors, the eponymous Circle seeks "full access to all data about every person" (ibid.: 483), employing a growing variety of sensors and services to create a society where "secrets are crimes" (ibid.), "privacy is theft" (ibid.: 303), and "all that happens must be known" (ibid.: 67). While the novel delves into the social and psychological consequences of extreme metrification and social media pressure, it also considers the effects of privately-controlled transparency on democracy, showcasing how the quest for improved political accountability can lead to an utter totalitarian nightmare. In general, it is this emphasis on the recurring gap between good intentions and bad solutions that makes *The Circle* an intriguing, provocative read: Child abduction is a horrible crime, but does this warrant the implantation of miniature tracking chips into each and every kid? More data may allow for better healthcare, but does this sanction a mandatory, round-the-clock monitoring of one's personal life? Back-room deals can undermine people's trust in public officials, but does this mean that officeholders should be forced to 'go clear', live streaming all of their meetings and conversations to an unrestricted public audience? What the Circle's followers dream of is a "new glorious openness, a world of perpetual light" (ibid.: 491); what they help create, however, is "the world's first tyrannical monopoly", a "private company [controlling] the flow of all information" (ibid.: 401).

Another dark vision of pervasive corporate surveillance is outlined in Ken Liu's (2012) short story *The Perfect Match*. In a world not too far from our own, people largely rely on personal AI assistants to manage their lives – what to eat, where to shop, how to relax, whom to date, Centillion's proprietary algorithms "always know best" (ibid.). The reason the software can provide such spot-on recommendations is its unrestricted access to personal information, including users' social media profiles, their complete search and purchase histories, any reviews they may give or receive, location data, audio and video streams, etcetera. All of these data are then used to compile user profiles, train personalized learning algorithms, and nudge people into commercial transactions, with the aim of boosting advertising revenue. What Liu's story highlights, however, is that most people give up their

data voluntarily, either unaware of the end user agreement terms and conditions, or else willingly accepting a loss of privacy and autonomy in exchange for better products and services. The life Liu describes is one filtered through the eyes of the machine, turning people into "docile", "obedient puppets" who have forgotten how to think and judge for themselves: "Look at you. You've agreed to have cameras observe your every move, to have every thought, word, interaction recorded in some distant data center so that algorithms could be run over them, mining them for data that marketers pay for. Now you've got nothing left that's private, nothing that's yours and yours alone. Centillion owns all of you. […] You buy what Centillion wants you to buy; you read what Centillion suggests you read; you date who Centillion thinks you should date. But are you really happy?" (ibid.).

4.2. Social Sorting

While processes of social sorting already play a role in the surveillance dystopias mentioned above, they figure more prominently in other fictions, a first example being Aldous Huxley's (2006 [1932]) canonical novel *Brave New World*. Whereas Orwell's *Nineteen Eighty-Four* envisions a violent totalitarian dictatorship, Huxley depicts a subtler, more mechanical system of social oppression. In a distant future, attempts to create an utmost efficient and stable society have led to an eradication of individuality and free will. People are bred into one of five castes, each chemically engineered and psychologically conditioned for specific, predefined tasks and duties: intellectually challenging jobs for smart-by-design Alphas, "semi-moron work" for artificially backward Epsilons – "the principle of mass production at last applied to biology" (ibid.: 7). Rather than forcefully suppressed, however, caste members are nudged into a life of self-indulgence and complacency, nurtured by a nonstop stream of moral propaganda and consumption-driving advertisement. In Huxley's vision, social order thus follows industry interests, the happy, goods-consuming citizen an effective solution to the need that the "machine turns, turns and must keep on turning" (ibid.: 42). Under-consumption, in turn, is seen as a "crime against society" (ibid.: 52), a failure to commit and conform to "what civilized people ought to do" (ibid.: 122). Sprinkled with a good dose of irony, what makes *Brave New World* a particularly commendable read is the attention devoted to the protagonists' emotional states as well as intense debates arguing the pros and cons of the imposed system, signifying a veritable culture clash. As the Savage, born and

raised in an Indian reservation, responds when reminded of the comfort the highly regulated life affords: "But I don't want comfort. I want God, I want poetry, I want real danger, I want freedom, I want goodness. I want sin. […] I'm claiming the right to be unhappy." (ibid.: 240)

The vision of a society ruled by measurements and metrics is also played out in the *Black Mirror* TV episode "Nosedive" (Wright, 2016). Using a centralized reputation system, the people in this fiction constantly rank and rate each other's behavior and social media feeds, the aggregated score becoming a public statement of one's status and standing within the community. A high score means not only social prestige, but also access to better jobs and exclusive discounts. A low score, on the other hand, can endanger one's very existence, threatening career prospects as well as acceptance in society. The episode pictures people's constant struggle to maintain their scores, the apparent need to create and hide behind a fake, always-smiling persona, and the relative ease with which carefully built reputations can be destroyed by a few unfortunate events. While the software technically enables this culture of vigorous grading, it is the people themselves who fuel the system's cruelty: Though friendly on the surface, they have turned into hypersensitive critics of their peers, obsessed with "play[ing] the numbers game" because "that's how the fucking world works" (ibid.). What is thus most interesting about "Nosedive" is not so much the details of the technology, but the story's inquiry into human nature, exposing our need to judge and compare, thereby always looking for approval and self-validation in the form of likes, upvotes, and 5-star ratings. And while the episode's cinematography paints the world and its people in bright pastel colors, it soon becomes clear that there is something very rotten about the state of this imagined civilization.

The theme of computerized social sorting is also explored in Robert C. Wilson's (2015) novel *The Affinities*. When the data-mining company InterAlia offers Affinity testing as a new service, the world starts to change. Based on a series of psychological assessments, people who qualify are placed in one of twenty-two Affinity groups where someone is "statistically more likely to trust others, to be trusted, to make friends, to find partners, in general to have successful social engagements" (ibid.: 15). While such groups can be physically, ethnically, and socially diverse, the members of an Affinity share certain character traits that facilitate interpersonal bonding, fellowship, and cooperation. Within but a few years, Affinity circles start to revolutionize societal dynamics: people distance themselves from their families and friends, preferring the company of their "tranchemates" instead; commercial networks

become increasingly tribal, with Affinity groups managing their own trusts and investments portfolios; political decision making, too, gets shaped by Affinity interests, a growing number of elected officials firmly in the pocket of powerful clan lobbies. What Wilson outlines is a post-national system of loyalty that shakes the very foundations of a society and culture. While also tackling questions of privatization and intellectual property, the book's central theme is that of algorithmic discrimination and social exclusion. Whereas acceptance to an Affinity group comes with a wide range of tangible benefits, those who do not qualify are left behind, locked out of the exclusive enclaves of their former peers: "You could put a hundred people together and they could live better, fuller, freer, happier, more collaborative lives – but only the *right* hundred people, not a hundred random people off the street. […] Not disputing it's nice inside, for anyone who can *get* inside. But think about what that means for all the people not included" (ibid.: 239f).

4.3. Prediction

The collection and statistical analysis of data is often driven by the goal to forecast trends and events. How likely is it that a particular recommendation will suit the recipient's tastes and preferences (*The Perfect Match*)? What are the probabilities that a person will fit into a specific group of likeminded people (*The Affinities*)? While predictive capacities play a role in many datafictions, there are some that treat the theme as a central subject matter. In Isaac Asimov's (2010 [1951, 1952, 1953]) *Foundation* trilogy, for instance, a scientist named Hari Seldon develops a branch of mathematics that allows him to accurately predict the future course of history. Having foretold the slow demise of the Galactic Empire, Seldon and his followers are cast into exile on a far-off planet where they attempt to preserve humanity's knowledge and rebuild the interstellar civilization. What is interesting is that only the first few chapters follow the actions of Seldon himself, the bulk of the books depicting crises and challenges of later generations, all perfectly predicted by the doctor's initial calculations. After a while, however, an unforeseen anomaly causes Seldon's predictions to fail, casting doubt on what was considered a proven destiny. The space saga then engages in a critical examination of a probabilistic science that assumes that "the laws of history are as absolute as the laws of physics" (ibid.: 283), pondering issues of blind trust, free will, and the limits of mathematical (pre)calculation. Importantly, however, Asimov's series is not a cautioning

dystopia, but a complex fictional treatment of the longstanding human aspiration to know ahead and plan in advance. And it is arguably this absence of overt moralizing that makes the series a particularly productive, albeit controversial (Krugman, 2012), discussion piece.

The theme of prediction and foresight also plays a central role in Philip K. Dick's (1991 [1956]) short story *The Minority Report*. In a fictional New York City, police forces make use of the Precrime system to detect and detain citizens who are prophesied to commit a crime in the near future, thereby "successfully abolish[ing] the postcrime punitive system of jails and fines" (ibid.: 72). As a result, felonies have been cut down by almost one hundred percent, creating an all but crimeless society in which acts of violence and aggression are largely unknown. But the practice of pre-emptive crime prevention raises some difficult questions: What are the rights of those incarcerated as the police is effectively "taking in individuals who have broken no law" (ibid.)? How accurate are the predictions, how secure is the system, and is there a danger of corruption? As a society, what do we value more – personal integrity or statistical safety? What Dick's thriller-like story furthermore outlines is the immense controlling power of the analytical machineries and those in charge of them. The situation is worsened as there seem to be hardly any accountability structures in place, the police and the military being the only organizations with access to the system's forecasts. Citizens are thus subjugated to a black-boxed security regime with no means to appeal or get information about the presumed future offense. Given the recent rise of predictive policing all over the world (Ferguson, 2017), Dick's *Minority Report* no longer seems like a distant fiction, but should rather be regarded as a very timely and relevant read.

Another example for society-wide pre-crime analysis can be found in the anime TV series *Psycho-Pass* (2012). In a dystopian 22nd-century Japan, people are constantly monitored by the Sibyl system, a super computer architecture that analyzes people's psychological profiles to determine their crime coefficient. If the estimated score is above a certain threshold, the person registers as a latent criminal with all rights being immediately revoked. The system then commands "enforcement action", essentially ordering authorities to eliminate the subject even if no crime has yet been committed. While visualizing the violent cruelty of the system in genre-typical extremes, the series offers thoughtful reflections on issues of responsibility in complex human-machine entanglements. As an officer concedes after nearly having shot a panicked hostage the Sibyl system deemed to be a dangerous threat: "I've been an enforcer for a long time now. [...] Following orders and taking down

preys engrained in me. My hands only know how to do what they are told, I've always obeyed […]. I've taken down many latent criminals. I did what was best for society, blindly accepting that idea. And one day I just stopped thinking about what I was doing. It all become perfunctory to me. As natural an act as breathing." (ibid., episode 2) Alongside more general questions of justice and due diligence, the series also focuses on the importance of public trust, stressing that the system's acceptance crucially depends on people's belief in its flawless working and perfect objectivity: "They promised a society managed by fair and impartial machine intelligence, a law free of petty human ego. People only accepted Sibyl because that's what they thought they were getting." (ibid., episode 17) Following the show's dark plot, it soon becomes clear that this is actually not quite what people were getting.

4.4. Advertising and Corporate Power

Another central topic is that of targeted advertising and hyper-commercialization, Egger's (2013) *The Circle*, Liu's (2012) *The Perfect Match*, Huxley's (2006 [1932]) *Brave New Word*, and *Black Mirror's* "Nosedive" episode (Wright, 2016) being but a few examples already mentioned above. A particularly well-known depiction of targeted commercial content is shown in Steven Spielberg's film *Minority Report* (2012) – a loose adaptation of Dick's same-named short story – where the main protagonist is personally addressed by surrounding billboards that seem to be able to detect his emotional state (extreme stress) and suggest various forms of relief (having a drink, going on vacation). The idea of highly personalized advertising is also explored in Thomas Sweterlitsch's (2014) novel *Tomorrow and Tomorrow*, a gory crime story taking place ten years after a nuclear terrorist attack has completely obliterated the city of Pittsburgh. People in this fiction have started to wear brain-wired augmented reality interfaces that project overlays onto supplementary retinal lenses. While these interfaces are used to watch video streams and revisit memories of the past, the system constantly tracks people's location and field of view, serving up ads that fit the current situation: "Waverly dips a biscuit into his cappuccino. Illy pitches espresso in the Adware – I consent and soon our waiter brings a fresh cup and biscotti [...]." (ibid.: 55) But the world of *Tomorrow and Tomorrow* is one defined by voyeuristic sensationalism and moral decline, the ads often linking to pornographic content matched to one's personal preferences: "I watch her a heartbeat too long after she returns to her game, her jersey dress rising with every

punch, and my Adware fills with pop-ups and redirect streams to escort services and live companions, to cam girls in lingerie who coo they want to meet me. […] 'I don't want any, I don't want them,' but the ads are better at knowing what I want than I am and ranks of girls march for my approval, all slight variations of Twiggy […]" (ibid.: 22).

While fictions such as *Tomorrow and Tomorrow* focus on people's individual experience of personalized adverts, Pohl and Kornbluth's (2003 [1953]) classic novel *The Space Merchants* features a critical portrayal of the advertising industry as such, and the substantial influence and power this sector can wield. The story is as follows: On an overpopulated planet Earth, ad agencies who have become the de facto rulers of modern society are looking to expand their customer base, following the credo "more people, more sales" (ibid.: 94f). But space on Earth is running out, with even the wealthy living in small, box-like apartments and both water and natural proteins being scarce, expensive resources. Marketing experts, though, have come up with a solution: To colonize Venus and turn the planet's inhabitable atmosphere into the home of a thriving, and thus exploitable, new civilization – an entire world to loot, gut, and plunder. To achieve this, however, people must be nudged into wanting to go to Venus, persuading prospective colonists but also society at large that "the grass is greener far away" (ibid.: 16). Against this background, the book's twist-rich plot deals with a range of issues: the ad industry's constant hunger to grow and expand; a professional ethos and ethics guided only by the proverbial "god of Sales" (ibid.: 8); the conflicting interests of producers and consumers in an economy geared toward profit maximization; a culture of lobbying where public officials are but puppets dancing to the strings of their corporate masters. *The Space Merchants* is not a recent publication, but in a world where the revenues of digital marketing companies dwarf the GDP of many nations (see Khanna, 2016), and where ad-selling social media platforms aim to connect ever more people to their version of the Internet (Shearlaw, 2016), the book seems more topical than ever.

The political influence of powerful tech corporations is also explored in Daniel F. Galouye's (1999 [1964]) novel *Simulacron-3*. In a society completely overrun by human pollsters, a company called Reactions, Inc. develops a computer simulation in which electronic avatars can be surveyed for their thoughts and opinions on advertised products. Since the simulated environment closely mimics that of its creators, the sampled reactions are indicative of a product's real-world performance: "Before we market a product, we want

to know who's going to buy it, how often, what they'll be willing to pay" (ibid.: 8). What soon becomes clear, however, is that the company's ambitions reach far beyond mere market research. First, by reprogramming the simulation from a market-focused to a politically-oriented environment, the company aims to support political leaders in their campaigns, telling them "which cards to play – in every national and local election and on every issue" (ibid.: 38). Second, using the simulator for political advice, the company's CEO eventually plans to assume leadership himself, replacing current office holders with an industry-friendly cabinet: "Hall, I think you're observant enough to know I'm a man of no small ambitions. […] What we want is the most capable national leadership available! Can you think of a bigger financial empire than the one I've created? Is there anyone more logically qualified to sit in the White House?" (ibid.: 56) Although published decades ago during the early years of the computer age, Galouye's book remains incredibly timely: Though the White House is not yet controlled by the CEO of a major tech company, media reports have documented the crucial role that consultancy firms have played in recent elections (Cadwalladr and Graham-Harrison, 2018), employing technology that is not too far from what is envisioned in Galouye's fiction. Given that *Simulacron-3* also addresses the issue of automation-caused mass unemployment, the book certainly should not be missing from any datafictions reading list.

4.5. Hubris, Breakdown, and The End of Big Data

Continuing with another classic of speculative fiction history, Edward M. Forster's short story *The Machine Stops* (2013 [1909]) envisions an extreme case of unmitigated technology dependence. In Forster's fiction, mankind has largely retreated to a subterranean habitat, with each individual residing in its own fully automated room, serviced by a giant machine structure. Direct personal interaction has become rare, with people mostly communicating through sound and image transmitting devices, unwilling to leave the comfort of their homes, where the machine provides them with any amenities needed – from food and clothing to medicine and hot baths to music and literature. In turn, though religion is generally frowned upon, people have started to worship the benevolent machine, gradually forgetting that it was man, not God, who once created it. Thus, while people have grown increasingly reliant on using and interacting with the machine, they no longer understand the apparatus as a whole, gradually losing the ability to control it: "We created the Machine, to do our will, but

we cannot make it do our will now. It has robbed us of the sense of space and of the sense of touch, it has blurred every human relation and narrowed down love to a carnal act, it has paralyzed our bodies and our wills, and now it compels us to worship it" (ibid.: 30). Ultimately, Forster's vision is a gloomy mediation on what is lost and left behind in the name of progress if thought of only in terms of technical advancement and the progress of the machine. It is a story about a decadent, complacent society that is swallowed by its own inventions: "Man, the flower of all flesh, the noblest of all creatures visible, man who had once made god in his image, and had mirrored his strength on the constellations, beautiful naked man was dying, strangled in the garments that he had woven" (ibid.: 39).

The theme of scientific hubris and out-of-control technology is also explored in Mary Shelley's (2012 [1818]) horror novel *Frankenstein*. After being forsaken by its creator and shunned by humanity, an unnamed artificial creature, kind-hearted and affectionate at first, turns into a murderous monstrosity, driven by feelings of hatred and revenge. While some chapters are told from the monster's perspective, the story mainly progresses through the eyes of its tormented creator, Victor Frankenstein, whose neglect and failure to assume responsibility greatly contribute to the creature's misery and wickedness. By pondering this father-child relationship, the tale raises questions that remain of crucial relevance in today's science and engineering culture: Are there any limits to what science and technology ought or ought not to do? What are the moral and ethical duties of a creator towards her invention and the possible effects this invention may have on society? Also, as for the most part it is only Frankenstein who is aware of the creature's evil deeds: What are the ill-effects of secrecy and non-disclosure? What structures of accountability could counteract such tendencies towards systematic concealment and opaque operations? Shelley's novel can thus be read as a call for prudence and caution, but also as a reminder that technologies, once unleashed upon society, must continually be watched and cared[5] for should we not wish to experience their broken-hearted wrath: "'Remember, that I am thy creature; I ought to be thy Adam, but I am rather the fallen angel, whom thou drivest from joy for no misdeed. […] I was benevolent and good; misery made me a fiend. Make me happy, and I shall again be virtuous.'" (ibid.: 85).

[5] For a more detailed discussion of *Frankenstein* and the significance of care, see Halpern et al. (2016). For an analysis of collective imaginations and practices of responsibility in collaborative knowledge production, see Völker (2018, forthcoming).

A final recommendation would be Bridle's (2016) short story *The End of Big Data*. In his text, Bridle sketches a world in which the collection and exchange of personal data has been banned by law. A sophisticated satellite system monitors earth, scanning infrastructure and transmissions to ensure that "all the mainframes that used to track and store every detail of [people's] lives are turned off, and stay off" (ibid.). This is necessary because a new black economy aims to profit from whatever data are left to scrape, trading user profiles to sustain the remnants of the once thriving advertising industry. The ban of personal data was caused by a complete breakdown of the Web's networks, followed by large-scale data breaches: social security numbers, passwords, emails, text messages, browsing histories, credit card transactions, contracts, medical records – all public and for everyone to see. Everyone became vulnerable and legislature decided to act: "Nothing identifying. No dossiers, no manila files, no cookies or patterns of life or digital signatures, nothing that could link anyone to anything." (ibid.) Bridle's story is a thought experiment envisioning the end of the digital world as we currently know it, paying particular attention to the sites and physical objects left behind: decommissioned data centers, disassembled routers and cooling vents, drums of redundant Ethernet cable, recycling crews that scrap computer parts for metals, minerals, and magnets. If a personal data repository is discovered, it will be destroyed, for the people in Bridle's tale have experienced and rebelled against its power: "Data is power. It's something to take from and hold over somebody else; quantified dominion. The more you have on someone, the more you have over them. The more personal it is, the more power, until you've eaten right through the skin of social relationships and into the flesh itself. The [personal data repository] is an ark of unqualified dominion." (ibid.)

5. Final Considerations

As indicated, the above overview of speculative datafictions is neither comprehensive nor conclusive. Rather, it is meant as an initial guide and orientation, an invitation to dig deeper and explore a growing body of literary and audio-visual texts relevant to the subject matter. What also needs to be noted is that the proposed sub-categories (surveillance, social sorting, prediction, etc.) should be treated as only a suggestion, as but one attempt to cluster works of fiction that are so varied and rich that they could be interpreted, linked, and grouped in many different ways. Thus, while in a sense we engage in the precarious process of "sorting

things out" (Bowker and Star, 1999), we seek to do so in an open and preliminary manner, providing a basis for further discussion rather than a definitive scheme or structure.

Questions of categorization aside, the overview should have conveyed a sense of the thematic breadth and scope covered by datafictions, but also of the ways these visions connect to issues around Big Data and algorithmic decision making. Going forward, we see two main application areas for this kind of sci-fi engagement: Frist, as often argued in the academic literature (e.g., see Berne and Schumer, 2005), narrative fiction can prove fertile ground for exploring social, ethical, and political issues around new and emerging technologies, thereby supporting educational efforts and the cultivation of reflexive capacities in students, scholars, and practitioners. As Burton and co-authors (2015: 35) explain, by "reframing recognizable human situations and problems in terms of unfamiliar settings and technology, science fiction […] can be an ideal medium for raising and exploring ethical concerns", making "visible the alarming and problematic aspects of a given situation that have become invisible in the mundane world because they have come to be regarded as ordinary or inevitable." From this perspective, works of fictions are primarily conceived as a tool for critical thinking, a way to draw people in and overcome the intellectual resistances some might feel towards the subject of ethics (see Pease, 2009). We concur with this perspective and would add that there is further need for shared collections and repositories that could facilitate the preparation of e.g. workshops and courses. The present article can be seen as an attempt to contribute to this effort. There is, however, a second area of application that tends to receive far less attention. Educational efforts aside, speculative fictions can also be used to examine how techno-scientific innovations are perceived at a given point in time as well as what is ultimately regarded as a desirable or undesirable future. Watching episodes of the British TV series *Black Mirror*, for instance, may give an impression of how specific technologies are culturally imagined and what kind of futures are considered plausible given current implementations and trajectories. In this regard, we recognize speculative fictions as valuable objects of research, conceptually no less legitimate and credible than other forms of future-oriented technology assessment such as mapping exercises, scenario development, or Delphi surveys. On the contrary, as a relatively unrestricted form of creative visioning, such narratives can expand the spectrum of what can reasonably be imagined, enabling comparative analysis between dominant and counter discourses (see Hajer, 2006), and possibly even serve as a necessary corrective to political and industry hyperbole (e.g., see

Rieder, 2018). In that sense, speculative fictions do not just imagine futures, but envision alternative paths and realities that can challenge the agenda-driven storylines of public-private stakeholders. We thus believe that a responsible research and innovation process should take into consideration the visions and worries put forth in speculative fictions, acknowledging the genre as a fruitful site for investigating broader cultural narratives and imaginations.

As mentioned, we hope that this paper marks only the beginning of more consistent academic interest in what we propose to call datafictions. The time seems right as processes of computerization and datafication continue to pose considerable social, ethical, and political challenges – from security to privacy to algorithmic bias and discrimination. There is no lack of relevant material: besides a proliferation of genre-specific publications, the fictional treatment of speculative data cultures has entered the mainstream. Jonathan Franzen's (2015) novel *Purity*, for example, can easily be regarded a datafiction from a 'serious', critically acclaimed author whose work would usually not be placed in, or even near, the science fiction genre. If one also takes into consideration that recent movies from the Marvel superhero universe feature visually striking examples of data-driven decision making – think of Project Insight in *Captain America: The Winter Solider* (2014) or Peter Parker's AI suite in *Spider-Man: Homecoming* (2017) – it becomes clear that datafictions no longer linger at the periphery of popular culture, but right at the center of modern society's collective imagination.

References

Adams, V., Murphy, M. and Clarke, A. E. (2009). Anticipation: Technoscience, Life, Affect, Temporality. *Subjectivity*, 28, pp. 246-265.

A.I. Artificial Intelligence. (2001). [film] USA: Steven Spielberg.

Andrews, M. A. (2017). Sales of Orwell's '1984' Spike After Kellyanne Conway's 'Alternative Facts'. *The Washington Post*. [online] Available at: http://wapo.st/2jdDYYk?tid=ss_tw&utm_term=.382224a84ac1 [Accessed 1 Nov. 2017].

Asimov, I. (2010 [1951, 1952, 1953]). *Foundation*. *Foundation and Empire*. *Second Foundation*. New York: Random House.

Atwood, M. (2011). *In Other Worlds. SF and the Human Imagination*. New York: Doubleday.

Atwood, M. (2005a). Aliens Have Taken the Place of Angels. *The Guardian*. [online] Available at: https://www.theguardian.com/film/2005/jun/17/sciencefictionfantasyandhorror.margaretatwood [Accessed 1 Nov. 2017].

Atwood, M. (2005b). *Writing with Intent. Essays, Reviews, Personal Prose: 1983-2005*. New York: Caroll & Graf Publishers.

Ball, P. (2002). The Physical Modelling of Society: A Historical Perspective. *Physica A: Statistical Mechanics and its Applications*, 314(1-4), pp. 1-14.

Barben, D., Fisher, E., Selin, C. and Guston, D. H. (2008). Anticipatory Governance and Nanotechnology: Foresight, Engagement, and Integration. In: E. J. Hackett, O. Amsterdamska, M. Lynch and J. Wajcman, eds., *The Handbook of Science and Technology Studies*, 3rd ed. Cambridge, MA: MIT Press, pp. 979-1000.

Bergstein, B. (2017). Fiction that Gets AI Right. What to Watch and Read Before the Robots Take Over. *MIT Technology Review*. [online] Available at: https://www.technologyreview.com/s/609131/fiction-that-gets-ai-right/ [Accessed 1 Nov. 2017].

Berne, R. W. and Schummer, J. (2005). Teaching Societal and Ethical Implications of Nanotechnology to Engineering Students Through Science Fiction. *Bulletin of Science, Technology & Society*, 25(6), pp. 459-468.

Bowker, G. C. and Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Bridle, J. (2016). The End of Big Data. *Motherboard*. [online] Available at: https://motherboard.vice.com/en_us/article/vv7ekm/the-end-of-big-data [Accessed 1 Nov. 2017].

Burton, E., Goldsmith, J. and Mattei N. (2015). Teaching AI Ethics Using Science Fiction. *Papers from the 2015 AAAI Workshop on Artificial Intelligence and Ethics*. [online] Palo Alto: AAAI Digital Library. Available at:

https://aaai.org/ocs/index.php/WS/AAAIW15/paper/view/10139 [Accessed 1 Nov. 2017].

Cadwalladr, C. and Graham-Harrison E. (2018). Revealed: 50 million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach. *The Guardian*. [online] Available at: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election [Accessed 1 Apr. 2018].

Callon, M., Lascoumes, P. and Barthe, Y. (2009 [2001]). *Acting in an Uncertain World: An Essay on Technical Democracy*. Cambridge, MA: The MIT Press.

Captain America: The Winter Soldier. (2014). [film] USA: Anthony Russo; John Russo.

Chollet, F. (2017). The Impossibility of Intelligence Explosion. *Medium*. [online] Available at: https://medium.com/@francois.chollet/the-impossibility-of-intelligence-explosion-5be4a9eda6ec [Accessed 20 June 2018].

Citron, D. K. and Pasquale F. (2014). The Scored Society: Due Process for Automated Decisions. *Washington Law Review*, 89(1), pp. 1-33.

Collingridge, D. (1980). *The Social Control of Technology*. London: Frances Pinter.

Collins, W. (1993). *Computer One*. London: No Exit Press.

Colossus: The Forbin Project. (1970). [film] USA: Joseph Sargent.

Del Prado, G. M. (2015). Intelligent Robots Don't Need to Be Conscious to Turn Against Us. *Business Insider*. [online] Available at: http://www.businessinsider.com/artificial-intelligence-machine-consciousness-expert-stuart-russell-future-ai-2015-7?r=US&IR=T&IR=T [Accessed 20 June 2018].

DLA Piper (2017). *Data Protection Laws of the World*. Handbook. [pdf] Available at: https://www.dlapiperdataprotection.com/system/modules/za.co.heliosdesign.dla.lotw.data_protection/functions/handbook.pdf?country=all [Accessed 1 Nov. 2017].

Dick, P. K. (1991 [1956]). The Minority Report. *The Collected Stories of Philip K. Dick. Volume 4*. New York: Citadel Press, pp. 71-102.

Egan, M. F. (1889). Book-Talk. *Lippincott's Monthly Magazine*, 44, Oct. 1889, pp. 597-598.

Eggers, D. (2013). *The Circle*. New York: Random House.

Ex Machina. (2014). [film] UK: Alex Garland.

Felt, U. (2015). Keeping Technologies Out: Sociotechnical Imaginaries and the Formation of a National Technopolitical Identity. In: S. Jasanoff and S.-H. Kim, eds., *Dreamscapes of Modernity. Sociotechnical Imaginaries and the Fabrication of Power*. Chicago: University of Chicago Press, pp. 103-125.

Ferguson, A. G. (2017). *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York: New York University Press.

Forster, E. M. (2013 [1909]). *The Machine Stops*. Rockville: Wildside Press.

Franzen, J. (2015). *Purity*. New York: Picador.

Galouye, D. F. (1999 [1964]). *Simulacron-3*. [Amazon Kindle Edition] Rockville: ArcManor.

Goodman, M. (2015). Chapter 4: You're Not the Customer, You're the Product. In: *Future Crimes. Inside the Digital Underground and the Battle for our Connected World*. New York: Anchor Books, pp. 54-80.

Hajer, M. A. (2006). The Living Institutions of the EU: Analysing Governance as Performance. *Perspectives on European Politics and Society*, 7(1), pp. 41-55.

Halpern, M. K., Sadowski, J., Eschrich, J., Finn, E. and Guston, D. H. (2016). Stitching Together Creativity and Responsibility: Interpreting Frankenstein Across Disciplines. *Bulletin of Science, Technology & Society*, 36(1), pp. 49-57.

Heinlein, R. A. (1949). March 4, 1949: Robert A. Heinlein to Lurton Blassingame. [letter] In: R. A. Heinlein and V. Heinlein, eds., (1989). *Grumbles from the Grave*. New York: Ballantine Books.

Huxley, A. (2006 [1932]). *Brave New World*. New York: HarperCollins.

I, Robot. (2004). [film] USA: Alex Proyas.

Jaruzelski, B., Staack, V. and Shinozaki, A. (2016). Global Innovation 1000: Software-as-a-Catalyst. *strategy+business*, Winter 2016, 85, pp. 85-105.

Karinen, R. and Guston, D. H. (2010). Toward Anticipatory Governance: The Experience with Nanotechnology. In: M. Kaiser, M. Kurath, S. Maasen and C. Rehmann-Sutter, eds., *Governing Future Technologies. Nanotechnology and the Rise of an Assessment Regime*. Dordrecht: Springer, pp. 217-232.

Khanna, P. (2016). These 25 Companies Are More Powerful than Many Countries. *Foreign Policy*. [online] Available at: http://foreignpolicy.com/2016/03/15/these-25-companies-are-more-powerful-than-many-countries-multinational-corporate-wealth-power/ [Accessed 1 Nov. 2017].

Krugman, P. (2012). Paul Krugman: Asimov's Foundation Novels Grounded my Economics. *The Guardian*. [online] Available at: https://www.theguardian.com/books/2012/dec/04/paul-krugman-asimov-economics [Accessed 1 Nov. 2017].

Kuźnicki, S. (2017). *Margaret Atwood's Dystopian Fiction. Fire Is Being Eaten*. Newcastle: Cambridge Scholars Publishing.

Le Guin, U. K. (2009). The Year of the Flood by Margeret Atwood. *The Guardian*. [online] Available at: https://www.theguardian.com/books/2009/aug/29/margaret-atwood-year-of-flood [Accessed 1 Nov. 2017].

Liu, K. (2012). The Perfect Match. *Lightspeed*, (31). [online] Available at: http://www.lightspeedmagazine.com/fiction/the-perfect-match/ [Accessed 1 Nov. 2017].

Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, Consequences, Critique. *Big Data & Society*, 1(2), pp. 1-13.

The Matrix. (1999). [film] USA: Lana Wachowski; Lilly Wachowski.

Mayer-Schönberger, V. and Cukier, K. (2013). *Big Data. A Revolution That Will Transform How We Live, Work, And Think*. New York: Houghton Mifflin Harcourt.

Michelson, E. S. (2016). *Assessing the Societal Implications of Emerging Technologies. Anticipatory Governance in Practice*. Abingdon; New York: Routledge.

Milkoreit, M. (2017). Imaginary Politics: Climate Change and Making the Future. *Elementa: Science of the Anthropocene*, 5(62), pp. 1-18.

Miller, R.-E. L. (2015). Impact Assessment and Science Fiction: Complementary Ways to Ask "What happens if…?" and the Delineation of a New Sub-genre. *Journal for Future Studies*, 20(2), pp. 105-112.

Miner, M. (1991). "Trust Me": Reading the Romance Plot in Margaret Atwood's The Handmaid's Tale. *Twentieth Century Literature*, 37(2), pp. 148-168.

Minority Report. (2002). [film] USA: Steven Spielberg.

Onyett, N. (2016). *Study and Revise for AS/A-level: The Handmaid's Tale by Margaret Atwood*. London: Hodder Education.

O'Neil, C. (2016). *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

Orwell, G. (1987 [1949]). *Nineteen Eighty-Four*. London: Penguin Books.

Pasquale, F. (2015). *The Black Box Society. The Secret Algorithms that Control Money and Information*. Cambridge, MA: Harvard University Press.

Patch, J. (2014). Science and Fiction. A Polemic on the Role of Imaginative Fiction in Civics and the Economy of Innovation. In: P. Paugh, T. Kress and R. Lake, eds., *Teaching towards Democracy with Postmodern and Popular Culture Texts*. Rotterdam: Sense Publishers, pp. 179-199.

Pease, A. (2009). Teaching Ethics with Science Fiction: A Case Study Syllabus. *Teaching Ethics*, 9(2), pp. 75-81.

Pentland, S. (2014). *Social Physics. How Good Ideas Spread – The Lessons from a New Science*. New York: The Penguin Press.

Pereira, Â. G. and Funtowicz, S., eds., (2015). *Science, Philosophy and Sustainability: The End of the Cartesian Dream*. London; New York: Routledge.

Pohl, F. and Kornbluth, C. M. (2003 [1953]). *The Space Merchants*. London: Orion Publishing.

Psycho-Pass. (2012). [television series] Japan: Tow Ubukata.

Reggia, J. A. (2013). The Rise of Machine Consciousness: Studying Consciousness with Computational Models. *Neural Networks*, 44, pp. 112-131.

Rieder, G. (2018). Tracing Big Data Imaginaries through Public Policy: The Case of the European Commission. In: A. R. Sætnan, I. Schneider and N. Green, eds., *The Politics and Policies of Big Data: Big Data, Big Brother?* New York; London: Routledge, pp. 89-109.

Shearlaw, M. (2016). Facebook Lures Africa with Free Internet – But What Is the Hidden Cost? *The Guardian*. [online] Available at: https://www.theguardian.com/world/2016/aug/01/facebook-free-basics-internet-africa-mark-zuckerberg [Accessed 1 Nov. 2017].

Shelley, M. (2012 [1818]). *Frankenstein*. New York: Barnes & Noble.

de Smedt, J. and de Cruz, H. (2015). The Epistemic Value of Speculative Fiction. *Midwest Studies in Philosophy*, 39(1), pp. 58-77.

Spider-Man: Homecoming. (2017). [film] USA: Jon Watts.

Strogatz, S. (2003). The Physics of Crowds. *Nature*, (428), pp. 367-368. [online] Available at: https://www.nature.com/articles/428367a [Accessed 1 Nov. 2017].

Sveck, M. and Winiski, M. (2013). SF and Speculative Novels. Confronting the Science and the Fiction. In: P. L. Thomas, ed., *Science Fiction and Speculative Fiction. Challenging Genres*. Rotterdam: Sense Publishers, pp. 34-57.

The Terminator. (1984). [film] UK; USA: James Cameron.

Thomas, P. L. (2013). Introduction Science Fiction and Speculative Fiction. In: P. L. Thomas, ed., *Science Fiction and Speculative Fiction. Challenging Genres*. Rotterdam: Sense Publishers, pp. 1-13.

Toffler, A. (1970). *Future Shock*. New York: Random House.

Völker, T. (2018, forthcoming). Re-ordering Science-Society Relations and Re-distributing Responsibility? On the Participatory Production and Circulation of Anticipatory Knowledge in Transdisciplinary Sustainability Research. In: A. Delicado, F. Crettaz Von Roten, A. Tupsala and I. van der Weijden, eds., *Science and Society Relations Today: A Sociological Gaze*. New York: Springer.

Westworld. (2016). [television series] USA: Bad Robot; Jerry Weintraub Productions; Kilter Films; Warner Bros. Television.

Wilson, D. H. (2011). *Robopocalypse*. New York: Doubleday.

Wilson, R. C. (2015). *The Affinities*. New York: Tom Doherty Associates.

Wright, J., dir., (2016). "Nosedive", *Black Mirror*, UK: Netflix.

CURRICULUM VITAE

Gernot Rieder

born July 12, 1984
in Klagenfurt, Austria

Email: gernot.rieder@univie.ac.at
Website: https://itu.academia.edu/gernotrieder

EDUCATION AND DIPLOMAS

-- 2015 – 2018	PhD Fellow, IT University of Copenhagen, Technologies in Practices Research Group
-- 2013	Master's Degree in Science-Technology-Society (with distinction), University of Vienna, Department of Science and Technology Studies
-- 2013	Bachelor's Degree in Theatre, Film and Media Studies (with distinction), University of Vienna, Department of Theatre, Film, and Media Studies
-- 2007	Bachelor's Degree in Journalism and Communication Studies (with distinction), University of Vienna, Department of Communication
-- 1994 – 2002	Secondary Education at the BG/BRG Ingeborg Bachmann in Klagenfurt, Austria
-- 1990 – 1994	Elementary School in Klagenfurt, Austria

SCHOLARSHIPS AND GRANTS

-- 2015 – 2018	Tuition Free Stipend, IT University of Copenhagen
-- 2011 – 2012	Research Scholarship, Austrian Science Fund
-- 2008	Scholarship of Excellence, University of Vienna
-- 2007	Scholarship of Excellence, University of Vienna

THESES

-- Doctoral Thesis	"Big Data: Or, The Vision That Would Not Fade" Copenhagen (2018), Advisor: Judith Simon
-- Master Thesis	"Making Futures Public: On the Modalities and Intricacies of Qualitative Social Science Nano Research" Vienna (2013), Advisor: Ulrike Felt
-- Bachelor Thesis	"Dimensionen einer Mensch-Computer Interaktion. Bildhafte Darstellung und symbolische Vermittlung als bezeichnende Elemente einer modernen Interface Kultur" Vienna (2007), Advisor: Frank Hartmann
-- Bachelor Thesis	"Internet Governance oder: Wer regiert den Cyberspace? Eine Darstellung der Entwicklung der administrativen Verwaltung des Internet und die aktuelle Diskussion zur Problematik der Domainvergabe" Vienna (2006), Advisor: Erich Geretschläger

TEACHING

-- Spring 2017	Co-Lecturer: *Big Data Processes*, IT University of Copenhagen
-- Spring 2016	Co-Lecturer: *Big Data Processes*, IT University of Copenhagen
-- Spring 2015	Assistant Lecturer: *Reflections on IT*, IT University of Copenhagen
-- 2007 – 2008	Tutor at the Department of Communication Studies, University of Vienna (1 to 2 courses per semester), Topics (example): Interface Communication, Web 2.0, Science Utopia

LANGUAGE SKILLS

-- German	mother tongue
-- English	excellent, written and spoken
-- French	advanced
-- Mandarin	notions