IRIE
International Review of Information Ethics Vol. 24 (05/2016)

Vol. 24 (05/2016)

Ethics of Big Data

edited by Klaus Wiegerling, Michael Nerurkar, Christian Wadephuhl

Editor of this issue:

Prof. Dr. Klaus Wiegerling, Dr. Michael Nerurkar, Christian Wadephuhl

Karlsruhe Institute of Technology, Germany

Institute for Technology Assessment and Systems Analysis (ITAS)

Editors of IRIE

  ISSN 1614-1687

Vol. 24 (05/2016)

Content:

  ISSN 1614-1687

Editorial: On IRIE Vol. 24

Does 'big' in the notion of Big Data also imply new' – a qualitative 'new'? Is 'Big Data' a legitimate new area of dedicated ethical analysis or do the rules of the classical 'Ethics of Data' simply apply to the 'Ethics of Big Data' as well?

Hegel has argued in his 'Logic' that from a certain point quantitative change leads to a qualitative shift. The prime example is water heated up vaporizing at 100 degrees centigrade thus becoming steam. Without indulging into the discussion of Hegel's concept especially not in the dialectic reception of Marx the question remains: is or from which point on is 'big' also 'new'? For a first approach let's see how big 'big' is:

- In 1995 the world wide web was considered to contain some 30 GB of data. That would fit on a regular USB stick today.
- In 2007 the total amount of data stored was calculated to 295 exa bytes. If we were to take all that information and store it in books, we could cover the entire area of the US or China in 13 layers of books.
- Facebook today stores, accesses, and analyses some 30+ Petabytes of user generated data i.e. thousand times more data than stored in the U.S. Library of Congress
- And the amount of data generated will continue to double every 2 years leading to a production of 40 zeta bytes of new data per year in 2020. If there was a star for every byte of data there would be a galaxy of data for every sand corn on the planet by then.

This technological development grows faster than everything else humanity has faced before and its social impact will be massive as well as pervasive reaching deeply into our daily life. So let us not investigate if the Hegelian concept applies. Let us analyze thoroughly how we shall handle this significant phenomenon from an ethical perspective. The editors of this issue have drawn together some very good suggestions. We thank them for thus contributing not only to the vast ocean of data produced in 2016 but also to the very discussion of its ethical relevance for us. We hope you appreciate it.

Sincerely yours,

the editors of IRIE

ISSN 1614-1687

Michael Nerurkar, Christian Wadephul, Klaus Wiegerling:

Ethics of Big Data: Introduction

Big Data has been one of the trending ideas in the information technology business in the recent years. It has (or already had) its place in the Gartner Hype Cycle for Emerging Technologies[1] and the phrase is notorious for oscillating between being used as a buzz word and being a meaningful technological concept. There is some broader interest in Big Data, as the many attempts to give an answer to the question "What is Big Data?" indicate.[2] The idea that when speaking of Big Data we are dealing with more than just one of the current coins in marketing jargon is also supported by the fact that the term has been admitted into venerable dictionaries like the Oxford English Dictionary and Merriam-Webster. The widespread interest in the nature of Big Data, though, also hints that there lie some difficulties in giving a concise and informative definition of the concept. This, of course, is not a distinctive feature of Big Data technologies, but is the case with modern technologies in general, especially with information and communications technologies: They evolve rapidly and sometimes are subject to technological leaps or redefinitions that transcend concepts in use at a time (the computer becoming a computer in a network, the mobile phone becoming the smart phone, glasses becoming augmented reality glasses, the automobile becoming the driverless car etc.).

So what is Big Data?[3] Big Data is very large volumes of data of various types, collected massively from heterogenous sources (sensors, cell phones, social networks etc.). Big Data is also the new technologies (algorithms/software and hardware) used to collect, store and process this data with high degrees of velocity. Technology forecasters and marketers envision Big Data to allow for entirely new kinds of applications and products to be developed in the field of information and communications. Big Data is marketed as a solution to many problems in informatics that could not be tackled before, or only with high costs and effort. Big Data is said to promise an increase in efficiency and productivity, to lower costs in industry and business, to enable new methods and knowledge in science, and to enhance control and regulation in personal life and in governance.

As regards those technologies as such it could seem that only technical disciplines would take an academic interest in them: It is a question for informatics whether envisioned Big Data applications actually are realizable, or whether they offer truly new approaches and solutions to classical problems. And it is a question for economics whether, for example, it would be economically rational to make use of Big Data technologies in business processes, or whether products built on them might find their markets. As regards ethics and the humanities it might seem not clear at first how Big Data technologies should actually raise any specifically *new* questions to think about. Are not problems of misuse and of harmful consequences of the use of technologies simply problems for a general ethics or an ethics of technology and lie in the responsibilities of individuals, public discourse and politics? Why, therefore, an "Ethics of Big Data"?

Outside of the technical literature the phrase "Big Data" (like, for example, its close associate "algorithm") is used in the media and a broader public to refer in a more or less technically informed way to "some technologies" that are more and more becoming part and shaping factor of our everyday life and that are thought to be potentially disruptive. Technology foresight scenarios give the impression that Big Data technologies can not only support or improve specific processes and activities but that they also enable new levels of regulation and control of individual, social, political und economical affairs and decisions. In this respect, "Big Data" would mean technologies that are socially contextualised and socially effective to an

[1] See the 2013, 2014 and 2015 versions of Gartner's Hype Cycle for Emerging Technologies: http://www.gartner.com/newsroom/id/2575515, http://www.gartner.com/newsroom/id/2819918, and http://www.gartner.com/newsroom/id/3114217.

[2] See for example http://datascience.berkeley.edu/what-is-big-data/.

[3] This is, of course, also not meant to be a definition of the term "Big Data" since it clearly is only very vague (regarding the technologies in question) and is invoking terms ("large", "new", "high degree" etc.) that are meaningful only relative to a contemporary state of the art. We are only trying to sum up here what is circulating in the non technical literature and on the web about the nature of Big Data.

ISSN 1614-1687

unusually high degree. A major effect (or benefit) of Big Data (as of any technology in general) might be to make things easier, more efficient, more rational, more economical etc. But the deployment and use of technologies currently referred to as "Big Data" could also cause a fundamental shift in our understanding of ourselves and our individual, social, cultural and political life. This invasive feature of Big Data should be subject to public, political and academic discourses. In this sense, "Big Data" as in "Ethics of Big Data" serves as an integrative concept. It does not express one single strict technological definition, but functions to integrate many different perspectives on a bigger phenomenon or paradigm shift that is technological in a narrower sense only at its core, but social and cultural in its impact. This throws up problems and questions in the various perspectives of disciplines like ethics, philosophy, epistemology, psychology, sociology, politics, jurisprudence or economics: Questions that are "ethical" in a wide sense, or "practical" in a classical sense: For example, autonomy and choice are concerned where supposedly highly accurate knowledge about persons "mined" from large sets of personal data leads to systems anticipating and pre-acting on our (presumed) desires, and "liberating" us from having to make our own decisions. Privacy is in question as regards how and with what amount of informed consent such individual data are collected and personality profiles are generated, stored, sold, made accessible to government agencies etc. Privacy is itself a necessary condition of autonomy since only when a private realm is guaranteed, and felt and known to be safe and free from external observation and intervention, one can engage in truly free processes of deliberation and decision making. Political freedom is at stake when government agencies practice data retention to profile citizens under the prospect of more security or more efficient regulation.

When talking about "data" one is led quickly to also talk about "information" and "knowledge", furthermore of "theory" and "science". In this regard, Big Data throws up not only ethical but epistemological and philosophical questions as well – to name just a few: What is data, ontologically? What is the relation between data, information, knowledge and theory, and how does one lead to the other? What sense is in speaking of data as "raw material" and of "mining" it? Big Data comes with claims to producing "better" knowledge, sometimes even with visions of making obsolete scientific models and theories, which clearly has epistemological and metaphysical implications since it purports to give us direct and "undistorted" access to reality. Those are claims that deserve thorough epistemological examination and critique. With Big Data we are given yet another label for self description of modern western type civilizations of the 21st century: Is it actually more appropriate to speak of "data societies" instead of "information societies" or of "knowledge societies"? – The contributions to this issue, however, are primarily dedicated to ethical/practical aspects of Big Data, though partly questions of theoretical/epistemological nature are also discussed.

Concerning the dominant optical metaphors within the Big Data discourse *Regine Buschauer* suggests a conceptual approach that allows clearer distinctions between Big Data as "visions" and as data technologies. Buschauer outlines three perspectives on the nexus between data and vision(s). Following Bruno Latour's counter-image of "oligoptica" Buschauer argues, more generally, in favor of a conceptual framework that understands Big Data as a sociotechnical infrastructure. Drawing on more recent studies, she discusses how such an approach allows to address social and ethical implications of present data technologies and practices in a more differentiated way.

In her contribution *Jessica Heesen* aims to explain, on one side, how a normative concept of the public sphere could be infiltrated by Big Data. Furthermore, she discusses how, on the other side, participative processes and common wealth can profit from a thorough use of Big Data analysis. Heesen introduces two important concepts: the numerical public (as a public that is constituted by machine-communication) and total politicisation (as a loss of negative freedom of expression).

Arne Manzeschke, Galia Assadi and *Willy Viehöver* consider Big Data to be a new form and instrument of biopolitics addressing both the categories of body and space. It is expected to fundamentally transform health care systems, domestic environments, and practices of self-observation and -reflection. Thus, the paper points out some problems and pitfalls as well as open questions that have emerged in the field of Ambient Assisted Living and that merit more attention in public and academic discourses.

Harald Weston asks whether predictive analytics enables businesses for purposes of lending money or extending credit, for example, to better assess characters of people beyond current objective measures of

credit scores and standard financial metrics. Can individual character be measured and predicted with predictive analytics? The pervasive data surveillance of people that goes with Big Data and predictive analytics is not only an invasion of privacy in general, but an impairment of the aspect of privacy called autonomy that will constrict and alter a person's choices and development of self.

In their article, *Andreas Kaminski* and *Philipp Richter* do not ask for possible losses of individual privacy and freedom but rather what Big Data visions mean for the concepts of self and human being. If "being oneself" according to Martin Heidegger is to live within a critical distance to predefined options, how can I know for sure then that I am "truly myself" and not just imitating behavior of others? Can Big Data help eliminate subjective distortions and illusions about our selves?

In his contribution, *Bruno Gransche* argues that the results of Big Data Analysis can be seen as today's oracle. The uncertainty of the modern, complex world can be felt as a threat and lead to a demand for better foreknowledge. Part of the promise of Big Data is to grant better foreknowledge while omitting problems of scientific theory and modeling. Gransche discusses that people would simply have to believe in the results of such data anlysis which would make Big Data based outcomes a matter of faith.

We thank the authors for their contributions.

The Editors of this Volume:

Prof. Dr. Klaus Wiegerling / Dr. Michael Nerurkar / Christian Wadephul, M. A.:

- Institut für Systemanalyse und Technikfolgenabschätzung (ITAS), Karlsruher Institut für Technologie, Postfach 3640, 76021 Karlsruhe, Germany
- 💻 https://www.itas.kit.edu
- ✉ klaus.wiegerling@kit.edu / nerurkar@kit.edu / christian.wadephul@kit.edu
- ☎ +49-721-608-23839

Regine Buschauer:

Datavisions – On Panoptica, Oligoptica, and (Big) Data

Abstract:

In focusing on relations between data and vision and proposing to address big data in terms of currently dominant optical metaphors (and, quite literally, in terms of 'visions'), the paper makes a case for an approach that allows for clearer distinctions between big data as 'visions', and data technologies. (Re)assessing (present and past) notions and visions of panoptic data technologies, I outline three perspectives on the nexus between data and vision(s). Following Bruno Latour's counter-image of "oligoptica", I argue, more generally, in favour of a conceptual framework that understands big data as a sociotechnical infrastructure, and discuss, drawing on more recent studies, in which ways this approach allows to address social and ethical implications of present data technologies and practices in a more differentiated way.

Agenda:

Author:

Dr. Regine Buschauer

- ✉ rebus@bluewin.ch
- Relevant publications:
  - (Very) Nervous Systems. Big Mobile Data, in: R. Reichert, (Ed.), Big Data. Bielefeld: transcript 2014, pp. 405-436.
  - Mobile Räume. Bielefeld: transcript 2010, 361 p.

ISSN 1614-1687

Introduction

What do we talk about when we talk about big data? Amongst the growing number of attempts to define this keyword of today's sociotechnical landscape, "Big Data" has been aptly identified as a buzzword that refers to a field who's object is difficult to narrow down and incites the (re-)introduction of metaphors – from oil and mining to footprints and data shadows – and of visual concepts. The debates on big data, and on surveillance in particular, have given rise to renewed recourses to the concept of panopticism developed by Michel Foucault (1975) based on his discussion of Jeremy Bentham's "Panopticon" (Bentham [1791]). As Tom Boellstorff (2013) pointed out, the optical metaphor provided by Bentham's architectural model of an 'all-seeing' surveillance seems, indeed, "prescient when an NSA surveillance program can be code-named 'prism'". It resonates, moreover, with habitualized metaphors used in the debates that surround data technologies, including metaphors of transparency such as the 'glass human being' or the transparent "panopticized state" (Fenster 2010), which are enmeshed with the notion of the Panopticon. On the other hand, uses of Foucault's or Bentham's concept as metaphors of a 'data society' have been widely criticized, rejected or replaced by new metaphors. Next to the revisions or modifications that have been proposed with concepts such as the "synoptical" (Mathiesen 1997) or "post-panoptic" (Boyne 2000), they have been more fundamentally challenged by Bruno Latour and his metaphor of the "oligopticon", coined as a theoretical counter-image to the Panopticon (Latour 2005).

To address (big) data in terms of these optical metaphors and, quite literally, of 'visions', as I want to propose, allows to consider some underlying assumptions that often remain unquestioned. In what follows the notion of a panoptical technology will thus be taken as the starting point of a discussion of (past and present) visions and practices of data and "datafications" with a particular focus on the relation between these metaphors and digital media and data. My aim will not be to (re-)examine Foucault's or Bentham's concepts, or the theoretical and historical issues they raise,[4] but to focus on how these concepts translate into notions of digital technologies, and in approaching big data through the 'lens' (so to speak) of presently dominant metaphors and their nexus between data and vision. This approach draws on conceptual considerations grounded in media studies as well as on more recent interdisciplinary debates on (big) data (see e.g. Kitchin 2014, Reichert 2014). More generally, I argue that the exploration of ethical and societal implications of an emerging field of data technologies cannot be separated from the question of how this field is conceptually framed and presents itself as a sociotechnical field in practice.

I will sketch out this argument proposing three perspectives on the relations between this field of metaphors and data. From a historical perspective, I will discuss relations between vision and digital data made in past visions of "smart machines" at work and their association with panopticism since the 1980s, arguing for an approach that situates big data in a history of information technology and its visions. I will then focus on these visions and optical metaphors in terms of visibility and vision, drawing on Latour's arguments in favour of a counter-image. Further developing this discussion, I will propose a perspective that follows Rob Kitchin's use of Latour's term and shifts the attention from visions of big data towards a notion of data technologies as a sociotechnical infrastructure. Based on some more recent studies, I will discuss in which way this perspective allows to address social and ethical implications of present data technologies and practices in a more precise way and discuss some findings and salient aspects of present day debates.

Visions of Smart Machines

A recent publication explaining IBM's new vision of an "era of cognitive computing" is entitled "Smart Machines" (Kelly/Hamm 2013) and directly echoes the title of Shoshana Zuboff's study published almost thirty years ago on the "future of work and power", in which she noted the arrival of a "panoptic power of

[4] For a recent discussion of both Foucault's and Bentham's concepts of the Panopticon see e. g. Laval 2012.

information technology" in the working world.[5] Reconsidered at the light of a present 'era of smart machines', Zuboff's book provides an example of a "history of the future" (Carey/Quirk [1989]) and allows to address current visions of data technologies from the perspective of a history of visions of panoptic smart machines.

"Information Panopticons"

From the research she conducted on "smart machines" during the years 1981 to 1986 Zuboff concluded that digital information systems, "that translate, record, and display human behavior", could

"provide the computer age version of universal transparency with a degree of illumination that would have exceeded even Bentham's most outlandish fantasies. Such systems can become information panopticons that […] do not depend upon the physical arrangements of buildings or the laborious record keeping of industrial administration. They do not require the mutual presence of objects of observation. They do not even require the presence of an observer."[6]

This conclusion translates Bentham's model of the Panopticon – or "Inspection House"[7] – into a metaphor for transformations Zuboff observed in industries such as paper production, pharmaceuticals or finance. In contrast with their potential to give rise to a knowledge based 'informated' future, she argued – drawing on visions such as Daniel Bell's "Information Society" –, computerization and automation tended to generate "information panopticons". Above all they provided new ways to capture data on behaviour as, e.g., in the case of a production control system she observed in a paper mill: This newly introduced system – able to capture and store data every five seconds from across the production process – provided data from which to analyse not only the operations per se, but of the operator's performance as well. As Zuboff notes, it developed in practice into a panoptic medium 'by accident'. Another case drawn from her study concerns the communication system within a pharmaceutical company which developed from an informal communication platform of the company's research unit to a system increasingly geared towards management control, finally replaced by a more formal communication system. As Zuboff underlines, this shift was reflected by rather self-protecting or self-promoting uses of the new system "as a way of going 'on record'".[8] Such systems of – in today's terms – 'datafication' thus implied a "panoptic" scheme and an intimation to comply which Zuboff found to parallel Bentham's model of the Panopticon – the "counterpart" of Bentham's "central tower" being "a video screen".[9]

Discussing the rise of "Pervasive Computing" and the need for an "ethics of forgetting", Martin Dodge and Rob Kitchin (2007) have pointed out that capturing digital behaviour data at the workplace, under the form of, e.g., swipe cards or keystroke counts, has already been a widespread practice for a long time. Zuboff's study provides an account of these practices from the historical phase of the 'computerization of work' which she envisioned along the lines of two alternative futures (i.e. a future of opportunities, knowledge and learning, and a future of threats and enhanced control). Her study, in this respect, leads back to practices as well as to future visions of the 1980s, including such ideas of a new era of "smart machines", its promises and threats, that reappear today. Against this background, the notion of "panoptic" technologies, in Zuboff's study as well as often today, is associated to a kind of counter-image to the promises of an 'informated' future, as it were, and to a model of centralized control.

This notion, though, fails already in regard to Zuboff's observations, which include what she referred to as "horizontal visibility": Information systems could allow individuals "to see their own behavior reflected in the

[5] Zuboff: In the Age of the Smart Machine, p. 322. On cognitive computing see also http://www.ibm.com/cognitive.

[6] Zuboff: In the Age of the Smart Machine, p. 322.

[7] Bentham's plan of the Panopticon (prototypically designed as a prison) envisioned single cells in a circular building around a central logde or tower from where a guard or "inspector" could oversee all of the cells without being seen; the "fundamental advantage", as Bentham puts it, being "the *apparent omnipresence* of the inspector (if divines will allow me the expression,) combined with the extreme facility of his *real presence*." Bentham: Panopticon, or the Inspection-House, p. 45.

[8] Zuboff: In the Age of the Smart Machine, p. 384.

[9] Ibid., p. 322.

system, while knowing that others (peers, subordinates, and superiors) could see it as well"; introducing a panopticon, "in which 'the many' view themselves and each views 'the other'".[10] Furthermore, she found that data shape organizations as a whole. An example is provided by a "cheating" case, concerning the work allocation and monitoring system of one of her study cases, in which the data was frequently manipulated (in order to meet the high performance expectations) in a practice that was shared across hierarchical levels. The point Zuboff noted as "striking" is that the system's data continued to be treated "as a legitimate reflection of the workplace". The system thus could "fulfill its role" irrespective of whether its data was known to be 'correct' or to be 'wrong'.[11]

Such examples, understood by Zuboff as reflecting an "evolution of the original concept", point towards issues that have been more broadly discussed in the light of increasingly pervasive technologies and of a rapidly emerging landscape of digital data and datafication. In Dodge and Kitchin's (2007) discussion of Pervasive Computing mentioned above, this change is referred to as a "shift from surveillance to sousveillance", drawing attention to the complementation of "external" data capture (in areas such as work, mobility, shopping or finance) by self-monitoring practices, and to the rise of a new and potentially panoptic "regime of recording".[12] Other authors have reassessed notions of "panoptic" technologies focusing more closely on the work sphere in a present "post-panopticism".

"Post-Panoptic[isms]"

"Post-panoptic" shifts have been identified in regard to changes in working environments related, e.g., to "networks of surveillance" – including social media and their connections to the sphere of work – and peer-oriented "lateral" surveillance (Brivot/Gendron 2011). They have also been related, more generally, to the potentially limitless (spatial and temporal) extensions of panoptic 'visibilities' granted, e.g., by mobile technologies. Summarizing their study of a knowledge management system in a large accounting company in the years 1999-2008, Marion Brivot and Yves Gendron stress the complexity of changes in which the "normalizing agenda" of the system was both circumvented and actively embraced while at the same time resulting in "lateral surveillance" among the employees. Similarly, in their study of mobile information systems in a smaller sized consulting company, Leclercq-Vandelannoitte et al. (2014) refer to "distributed control" and coin the term "free control", emphasizing how mobile connectivity, in the case they studied, was both shaped by and shaping a nonstop-working environment characterized by real-time control, a shared morale of coercive reactivity and performance expectations and a "new control logic, based on the total involvement of the professional".[13]

Such findings reflect contemporary 'visibilities' and a digital change that has indeed become pervasive, but that also points back to observations made by Zuboff; further examples e.g. of work-related data profiling and prediction practices add to this mixed picture.[14] What the above mentioned studies describe as "post-panoptic" or "beyond panopticism" is thus related to complex sociotechnical changes – expressed through keywords such as "bureaucratization", "normalization", or "self-monitoring" – and to a redefinition of "knowledge" as introduced by and framed within the confines of "knowledge management systems". These changes reinforce schemes of 'visibility', according to Brivot and Gendron who found that the "common denominator" of the changes studied was the "central role played by the notion of visibility – visibility to others, visibility to the self, as well as the fear or preoccupation of being visible or invisible to others".[15]

---

[10] Ibid., p. 350f.

[11] Ibid., p. 354.

[12] Dodge and Kitchin: 'Outlines of a world coming into existence', p. 432; 434.

[13] Leclercq-Vandelannoitte et al.: Mobile Information Systems and Organizational Control, p. 554.

[14] An example in this regard is the claimed ability to forecast whether an employee is inclined to look for a new job; see, e.g., "The Algorithm That Tells the Boss Who Might Quit", The Wall Street Journal, 13.3.2015; cf. Kitchin: The Data Revolution, p. 178.

[15] Brivot and Gendron: Beyond panopticism, p. 153.

From a theoretical point of view, it can be debated whether such findings indicate a "post-panopticism" particularly when one considers that "panopticism" is a term that arises from Foucault's discussion of Bentham and which already indicates a move 'beyond the Panopticon'. What is described in these studies indeed points back to issues also raised in Foucault's discussion (Foucault 1975). "Post-panopticism" may be considered more fundamentally as an ambiguous or even as a "dubious concept"[16], following Roy Boyne (2000) who introduced the term to denote the puzzling role played by panopticism he found in his survey of social theory and recent sociotechnical change. On the one hand, the concept could be considered as outdated while, on the other hand, remaining topical. Boyne's term therewith summarizes an ambivalence that can be seen in the broad theoretical debate on panopticism and in a range of newer derivative terms. As he pointed out, an increasing number of comparably unspectacular digital systems could appear to support the concept, while there were also powerful arguments against it; such as the link between surveillance and simulation, that alters the temporality of the panoptic frame; or a 'visibility' that Thomas Mathiesen (1997) termed "Synopticism", of the "few" by the "many", corresponding to a reverse panopticon.[17]

Taken together, these (more or less historical) visions and their association with different digital "smart machines", systems and practices, suggest a history of visions of "panoptic" data that indeed brings to the fore 'visibility' as the common denominator linked to, in Zuboff's terms, the idea of "universal transparency". This idea encompasses both surveillance and transparency, intertwined in a "visibility" that may be seen today, as Shiv Ganesh (2016) argues, as a "major 21st century trope".[18] As such, it underlies seemingly opposite concepts and metaphors including the 'glass human being', the 'transparent state' or, as has been critically pointed out, "the panoptiziced state".[19] It also underlies the characterization of the internet as, respectively; a "medium of transparency par excellence" or as something "structurally resembling a Panopticon",[20] resonating with Boyne's finding that 'somehow' panopticism is not outdated, but topical. The concluding question Boyne raised is what that 'somehow' means, and what this notion refers to. Emphasizing a difference "between actual social contexts and an updated ideal type [...] in contemporary technoculture", he came up with the idea to "draw a line" through terms such as "Panopticon"; therewith allowing this vision "to be seen at the same time as denying its validity as description".[21] Boyne thus highlighted a distinction that shifts the attention from the uses of metaphors of data technologies to the question of their vision.

Oligoptica and (big) data

'What do we see?'

Using the example of a traffic monitoring system on Paris' Boulevard Périphérique, Bruno Latour (1998) emphasized that digital technologies actually give rise to something quite different from a Panopticon, which he called "Oligopticon". The oligoptic view, he argues, "is not what sees everything, but what sees a little bit", or related to "places which have a total view under a very very small perspective". This counter-image is further described by Latour (2005) as opposite to the "absolutist gaze" and "utopia" of the Panopticon: Oligoptica, he explains, provide "sturdy but extremely narrow views of the (connected) whole".[22] Asking 'what we see', Latour thus arrives at a metaphor that puts in question the vision of an all-seeing 'panoptic' technology, associated with 'universal transparency' or a digital 'illumination', and replaces it by a theoretical and analytical concept

[16] Boyne: Post-Panopticism, p. 303.

[17] Mathiesen: The Viewer Society, p. 218f. The principle of a reverse panopticon is also worked out in Bentham's "Constitutional Code" (1830). Cf. Brunon-Ernst: Deconstructing Panopticism into the Plural Panopticons, p. 24.

[18] Ganesh: Managing Surveillance, p. 164.

[19] Fenster: Seeing the State, p. 668.

[20] Cf. Baumann: Die schöne Transparenz-Norm und das Biest des Politischen, p. 399 ("Transparenzmedium schlechthin", quot. Zehnpfennig 2013), p. 401 ("Das Internet selbst gemahnt von seiner Struktur her an ein Panoptikum").

[21] Boyne: Post-Panopticism, p. 303.

[22] Latour: Reassembling the Social, p. 181.

that focuses on how data and datafication take place and work in practice. In their discussion of Pervasive Computing, Dodge and Kitchin (2007) draw on Latour's terminology and raise precisely the question of this difference between oligoptic data technologies and a panoptic scheme particularly associated with the emergence of sousveillance. In his analysis of (big) data, Kitchin (2014) further develops this approach. Reconsidering (small and big) data and examining the ramifications of (big) data technologies, he dissociates digital systems from an overall vision and from metaphors of transparency. The notion of data as oligoptic, in contrast, raises the question of how data – collected or captured, processed and presented in particular ways – are produced and translated into 'what is seen'. This shift of attention to data and data technologies conceptually reframes big data in terms of "assemblages" or of ensembles that make up a complex sociotechnical infrastructure. It provides a perspective on the interplay between data technologies and their intertwined technical and social implications that focuses on the practice of big data, taking into account the actually different ways in which data technologies are shaped and shape today's sociotechnical landscape.

Latour's counter-concept to the "absolutist gaze" and "utopia" of the Panopticon also questions the idea, or contemporary trope, of visibility, whether it is linked to digital data, to the internet or to computing in general. This relation is indeed somewhat dubious when considering the basic fact that digital data have little to do with vision. Although digital media include visual media (as far as this categorization makes sense),[23] neither data not computer – even the "cognitive computer", up to now[24] – are traditionally considered as having the ability to 'see', and computation is, in itself, quite distinct from vision. Visual metaphors are, of course, useful, even needed due to a lack of visibility of computing. Articles on big data make use of them, and the vision of an "era of cognitive computing", as communicated online by IBM, is presented in videos; thus contributing to a nexus between vision and computing.[25] To refer to eyes, focus or perspectives as metaphors is also a common practice, including Latour's counter-metaphor. What his opposition brings to the fore, however, is how a vision of data or computing associated with an idea of universal transparency ends up being remote from usual metaphors or images of vision. It would, by contrast, correspond to no imaginable vision and no perspective at all, but to a 'vision' (or "gaze") of data by (or to) which everything would be immediately and incontestably evident.[26] A trope of visibility associated with such an "utopian" vision and transparency would thus associate both data and vision per se with incontestable evidence. In practice, such ideas add to the risk of "data determinism",[27] a term used by Kitchin to point to social and ethical issues arising from unquestioned general assumptions without consideration of the particularities of data and data technologies, or of the particular insights they provide.

Big data as a sociotechnical infrastructure

Such considerations suggest that issues of big data may be more appropriately addressed in terms of a complex sociotechnical infrastructure, apart from a trope of visibility. The question then becomes not only 'what is seen', metaphorically speaking, but also, what is in which way connected within the dynamic ensembles of an emerging infrastructure of data technologies and practices. This shift draws attention to data in the vast array of ways in which they are part of social practices (of classification, ordering or signification), as well as to the ways in which data and data systems are – and become ever more – connected. As the examples above

[23] Cf. Mitchell: There Are No Visual Media. Mitchell argues that, first, there are only mixed media, and, second, there is no such thing as a pure visual perception or an unmediated vision.

[24] See http://asmarterplanet.com/blog/2015/08/seeing-believing-bringing-cognitive-image-analytics-healthcare.html.

[25] Wendy Chun has emphasized the "decline in and frenzy of visual knowledge" computers – "[o]paque yet transparent" – have given rise to. Chun: Programmed Visions, p. 15. In the presentation of cognitive computing and image analytics mentioned above both, the "prolieferation of digital images (new media as 'visual culture')" and "'total information' systems (new media as 'transparent')" (ibid.) can be seen, and conflate with the topic featured – image analytics – in one video. The question 'what we see' is, here, not easy to answer.

[26] This would correspond to "digital metaphysics" following the term coined by Raffael Capurro. Cf. Capurro: Beyond Humanisms, p. 11. A further discussion of 'vision' and 'cognition' might be opened by Mitchell's reference to Oliver Sacks' "To See and Not See"; and by Bruno Latour's notion of "Visual Culture": As Latour has pointed out, a "new visual culture redefines both, what it is to see and whatthere is to see"; Latour: Visualization and Cognition, p. 9; cf. Mitchell: There Are No Visual Media.

[27] Kitchin: The Data Revolution, p. 45.

underline, big data and the social and ethical issues that arise from the field of connected data and datafication can be seen against the background of digital change and thus in a frame that differs from the "before-after model"[28] underlying some of the present debates. Drawing on studies by Kitchin and other authors, three focal points of such a frame may be distinguished, corresponding to different aspects of big data as a sociotechnical infrastructure related to personal data and data capture; to algorithms and their 'opacities'; and to the increasing number of connections between previously separate data and data technologies.

Data, or "capta" (Agre 1994), as has been pointed out by many authors, have not only become a "data deluge", but have also changed in terms of measurement and production, capture, collection, and classification of personal data. Produced ever more as a 'by-product' of activities and behaviour (from social media to smartphones, smart cards or sensors in cars), (big) "data" today refers to a variety of data, or of ways of producing data which have broken down traditional clearly defined borders. From the point of view of quantitative social science research, Plantin et al. (2015) refer to a "fracturing of the [traditional] control zone", suggesting that the "uncertainty about the provenance of data", rather than volume or variety, is the main characteristic of big data.[29] From the point of view of the individual, the capture of data by an array of applications and measurement devices results in heterogeneous 'traces' quite different from traditional notions of 'personal data', and produced in ways that depend on particular applications. What ends up being captured (or selected to be measured) how and in which framework of signification, is based on defined models and their underlying assumptions, independent from any personal involvement or conditions.[30] While in many cases data is captured without the individual's or the user's knowledge, sousveillance, or practices of self-monitoring and self-measurement on a voluntary basis, contribute in particular ways to a normalization of personal data capture – as already noted by Dodge and Kitchin (2007). Today, tracking-App's and Wearables, embedded in cultures of gamification, fitness and body-measurement, are popular 'persuasive' media that raise, at the same time, issues of the uses, the value and interpretation of the data, as well as of the thin line separating voluntary from "pushed", or even "imposed", self-monitoring.[31]

While (big) data are associated with a vision of 'transparency', big data technologies and algorithms are associated with "opacity", as Jenna Burrell (2016) notes, emphasizing a conflation that underlies this association between algorithms and the number of factors that contribute to an "inherent [...] uncertainty of big data", as Kitchin puts it, or to issues that arise from "a state of not knowing".[32] Burrell suggests that three different forms of "opacity" may be distinguished, resulting from protected access (in case of corporations or institutions); from a lack of specialist skills; and from a "mismatch between mathematical optimization in high-dimensionality characteristic of machine learning and the demands of human-scale reasoning and styles of semantic interpretation."[33] She thus makes a case for a more precise distinction between the mathematical logic of (different) algorithms and issues of their interpretation, commercial and institutional applications and use. Further aspects of interpretation may be added, in particular the role of visualization and visual data analysis "as a sense-making tool".[34] In practice, data technologies may be seen as (particular) applications that raise issues in precisely this mixture of 'opacities', while they are increasingly, as Dominique Cardon (2015) argues, shaping societal cognitive and cultural frames; including hypotheses underlying future predictions. Cardon focuses on the development of data technologies in the web and, from this point of view, distinguishes between four types of calculation, related to (1) votes and a principle of popularity (based on views); (2) meritocracy and a principle of authority (based on links); (3) benchmark and a principle of "gloriometric" reputation (based on likes); and (4) machine learning and a principle of prediction (based on traces).[35] His analysis thus brings

[28] Kaschuba: Die Überwindung der Distanz, p. 97 ("vorher-nachher-Modell").

[29] Plantin et al.: Big data is not about size.

[30] "Digital footprints" may thus rather be seen as particular "oligoptical views" following Kitchin: The Data Revolution, p. 167.

[31] Deborah Lupton quot. in Reichert: Digitale Selbstvermessung, p. 75.

[32] Burrell: How the machine 'thinks', p. 1; Kitchin: The Data Revolution, p. 102, cf. pp. 149ff.

[33] Burrell: How the machine 'thinks', p. 1f.

[34] Bollier: The promise and peril of Big Data, p. 9; cf. Kitchin: The Data Revolution, pp. 106ff.

[35] Cardon: A quoi rêvent les algorithmes, p. 17ff. The term "gloriomètres" (p. 30) is quot. from Gabriel Tarde.

to the fore the range of procedures of classification, ranking, sorting and scoring that are associated with the rise of big data.

Furthermore, as Burrell and other authors point out and as a more historical perspective makes clear, the rise of big data can not be isolated from earlier and parallel developments, particularly in the area of mobile and pervasive media, computing, and data collection techniques. In ever more environments of work as well as private life, data capture and processing are involved, and data technologies shape and continuously transform areas such as, among others, finance and retail, energy, mobility, health, education, and science. The role of increasingly automated decision making on individuals, in particular in regard to work and employment, has been associated with implications of "data discrimination" (Barocas/Selbst 2015), a keyword that embraces the socially discriminatory effects of the use of data mining applications.[36] In parallel, data are increasingly aggregated, analysed in real-time, repackaged and traded as a commodity. Data brokers collecting and selling personal data can have a significant impact, as has been shown, on decisions over individuals, including predictive analysis, without knowledge of the individuals concerned; thus leading to a "data determinism" in which individuals are classified and judged based on profiles not only of data about what they have, or maybe have done in the past, but also of predictions "of what they might do" in the future.[37] The increasing number of such multiple connections, reflected in the (above mentioned) characteristic uncertainty about the provenance of (big) data, and of the "re-scalings and reorderings" they imply require more attention, as Kitchin argues, and run counter to ideas that tend to 'reify' big data.[38] He suggests, in particular, that the implications of this increasing interconnection of previously separate, 'oligoptic' data, data technologies and systems, and of an increasingly pervasive data infrastructure raise issues that need be addressed.

Conclusion

'Big data' can be considered as a field of both major visions and major implications of present data technologies. As a rapidly emerging field of digital change, it raises questions which, as Kitchin points out, "we have barely begun to ask and answer".[39] The goal of this paper was to approach the field, not from the (range of) issues that arise from the definition of big data, but through shifting the attention to the role of optical metaphors and relations between data and vision that shape concepts and present debates of data and datafication. (Re)assessing notions and visions of panoptic data technologies, I have proposed to address (big) data from three perspectives. I have, first, argued for a historical perspective that situates big data in a history of digital change, allowing to more precisely reframe its newness, and in a "history of the future" – of promises/threats and visions of a future "already out there".[40] As shown, the notion of panoptic, all-seeing data technologies can itself be situated in a history of information technology and its visions. Second, I have proposed to address this notion quite literally in terms of vision, arguing in favour of a perspective that allows to separate data technologies from a vision associated with universal transparency, and from a contemporary trope of visibility. Finally, based on Latour's notion of vision, I have argued in favour of an approach that draws attention to issues of (big) data which have been overshadowed by this trope and proposed to address them in terms of a complex sociotechnical infrastructure. As discussed, such a perspective can provide a more precise view that goes beyond generalized notions of big data and allows to distinguish between different ways in which present data technologies and practices raise particular social and ethical questions. This includes different aspects of 'opacity' (or transparency, respectively) of present technologies of (big) data.[41] The notion of big data as an

[36] The authors note that, of course, by "definition data mining is always a form of statistical […] discrimination" in the sense that data mining is intended "to provide a rational basis upon which to distinguish between individuals", but suggest, in regard here to US-American law, that "[n]evertheless, data mining holds the potential to unduly discount members of legally protected classes and to place them at systematic relative disadvantage." Barocas and Selbst: Big Data's Disparate Impact, p. 7.

[37] Kitchin: The Data Revolution, p. 45; see, e.g., "The Dark Market for Personal Data", The New York Times, 17.10.2014.

[38] Kitchin: The Data Revolution, p. 99; cf. p. 181.

[39] Ibid., p. 99.

[40] Carey and Quirk: The History of the Future, p. 150.

[41] Cf. for the related discussion on auditing algorithms, Burrell: How the machine thinks; Cardon: A quoi rêvent les algorithmes, p. 83.

infrastructure, moreover, draws attention to the increasing interconnection between previously separate, different systems and societal domains, and to its implications.

References

Agre, Philip E.: Surveillance and Capture. Two models of privacy, Information Society 10 (1994), 2, pp. 101–127.

Barocas, Solon and Selbst, Andrew D.: Big Data's Disparate Impact, preprint (2015), California Law Review, Vol. 104 (2016) (forthcoming), available at http://ssrn.com/abstract=2477899.

Baumann, Max-Otto: Die schöne Transparenz-Norm und das Biest des Politischen. Paradoxe Folgen einer neuen Ideologie der Öffentlichkeit, Leviathan 42 (2014), 3, pp. 398–419.

Bentham, Jeremy: Panopticon; or, the Inspection-House; containing the idea of a New Principle of Construction applicable to any sort of Establishment, in which Persons of any Description are to be kept under Inspection [1791], in: The Works of Jeremy Bentham, John Bowring (Ed.), Vol. IV, New York: Russell & Russell 1962, pp. 37–172.

Boellstorff, Tom: Making big data, in theory, Firstmonday, 18 (2013), 10, available at http://firstmonday.org/article/view/4869/3750.

Bollier, David: The Promise and Peril of Big Data, The Aspen Institute, Washington DC 2010, available at http://www.aspeninstitute.org/publications/promise-peril-big-data.

Boyne, Roy: Post-Panopticism, Economy and Society, 29 (2000), 2, pp. 285–307.

Brivot, Marion/Gendron, Yves: Beyond panopticism: On the ramifications of surveillance in a contemporary professional setting, Accounting, Organizations and Society, 36 (2011), 3, pp. 135–155.

Brunon-Ernst, Anne: Deconstructing Panopticism into the Plural Panopticons, in: Beyond Foucault. New Perspectives on Bentham's Panopticon, Anne Brunon-Ernst (Ed.), Farnham and Burlington: Ashgate 2012, pp. 17–41.

Burrell, Jenna: How the machine 'thinks': Understanding opacity in machine learning algorithms, Big Data & Society 3 (2016), 1, pp. 1–12.

Capurro, Rafael: Beyond Humanisms, Journal of New Frontiers in Spatial Concepts 4 (2012), pp. 1–12.

Cardon, Dominique: A quoi rêvent les algorithmes. Nos vies à l'heure des big data, Paris: Seuil 2015.

Carey, James/Quirk, John: The History of the Future, in: Communication as Culture: Essays in Media and Society, James Carey [1st ed. 1989], New York and Abingdon: Routledge 2009, pp. 133–154.

Chun, Wendy Hui Kyong: Programmed Visions. Software and Memory, Cambridge MA and London: MIT Press 2011.

Dodge, Martin/Kitchin, Rob: 'Outlines of a world coming into existence': pervasive computing and the ethics of forgetting, Environment and Planning B, 34 (2007), pp. 431–445.

Fenster, Mark: Seeing the State: Transparency as Metaphor. Administrative Law Review, 617 (2010), available at http://scholarship.law.ufl.edu/facultypub/572.

Foucault, Michel: Surveiller et punir: Naissance de la prison. Paris: Gallimard 1975; engl. translation: Discipline and Punish: The Birth of the Prison, New York: Random House 1975.

Ganesh, Shiv: Managing Surveillance: Surveillant Individualism in an Era of Relentless Visibility, International Journal of Communication 10 (2016), pp. 164–177.

Kaschuba, Wolfgang: Die Überwindung der Distanz. Zeit und Raum in der europäischen Moderne, Frankfurt a.M.: Fischer 2004.

Kelly, John E./Hamm, Steve: Smart Machines. IBM's Watson and the Era of Cognitive Computing, New York: Columbia University Press/IBM Corporation 2013.

Kitchin, Rob: The Data Revolution. Big Data, Open Data, Data Infrastructures and Their Consequences, London etc.: Sage 2014.

Latour, Bruno: Visualization and Cognition: Drawing Things Together, in: Knowledge and Society. Studies in the Sociology of Culture Past and Present, Henrika Kuklick/Elizabeth Long (Eds.), Greenwich CT and London: Jai Press 1986, pp. 1–40.

Latour, Bruno: Thought Experiments in Social Science: from the Social Contract to Virtual Society. 1st Virtual Society? Annual Public Lecture, Brunel University 1998, available at http://www.artefaktum.hu/it/Latour.htm.

Latour, Bruno: Reassembling the Social, Oxford: Oxford University Press 2005.

Laval, Christian: From Discipline and Punish to The Birth of Biopolitics, in: Beyond Foucault. New Perspectives on Bentham's Panopticon, Anne Brunon-Ernst (Ed.), Farnham and Burlington: Ashgate 2012, pp. 43–60.

Leclercq-Vandelannoitte, Aurélie/Isaac, Henri/Kalika, Michel: Mobile Information Systems and Organizational Control: Beyond the panopticon metaphor? European Journal of Information Systems 23 (2014), 5, pp. 543–557.

Mathiesen, Thomas: The viewer society: Michel Foucault's 'Panopticon' revisited, Theoretical Criminology 1 (1997), 2, pp. 215–232.

Mitchell, W.J.T.: There Are No Visual Media, Journal of Visual Culture 4 (2005), 2, pp. 257–266.

Plantin, Jean-Christophe et al.: Big Data is not about size: when data transform scholarship, final draft version (2015), available at http://pne.people.si.umich.edu/articles.html.

Reichert, Ramón (Ed.): Big Data. Analysen zum digitalen Wandel von Wissen, Macht und Ökonomie, Bielefeld: transcript 2014.

Reichert, Ramón: Digitale Selbstvermessung. Verdatung und soziale Kontrolle, Zeitschrift für Medienwissenschaft 13 (2/2015), pp. 66-77.

Zuboff, Shoshana: In the Age of the Smart Machine. The Future of Work and Power, New York: Basic Books 1988.

Jessica Heesen:

Big Data for a Fairer Democracy?

Abstract:

Big data-analysis is linked to the expectation to provide a general image of socially relevant topics and processes. Similar to this, the idea of the public sphere involves being representative of all citizens and of important topics and problems. This contribution, on one side, aims to explain how a normative concept of the public sphere could be infiltrated by big data. On the other, it will discuss how participative processes and common wealth can profit from a thorough use of big data analysis. As important parts of the argument, two concepts will be introduced: the numerical public (as a public that is constituted by machine-communication) and total politicisation (as a loss of negative freedom of expression).

Agenda:

Author:

PD Dr. Jessica Heesen:

- University of Tübingen, Centre for Ethics in the Sciences and the Humanities, Head of Junior Research Group Media Ethics, Wilhelmstraße 19, 72074 Tübingen, Germany
- ☎ + 49 – 7071 2977516, ✉ jessica.heesen@uni-tuebingen.de, 💻 http://www.izew.uni-tuebingen.de/
- Relevant publications:
  - Computer and Information Ethics. In: Ruth Chadwick (ed.): Encyclopedia of Applied Ethics. Second Edition, Vol. 1, Academic Press: San Diego 2012, 538 – 546 p.
  - Ein Fels in der Brandung? Positionen der Medienethik zwischen verflüssigtem Medienbegriff und schwankender Wertebasis, in: Marlis Prinzing/Matthias Rath/Christian Schicha/Ingrid Stapf (Hg.): Neuvermessung der Medienethik. Bilanz, Themen und Herausforderungen seit 2000, Beltz Juventa: Weinheim 2015, 86 – 98 p.

ISSN 1614-1687

Introduction

The concept of "big data" indicates the interplay of different technological developments. On the one hand, big data stands for efficient storage and management of huge amounts of data. On the other hand, it depends on the always growing sources for digital information due to mobile applications, an "always on"-mentality, and the pervasion of machines, communication, and the whole environment with information technology (internet of things, ambient intelligence). But big data does not only mean that more data is gathered from more sources – additional important aspects are new methods for the evaluation of data pools. The outcome will be demonstrations of new correlations or - as expressed by the technical term "knowledge discovery"[42] - a new "knowledge" that is generated from data after a process of interpretation and modelling.

The ability to analyze large amounts of data is being discussed more and more often in the context of its usefulness for the improvement of political decision making and social welfare; for example for the prediction of epidemics[43], for smart policing (http://www.smartpolicinginitiative.com), traffic management, energy supply or the identification and publication of administrative affairs (e.g. http://www.dublindashboard.ie, http://www.d4d.orange.com/en/Accueil).

The MIT researcher Alex Pentland mentions in particular the richness of detail of big data analysis and the breadth of information it provides for all sectors of society. He has great expectations concerning the future benefits of big data: "We're going to end up reinventing what it means to have a human society."[44] That implies: potentially everything could be important and should be gathered and analyzed with the new comprehensive methods. At this point, an interesting linkage to the concept of the public sphere is emerging: big data-analysis is linked with the expectation of providing a general image of socially relevant topics and processes. Similar to this, the idea of the public sphere involves being representative of all citizens and of important topics and problems. This article wants to point out how a normative concept of the public sphere is challenged when a technology is used to support or substitute parts of its functions.

Information technologies as tools and media

Big data analysis uses a variety of data which provides information on everyday activities, the usage of internet-services and the communication behaviour of individuals. Because of their mobility and their intertwining with objects of every-day live, information technologies (IT) are ubiquitous. Yet at the same time, they keep their function as a "window to the world" – but in two directions: on the one hand, for example television, telephone and internet communication in general serve as media for information and communication; on the other, IT potentially reports the individual behaviour to an unknown circle of machines and/or human beings or, in the case of social networks, to a broader public sphere. Big data methods profit from the features of computers as universal machines with numerous users and contextual usages.

In this context, information technologies can be described as hybrids between tools and media. Therefore, sources for the gathering of data are, both, the witting respectively unwitting interaction with a technical system and the interaction between human beings by use of natural language. Only this latter "natural" communication is part of a discursive conception of the public sphere as it is justified in normative concepts of the public in philosophy and political theory. This normative concept includes the supposition that public discourse – to a certain extent – is guided by rational justification, refers to reciprocal expectations, and is focused on certain

[42] Cf. Fayyad et al. (1996).

[43] Cf. Ginsberg u.a. (2009).

[44] Pentland (2012).

topics.[45] In contrast, big data starts from other premises: Its "knowledge" is constituted from data material that is generated through the everyday routines of people.

Nevertheless, there could be good reasons for intensified usage of big data for decision making procedures. In the following section four reasons will be mentioned why big data analysis could have advantages over politics as usual.

(1) It might be the case that knowledge from data analysis can claim a higher proximity to reality. In many empirical investigations one finds that people often give a subjective or strategic evaluation of their behaviour. For example, when asked about their highest priority interest with respect to internet communication, many people answer "information", whereas statistics tell us that in reality, pornography is significantly more popular.[46] In this way analysis of data can produce a more accurate, "truer" empirical knowledge.

(2) Moreover, one could imagine that a comprehensive gathering of data takes into account all those people that are not part of public discourse and the publicity of (mass)media and its own laws (for example, so called marginal groups). Therefore a more intensive use of big data for politics could lead to more social equality.

(3) Also a general lack of interest of the population in the democratic process could be balanced by behavioural analysis. In many western democratic systems we have come to notice a decrease in voter turnout. Thus data analysis could offer a compensation for election and active participation.

(4) And as a last point it should be mentioned that big data includes a promise for more neutrality and rationality in political decision making, as pointed out for example by many advocates of so called information governance technology.[47]

Technocracy instead of democracy

A consistent application of big data analysis might therefore have advantages in terms of policy-making. Possibilities for datafying or respectively mediazation offer a lot of potential advantages regarding government and regulation. For example there is talk of "evidence-based policy"[48] and the aforementioned "information governance". These are phenomena of a new technocratic movement where governments and organisations try to strictly follow findings of science and research. Daniel Gillick, a Google Senior Research Scientist, states "'Big data' represents a cultural shift in which more and more decisions are made by algorithms with transparent logic, operating on documented immutable evidence."[49]

In this respect, one could speak of a newly optimized technocratic process. Initially, the term technocracy indicated a concept for the regulation of societal organization. According to that, "mechanical" elements of scientific and technological civilization assume the role of political arrangements. From this perspective, the advantage lies in an efficient social regulation that is immunized against inappropriate ideological decisions. In a technocracy, instead of the political sovereign, the optimized self-organization of the people through science and technology gains centrality.

Due to the established values of constitutional democracies, technocratic forms of government are refused; in particular because they undermine the role of the political sovereign and the principle of democratic self-organization.[50] Yet, constitutional democracies use significant technocratic elements of administration as well.

---

[45] Cf. Habermas (1996).

[46] Daily Infographic (2013).

[47] Information Governance Initiative, http://iginitiative.com/

[48] Head (2010).

[49] Dutcher (2014).

[50] Cf. Habermas (2013).

This becomes obvious for example in the definition of the goals for climate control: They are based on scientific forecasts, but they have heavy consequences for economic and political processes. Another example is the development of new forms of participation like e-democracy or open data, which are driven by new technological possibilities.

There is no doubt that it makes sense to use empirical and general scientific knowledge for political processes and decisions. How else could one come to an informed decision? Despite this, the advanced forms of data analysis develop their own dynamics that connect to the past lines of discussion about technocracy.[51] The special authority of the "pure number", the apodictic evidence of calculation models and the apparent neutrality of quantification and statistics were always reason for disputes about their validity in terms of political action (but also within the epistemological dimension).[52] Moreover, quantification is one of the methods to give a topic a high news value and, according to this, to gain a high level of public attention.[53] But what distinguishes the use of data analysis for social regulation on one side from a societal self-regulation by a political public sphere on the other?

Indicators for a normative concept of public communication are:

1. Conscious participation: this is connected to a specific cognitive mind-set of the participating persons in relation to their communication as public communication. On the one hand this attitude is reflected in the selection of the topic of communication. Accordingly, it is visible in a supra-individual choice of subject, or rather its relevance for a discussion of the res publica and public welfare. On the other hand this attitude is also reflected in the awareness of the individual behaviour: that means persons in a public situation will reflect on their impact on others. For example they will ask themselves if their behaviour is commendable and if they wish that anybody would act like they do.

2. Reciprocity: associated with this are also mutual expectations of communicators as they are formulated by discourse ethics. These expectations involve truth, rightness, and truthfulness of communication, which is oriented towards reaching understanding.[54]

3. Orientation: media publicity produces a "knowledge from the knowledge of the other". This orienting knowledge is given through the ability to monitor what is happening in mass media communication. This knowledge includes not only knowledge of facts or historical contexts, but also knowledge on the validity of normative rules. Not least it gives information about the emergence of social understanding as such. Thus, the orienting function of public communication is a prerequisite for the realization of an individual's successful social participation.

These forms of orientation, reciprocity, and conscious participation require a special normative definition of the public. This normative concept is concerned with the function of the public sphere in a democratic system and the preconditions for making the public a valuable element of a liberal constitutional state. In spite of the importance of a normative concept of the public for political theory, empirically existing public spheres are often shaped by the prevalence of power struggles and the dynamics of media communication. In this sense a normative definition of the public sphere is oriented to an often counterfactual ideal.

[51] In this regard, cf. the controversy between positivism and dialectic approaches (Adorno et al. 1989) as well as the debate about technocracy (Bell 1980, Ellul 1953, Lenk 1973).

[52] Cf. Porter (1996).

[53] Cf. Luhmann (1990: 178).

[54] Cf. Habermas (1991).

Numerical public and total politicisation

But in the case of big data, things are different: In contrast to the development of a counterfactual ideal, big data is based on other conditions. The use of big data for policymaking builds on the numerous traces of individual action that are to be detected following the penetration of everyday life with information technologies. Unlike in public discourse – where citizens can regulate their participation in the public sphere –, big data-analysis causes the gathering and publication, especially, of every day-activities and behaviour. Because of this, big data gains a numerical picture of the general public *and* the general private that is connected by correlations, not by interactions. In this sense, one can speak of a numerical public here, thus, a public that does not reflect on its own publicness and public responsibility.

On the one hand the individual involvement in the constitution of a big data pool can be called participative in a way, but on the other hand this kind of "participation" opens a technical way for the pervasion of private and everyday life with a political or rather ideological component. Total participation, enabled by context-aware media, ambient intelligence and big data, implies, at the same time, a total politicisation. In this case, politicisation must not be understood as reciprocal relationship with discourse and critique in terms of the definition of public welfare, but as an anonymous communitization[55] in relation to different social contexts like economy, traffic management or insurance business. This means that the actions of every single person are arranged and in the end assessed with respect to specific patterns. The single action is no longer an individual one but part of a "typical situation", the fulfilling of a specific stereotype in a similar group of people. On the one hand, this anonymous communitization is always combined with a comparison of individual behaviours with the current notion of normality and is an important factor for a feeling of surveillance and its negative consequences.[56] On the other hand, individual (private) actions gain importance for supra individual decision making.

To give a simple example: Maybe you have a weakness for superficial entertainment shows and from time to time you watch one of these shows with your smart television. Your internet provider, the broadcasting company and/or the smart TV-company will collect your data or rather your selection and maybe in the future more superficial entertainment will be offered for you and for the general public. But if somebody would have asked you individually for your opinion about the offers of TV-channels probably you would not have supported an increasing amount of entertainment shows. On the contrary you could have voted for more documentary films, news programmes or auteur movies because you think public communication should support intellectual involvement – even if you personally love superficial entertainment.

So, it can be noted that not the motivation for action is important for a data analysis, but the fact that you conduct it. In such a kind of data analysis it is not about finding reasons but about recognising patterns in data agglomerations. In this respect, a systematic use of big data for political action implies the exclusion of counterfactual and probably moral reasons from political reflection. Big data confines societal analysis to what is empirically existent. Instead of this, in public discourse citizens, journalists and stakeholders (and unfortunately also spin doctors) have the possibility for a deliberate (innovative or critical) discussion of topics. In this respect, data analysis only reproduces what exists anyway and reduces the possibilities of information technology to its affirmative – the societal conditions confirming – forms of usage.

Moreover, total politicisation implies that every individual action is registered, and, as a consequence, every individual person could start thinking about the political and ethical dimension of his or her action. In a society with a comprehensive use of big data, all areas of life gain importance for the public and the political and we get a "participation" by the analysis of behaviour. As another example: data from your private fitness program can be used by medical insurances to change their concept of solidarity.

[55] Cf. in a different context Hubig (2003).

[56] Cf. Haggerty/Ericson (2006).

Furthermore, for a total politicisation it is typical to have a dissolution of nonpolitical private and social spaces. Which in turn leads to the loss of a negative freedom of expression. But: in a liberal state you have the right not to express your opinion or attitude. So it can be concluded that (roughly) total behavioural analysis is quite the opposite of a seemingly non-ideological and neutral approach to reality – the aspect that is stressed as an advantage by the advocates of information governance. With total behavioural analysis, there are no longer any non-political and in this sense non-ideological actions.

How to use big data in a smart way

The foregoing analysis tried to describe the theoretical differences between a social organisation through big data on the one hand and public discourse on the other. At any rate, on an empirical level we cannot see a prevailing comprehensive use of data analysis yet. Nevertheless applications from information technology already determine areas of politics like security and energy politics. Hence public deliberation about the chances and risks of data analysis is essential to reflect on these new opportunities and to justify them as well informed decisions.

Big Data entails a lot of critical points to consider but on the other side there are a lot of valuable data for societal self-reflection and regulation that we should use. As a conclusion, the following section lists four assumptions that are important for a prudent "self-tracking of democracy":

1. Awareness of the advantages and limitations of data analysis:
   Don't have too much respect for "pure" data
2. Use of privacy by design to avoid a totalitarian accumulation of data:
   Don't forget the principles of data protection like informed consent, limited purpose, relevance
3. Provision of governmental data as open data:
   "Make public data available, protect private data" (CCC)
4. Integration of big data in public discourse:
   With the aim of ensuring well informed decisions and showing the possibilities and capabilities of data gathering for politics, civil society and public welfare, data journalism[57] starts to be established.

References:

Adorno, Theodor W. et al (1989): Der Positivismusstreit in der deutschen Soziologie 13th Edition. Darmstadt/Neuwied: Luchterhand.

Bell, Daniel (1980): Veblen and the Technocrats: On the Engineers and the Price System. In: The Winding Passage: Sociological Essays and Journeys 1960 - 1980, Cambridge: Harvard University Press, 69-90.

Chaos Computer Club (CCC): Hackerethik. https://www.ccc.de/de/hackerethik (13.1.206)

Daily Infographic (2013): The Stats on Internet Pornography. http://www.dailyinfographic.com/the-stats-on-internet-pornography-infographic.

Dutcher, Jennifer (2014): What Is Big Data? (3.9.2014). https://datascience.berkeley.edu/what-is-big-data/ (19.11.2015)

Ellul, Jacques (1954): La technique ou l'enjeu du siècle. Paris: Armand Colin.

[57] Gray et al. (2012).

Fayyad, Usama M. et al. (ed.) (1996): Advances in Knowledge Discovery and Data Mining. Amer. Assn. for Artificial Intelligence, Menlo Park, CA, USA, AI-Magazin, 37–54.

Ginsberg, Jeremy et al. (2009), Detecting influenza epidemics using search engine query data. In: Nature 457, 1012 - 1014, 19 February 2009, doi: 10.1038/nature07634; (Published online 19 November 2008; Corrected 19 February 2009).

Gray, Jonathan et al. (2012): The data journalism handbook. O'Reilly Media, Inc.

Habermas, Jürgen (1991): Erläuterungen zur Diskursethik. Frankfurt a. M: Suhrkamp.

Habermas, Jürgen (1996): Strukturwandel der Öffentlichkeit. Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft. Mit einem Vorwort zur Neuauflage 1990 5th Edition 1996. Frankfurt a. M: Suhrkamp..

Habermas, Jürgen (2013): Im Sog der Technokratie. Berlin: Suhrkamp.

Haggerty, K./Ericson, R. (2006): The New Politics of Surveillance and Visibility. Toronto: University of Toronto Press.

Head, Brian (2010): Evidence-based policy: principles and requirements. In: Strengthening evidence based policy in the australian federation, volume 1: proceedings01/2010. https://www.researchgate.net/publication/242709855_2_Evidence-based_policy_principles_and_requirements (10.12.2015)

Hubig, Christoph (2003): Selbstständige Nutzer oder verselbstständigte Medien – Die neue Qualität der Vernetzung.In: F. Mattern (ed.): Total vernetzt, Szenarien einer informatisierten Welt, Springer: Berlin/New York, 211-230.

Information Governance Initiative.he. http://iginitiative.com/ (13.1.2016)

Lenk, Hans (ed.) (1973): Technokratie als Ideologie. Sozialphilosophische Beiträge zu einem politischen Dilemma. Kohlhammer: Stuttgart.

Luhmann, Niklas (1990): Gesellschaftliche Komplexität und öffentliche Meinung. In: Soziologische Aufklärung 5. Konstruktivistische Perspektiven. Westdeutscher Verlag: Opladen, 170-183.

Pentland, Alex (2012): Reinventing Society in the wake of big data, A Conversation with Alex (Sandy) Pentland [8.30.12]. http://edge.org/conversation/reinventing-society-in-the-wake-of-big-data (30.9.2015)

Open Data City, https://opendatacity.de/ (28.10.2015)

Porter, Theodore M. (1996): Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton University Press: Princeton.

Arne Manzeschke, Galia Assadi and Willy Viehöver:

The Role of Big Data in Ambient Assisted Living

Abstract:

Big Data and biopolitics are two major issues currently attracting attention in public health discourse, but also in sociology of knowledge, STS Studies as well as in philosophy of science and bioethics. The paper considers big data to be a new form and instrument of biopolitics (Foucault) which addresses both the categories of *body* and *space*. It is expected to fundamentally transform health care systems, domestic environments and practices of self-observation and reflection. Accordingly the paper points out some problems and pitfalls as well as open questions that have emerged in the field of AAL, which merit more attention in future public and academic debate.

Agenda:

Authors:

Prof. Dr. theol. habil. Arne Manzeschke:

- Institut Technik-Theologie-Naturwissenschaften an der Ludwig-Maximilians-Universität München, Katharina-von-Bora-Straße 11, 80333 München
- ☎ +49 – (0)89 – 5595 602, ✉ arne.manzeschke@elkb.de, 💻 http://www.ttn-institut.de/node/1356

Dr. Galia Assadi:

- Institut Technik-Theologie-Naturwissenschaften an der Ludwig-Maximilians-Universität München, Katharina-von-Bora-Straße 11, 80333 München
- ☎ +49 – (0)89 – 5595 605, ✉ galia.assadi@elkb.de, 💻 http://www.ttn-institut.de/node/1894

Willy Viehöver:

- Institut Technik-Theologie-Naturwissenschaften an der Ludwig-Maximilians-Universität München, Katharina-von-Bora-Straße 11, 80333 München

☎ +49 – (0)89 – 5595 603, ✉ willy.viehoever@elkb.de, 💻 http://www.ttn-institut.de/node/1893

ISSN 1614-1687

Inroduction

Big Data and *Biopolitics* are two major issues currently attracting attention in public health discourse, but also in sociology of knowledge, STS Studies as well as in philosophy of science and bioethics. The following paper considers Big Data to be a new form and instrument of *biopolitics* which is expected to fundamentally transform health care systems, domestic environments and practices of self-observation and reflection. The merging of Big Data and assisted living based information technological devices is intended to generate, store, interconnect, distribute and allow huge amounts and even highly diverse stocks of quantitative data to be used by health and social care providers, public services as well as users in private households. The visions of Big Data not only point to a multidisciplinary research area, a further step in rationalization, scientification and datafication of everyday life[58] – by gaining critical insights from data collected, shared and aggregated via fixed in-home sensor technology and mobile devices (Reeder et al. 2014) – they also represent a challenging shift in the "politics of life itself" (Rose 2007). What proponents of Ambient Assisted Living (AAL) arrangements[59] promise is to enhance personal security in residences as well as to raise and sustain the quality of life of disabled or sick people or aging individuals. It also caters to specific population collectives as, for example, in the case of obesity, borderline personality disorder or Alzheimer's disease.

Big Data, smart homes and AAL are currently being hyped as a new technological and organizational revolution, a rapidly expanding market, and a promising business model which is expected to benefit from integrating data from AAL, remote monitoring by advanced electronic devices and related electronic patient and health records. Benefits of assisted living projects based on Big Data are projected not only for side of decision makers, managers and service providers, but also for consumers, promising reduced costs, real time analysis, early alert monitoring and thus more patient safety (Vimarlund/Wass 2014), although there is no empirical data and scientific evidence to justify these expectations so far. Currently, business and science policy have been pushing the concepts Big Data and AAL into the health and social care market (Groves et al. 2013). The available literature and studies on the topic focus mainly on the technological and socio-economic factors and related challenges (Vilmarlund/Wass 2014). Going beyond a mere technological and economical view on this topic, how could one more appropriately frame this new and puzzling blend of data and technology (Groves et al. 2013)? A second glance reveals that the issue of Big Data and AAL is fundamentally related to the idea of leading a good life, i.e. the 'best' life possible. With regard to this, we assume that AAL projects are currently taking the shape of a new form of *biopolitics* that attempts to take on the task of both governance of individual and collective bodies (embodied persons as well as sub-populations) and surveillance of residential spaces (smart homes). What is peculiar to the case of AAL is that the idea of a good life, and thus the fundamentally ethical question of the appropriate conduct of life, becomes deeply entangled with smart technologies generating, gathering, mining, distributing, sharing and interpreting big amounts of (quantitative) data.

Although this announced technological revolution is still at an early stage, one can already discuss the question whether and with which consequences and (possible) side-effects it indicates a new qualitative step in biopolitics. According to the current public, health policy and academic discourse, the technological devices as well as patient records and health data are considered to support and facilitate everyday tasks of senior citizens or patient in households and adjacent areas by safeguarding their personal autonomy at the same time. We presume that looking at the concept of Big Data in AAL-context from a biopolitical point of view will open up a different perspective for examining the benefits and possible negative side-effects and pitfalls around AAL sites equipped with advanced electronic devices.

[58] See Weber (1981) and Weingart (1983) on rationalization and scientification, Driscoll (2012) and van Dijck (2014) on dataism, datafication and dataveillance.

[59] The term Ambient Assisted Living (AAL) was created for the European Joint Program in 2007; cf. http://www.aal-europe.eu/. Although it appears English in origin, the term has primarily been used in the German-speaking countries, promoted especially by the German Federal Ministry for Research and Education (BMBF). In 2011, the term was phased out of the ministry's official documents and replaced by the term »Altersgerechte Assistenzsysteme« which could be translated by »age-appropriate systems of assistance«; cf. Manzeschke et al. 2015.

Big Data, sophisticated technology in residential areas and the ethos of a good life

What is behind the concept of Big Data?

The concept of Big Data is still a fuzzy one; this is basically due to the fact that what is 'big' is constantly shifting with technological innovation in the IT sectors. Big Data cannot be simply defined by datasets beyond a certain number of terabytes and the definition of what is estimated as big often varies by sectors or institutional fields. The amount of digital data doubles every two years. We know that currently every day some quintillions of data entries are generated by GPS locators, messages, computer-networks, pictures, wearables, etc. This data is supposed to become a great source of knowledge on the one hand, but at the same time the production, circulation, storage and use of this huge amounts of data is equally seen as a root of serious problems, which mainly relate to their reliability, transparency and to questions of data security as well as the protection of privacy (Zwitter/Hadfield 2014).

What are the prerequisites for the generation, interconnection, integration, storage and use of Big Data (Vilmarlund/Wass 2014)? (1) A first point to be mentioned is the need for a technological infrastructure. Devices for the collection of data can be located in the body, on the body, or they can be installed independently in residences. Among these smart technologies are wearable technologies, motion sensors, cameras, gateways, web-based applications (webcam, interactive computer systems). In this context it should be added that Big Data sets can either be static or real-time data, especially in the health sector there is an increasing demand for real-time data which allow for personalized diagnosis and treatment and to deliver health care tailored for specific populations.

(2) Big Data sets derive from different information streams, sources or data records. They can 1) be generated by humans (e-mail or notes by physicians), 2) they can derive from social media or the web, 3) they can be read from machine to machine (from wearables to clouds or platforms), 4) biometric data (fingerprint) and 5) from transaction data. A sophisticated technological infrastructure is thus needed not only for the collection of data but also getting access, mining, circulation and aggregating different and often unstructured digital data.

(3) Big Data becomes functional only when it is embedded in an effective organizational infrastructure or when a recognized structure is co-established with the creation of Big Data sets (e.g., in the case of health- and biobanks). One emerging field of research, development and application of a Big Data vision regarding electronic patient and health records, is represented by the concept of AAL (Knaup/Schöpe 2014).

Ambient Assisted Living

The idea of AAL is closely linked to the application of innovative state-of-the-art information & communication technology (ICT) products in private households or larger residences, be it wearable devices, sensors, computers, electronic goods and software as well as services (e.g., online information systems and platforms or services) (Wichert/Klausing 2015). The technological devices used in AAL sites in the first place focus on the collection of body related vital data. Sources of information can be embodied (e.g., chip cards, pace-makers), wearable technologies like smart watches, smart phones making use of health apps or smart glasses worn on the body or interactive computers systems (e.g., avatars), TV, and online information systems. Other devices are designed to monitor the residential space elderly citizens are living (Nitzsche et al. 2014).

What is now claimed to be new with Big Data in AAL is a change in the direction of information flow. Information does not longer flow from the public to the individual household; rather the flow is reversed or becomes multidirectional in order to take advantage of Big Data. In order to support senior patients at home and to deliver appropriate and in-time health and social services a sophisticated technological infrastructure is needed to monitor people at home including all physical activities as well as other daily routines and habits (e.g., nutrition, medication).

As with most of the key technologies hyped by science, industries and science policies the new blend of smart living and Big Data are more based on expectations than on evidence (TAB 2011). Thence, what are the "horizons of expectations" (Koselleck 1989)? One of the major arguments insinuates a coming efficiency revolution in health care, which is based on data records generated, distributed, stored by smart technologies, technical devices and services and retrieved and used by physicians, care services, insurances, etc. The central objective of Big Data is not collecting individual records but rather the aggregation and integration of data and data stocks from different technological and medical devices and sources, because it allows for a more comprehensive view on health (Knaup/Schöpe 2014: 150). The enhanced effectiveness proponents expect requires not only the integration of data from different sources, another prerequisite is also the interconnection of health care providers (nurses, physicians, care givers, health record banks, etc.) and the active patient or client as well. Concerning the latter, as Vilmarlund and Wass (2014: 146) stress, a motivation to accept and use the technological equipment is a necessary condition for the implementation of the AAL concept as well as a technological empowerment of the user is needed. What is often underscored by promoters of AAL projects is that the data as well as the data sources should be accessible and tailored for the health decisions related to the concerned individuals. Thence the question who has legal access to the Big Data and the conclusions drawn and the medical advices given on their knowledge base is another urgent question.

The ethos of a *good life* in AAL

What is the social and cultural background of this hype on Big Data based AAL projects? Of course we currently face a massive trend in the datafication of everyday life and a blossoming trend to dataveillance too, both fueled by an ideology of dataism (van Dijck 2014; Lupton 2014; Driscoll 2012). This, however, seems to be a latent prerequisite for expanding Big Data to the health service in AAL-settings. AAL promoters assert that modern developed societies are currently facing a fundamental demographic change. A constantly growing percentage of elderly or aging citizens corresponds with a lack of skilled employees in the medical and health sector (Demiris 2008; Demiris/Thompson 2011; Leone/Diraco/Siciliano 2011; Knaup/Schöpe 2014).

A second motive is the right for a self-determined life and societal participation of disabled or elderly people (cf. Convention on the Right of Persons with Disabilities, 2006). AAL concepts are supposed to enable citizens in increasingly individualized societies to live independently in their private apartments or houses and participate in society. Third, as far as the presumed severe demographic shift is believed to have severe consequences for the health sectors in contemporary societies, AAL is claimed to remove pressure from the health sector (Spitalewsky et al. 2013).

A fourth motive brought forward is the value of *health* itself. New smart technologies and Big Data are alleged to improve individual and public health substantially (Jacobsen et al. 2014), especially by constant telemonitoring of vital parameters (cf. Schmidt/Rienhoff 2013), be it in cases of patients or population segments with chronic diseases (hypertonic patient, chronic obstructive pulmonary diseases etc.) or patient with acute bodily disorders or those in periods of convalescence. Telemonitoring might, however, be used for preventive goals too (e.g., prevention of falls, heart-attack-patients) (Hilbe et al. 2010).

On the whole, we think, Big Data and AAL is not simple a promise for better health care and more safety. We believe even more that the idea of Big Data and AAL claims its own moral belief, that technology will improve the quality of life of elderly people in terms of health, safety, autonomy, and participation. Furthermore, it is supposed to have positive impact on national economy and research & development institutions. For these reasons, according to the promoters of AAL, the outlined technological efforts should be undertaken and are, indeed, without reasonable alternative. This seemingly impossibly ambitious venture requires sound empirical research and further ethical scrutiny in terms of individuals whose motivation, cooperation and collaboration as active citizens or patients are necessary. This research is also needed regarding the effects on society itself, which this notion would see transformed into an entity of ubiquitous and permanent surveillance, where digital data seem to become the most valued currency. Simply put, people become the object of biopolitics that is presumably rooted in a new form of benevolent paternalism (nudging). In the following chapter, we will view Big Data and ambient assistant living through the lens of the Foucaultian concept of biopolitics.

Big Data and AAL as a new form of biopolitics

We start off with a short description of the central tenets that constitute the Foucaultian idea of governing and managing the life of human individuals as well as human population by means of technologies of power, technologies of the self and biopolitical strategies deployed in order to shape the conduct of life – individual or collective. Biopolitics according to Foucault is basically the art of guiding, educating and administering individuals and collectives by means of specific knowledge based technologies. All of them are, in one way or another, *technologies of power*, i.e., they constitute asymmetrical social relations between those who exert and those who are subject to power, although these power relations may not be visible. Foucault furthermore makes an analytical distinction regarding the fact that governmentality refers either to the individual or to a population, defined and segmented according to knowledge-based criteria. (1) In view of the former, technologies of power discipline, order, educate and control embodied human subjects. Although Foucault is not always entirely clear on this point, these technologies of power are to be conceived as enabling mechanisms. They constitute, form and configure the subjects according to the desires and interests by those in power or respective positions. In his later work, Foucault introduces the concept of technologies of the self. These are mechanism of power that introduce capacities for self-education and capabilities for self-management. Biopower and biopolitics, distinct from (repressive) sovereign and legal instruments of power and domination, is a comparatively recent form of administering and governing human individuals and collectives (Foucault 1983). As Foucault has shown, the modern state and its various agencies have increasingly focused on life (*bios*) as a central resource of power (Foucault 1983, 2005; Foucault et al. 1993). Biopower and biopolitics became decisive mechanisms of forming and transforming *subjects*, their bodily features, body techniques, complex practices and everyday routines (Foucault 1983; see also Lemke 2007). Yet, biopolitics not only targets the individual human body and its capabilities and performances, but also defines and categorizes population or sub-segments to be governed. In the 19th and 20th century, the emerging health sector, e.g., the different hygiene policies of modern states, became a core arena of the politics of life (Foucault 1996). The development and expansion of biopolitics goes hand in hand with the growing scientification of modern societies. In fact, scientific knowledge is the main resource for the constitution and legitimation of biopolitical strategies and their objectives. Foucault's main point, however, is hardly to show that health policy has become a separate field of politics. Rather it is to underscore that by invoking health, biopolitics has permeated more and more societal fields, collective practices as well as individual routines in everyday life. Some even believe that the modern subject is in various respects a product of health related forms of biopolitics (Sarasin 2001).

(2) Populations or subpopulations first need to be identified and classified in order to become a malleable subject of technologies of power and biopolitical strategies. It has become a rather common view to allege that (neo-)liberal societies have intensified and expanded the politics of life, be it to reshape the lifestyle of societal subpopulations, as in the cases of obesity and alcoholism, or to govern more and more bodily aspects and biological dispositions on genetic and molecular levels (Rose 2007; Lemke 2000). Thus, one could maintain that the Big Data-based AAL project is simply another case of modern biopolitics, focusing on specific subpopulations to which certain features are ascribed by means of (evidence-based) scientific knowledge.

Even though the all-encompassing application of Big Data in AAL at least in its interconnected form is still (promising) science fiction, many research projects have begun to move in this direction. Thus, looking at the emerging field of data-based and -driven ambient technologies at an early stage makes for an interesting topic of social scientific and ethical investigation. Also, the promise of enhancing the quality of life of an aging population is not only a highly legitimate consideration, but also a resource for gaining research funds and economic return. What is the central concept behind the science based political economy of good life in AAL sites?

Data driven biopolitics of body, space and population

What insights may be achieved looking upon Big Data in AAL from a Foucaultian perspective? Though a general trend toward Big Data in rapidly globalizing societies can hardly be disavowed, one first has to acknowledge that Big Data in AAL residences, at least in Germany, is more a technological vision than reality. At the same

time, it invokes the values of modern liberal societies, such as autonomy, health, safety and the right to a self-reliant life. In this sense, one could even literally speak of a "moral economy" (Fassin 2005) of data based ambient living as a biopolitical conduct of conducts. Big Data in AAL offers a technological solution that seems to reconcile societal needs in terms of demographic pressure, the core values of liberal society and its guiding principle of a long, independent and healthy life as well as the interest and demands of a variety of social actors. In the following we list a few crucial issues that characterize the concept of Big Data in AAL that should be focused on in further public deliberation.

(1) Knowledge production: a first issue that needs further discussion are the characteristics of *digital knowledge production* itself. To be able to cope with self-set tasks, Big Data biopolitics must be multidirectional in character, not only in regard to the technological infrastructure required but also the coordination of actors involved in AAL projects, starting from caregivers, physicians, relatives, online-services, call centers, transport services and the residents themselves. With regard to the *individual,* we must explore the way Big Data based AAL projects target and shape the human body by asking what kind and quality of data is typically processed. To the extent that the individual's body is in the center of Big Data in AAL, biopolitics is clearly a form of politics of life, addressing both the *zoë* (biological life) and the *bios* (political life). With respect to the former, all vital data which related to disease, illness) and health are of major interest for data processing including the subsequent decision making, diagnosis and therapeutical strategies. We do not know much about the truth games and the ways future "raw data" (which in itself might be a very misleading term) will be aggregated, how the vital parameters will be categorized, by whom and for what reasons and to what ends. This affects especially the highly malleable modes of sorting information in or out. Neither do we know much about what kind of academic knowledge concepts will feed and shape the process of knowledge production, configuration and categorization. Third, we know even less about the (economic and health-political) interests of the actors (care providers, etc.) who will translate these scientific knowledge patterns into action. All in all these are vitals concerns of a democratic society which should be deliberated publicly and on time.

With regard to the organization of space, it is clear that Big Data governmentality does not end at the boundaries of the individual human body; instead, it has an eye on the practices the individual is embedded in. ICT in AAL also allow for the large surveillance of residential spaces and adjacent areas by means of context sensitive motion sensors, gateways, wearable technologies or web-based devices which passively monitor or actively instruct the user. With respect to this AAL extends surveillance to the entire living situation of elderly residents or disabled citizens. This will make the consensus of users necessary which tends to be more than a complex task. Big Data is at the same time opening up and promising opportunities by means of collecting data on body and space while introducing data generating devices in the lived spaces (as well as in and on bodies) and thereby shaping these spaces people live and move in.

This is not only a question of data safety in databases, clouds or on platforms and networks as such, but also concerns the say of the clients that in reality provide their private data to contribute to scientific knowledge production. Thus, besides safety related questions the topic of appropriate participation occurs, since AAL de facto transforms private residences into living labs. Big Data producer can hardly be conceived as an impartial and disinterested provider of evidence-based knowledge whose only intention is to enhance and maintain AAL user's health status. What is happening or will happen in Big Data based AAL projects is the performance of what Krohn/Groß/Hoffmann-Riem (2005) call real life experiments that is scientific experiments carried beyond the laboratory walls. With respect to that we have to take into account that Big Data in AAL is a powerful science based real life experiment. The embodied individuals are by no means mere clients of health care services. They are – wittingly or not – the subjects of research and knowledge production, a "privilege" they even have to pay for.

(2) Normalization and scientification: Big Data is a new biopolitical strategy that leads to an increased scientification of everyday life, but still, taking the example of contemporary health apps, there is also a considerable chance that it will oscillate between science and pseudo-scientific gamification and infotainment. Among the central problems are thus the quality and accountability of data and the processes of data mining and aggregation, the relevance of the categories in use, and the quality of the conclusions drawn from the data stocks by caregivers, insurance companies, health policy, public and private services, housing companies, the individual user and their relatives. It can also be expected that Big Data based biopolitics will cause a further

standardization of health/disease concepts and lifestyle or activity profiles. Big Data, in this sense, may lead to an enhanced process of normalization in future (Link 2013), i.e., it will enhance the process of creating images of the "normal human body" by means of aggregated statistical data from multiple sources. In order to do so, residents in AAL sites need to be equipped with sophisticated monitoring and communication devices that generate appropriate data streams. It will be of interest to see which kind of data will be really collected and used in future and why. No Big Data in AAL infrastructure can function beyond the categories established by the (bio-)medical and scientific concepts, practices and institutions they serve. Therefore, one has to carefully scrutinize and evaluate which kind of concepts of disease, illness and health will be used to select, aggregate and interpret Big Data records. Biopolitics, and here we follow Fassin (2005), always has a normative element. A normative rationale is also inherent in the scientific concepts used by (private) care providers, physicians as well as health policymakers. This poses the question: what are the consequences of standardizing and typifying health profiles, be it vital parameter clusters or personal activity patterns or collective habits, routines and practices? How can the extensive datafication of everyday life be legitimized?

(3) Virtual duplicate: a third important issue seems to be the fact that Big Data allows for a doubling of identities. Big Data is obviously based on real-life-persons. But beyond these identities we now have to deal with another identity, the digital duplicate created by Big Data. This is what Cheney-Lippold (2011) has coined an "algorithmic identity" which is the construction of net-based soft biopolitics as a new form of conduct of conducts. Besides questions of privacy, liability and data security, the more fundamental question of the form of subjectivity involved in regimes of algorithmic governmentality occurs. Big Data streams allow for the constitution of a *digital other* that informs, mirrors or shapes the flesh-and-blood selves of the citizens which is very difficult to interpret independently. This data can or even should be used by the resident for self-observation and reflection. In this sense, Big Data provides concepts for "technologies of the self" (Foucault). We expect the virtual body to become a central instrument of biopolitics inasmuch as it allows for a (popular) science-based *normalization* (Link 2013) of aging bodies. This virtual duplicate is presumably not fully visible or even accessible at all to the user. It seems to lead its own life, which is in fact a misleading idea provided that there is somebody who has full access and full control of the data. The emergence of the virtual duplicate is an ambivalent category, a mirror and means of self-inspection and self-monitoring, but it is also a new source of power and – legally speaking – a threat to the right to informational self-determination.

(4) Categorization of collectives: Big Data seemingly offers new opportunities for the definition, (re-)categorization, medicalization and politicization of specific (sub-)populations which can be classified and even treated as deviant. The surveillance and the monitoring of health related routines, movement profiles and nutritional habits of these collectivities becomes a further focus of algorithmic biopolitics. The crucial question is how Big Data mining and aggregation works in regard to the construction of health and lifestyle-related patterns? According to which criteria will a differentiation between an appropriate and a socially unacceptable (and potentially punishable) lifestyle be made? We assume the ways vital parameters digitally constructed of group members, the respective embodied person, (digital and real) surroundings, devices and collective actors will be interconnected to become a central source of biopower. Whether this will be a source of individual empowerment or control and subjugation is an empirical question to be answered by future developments and research.

(5) Transparency and data security: fifth issue to be addressed is the transparency and safety of data related to the storage, circulation and its accessibility. Big Data provides information about everyday life and thus expands control over the entire lifestyle. The benefits promoters of AAL projects expect significant changes in the health status of citizens. Big Data generated by sensor data and their visualization is supposed to provide a better basis for medical decision and forms a necessary prerequisite for future telemedical settings (Jacobson et al. 2013). With regard to this Knaup and Schöpe (2014: 150) assume that the provision of surveillance data from residential areas to physicians and care providers may need standardized communication technologies and specific organizational entities in order to manage the huge amounts of data, such as telemedical centers. On the other hand, using computer and web-based devices requires at least some appropriate training on the side of senior or disabled citizens. Another benefit alleged concerns the advantages of Big Data for the establishment of large scale (regional) health information centers (Gietzelt et al. 2014). Integrated sensor based networks in AAL are no longer document-based and provide continuous data streams, but the integration of multiple sensor data requires sophisticated coding, models and methods for interpreting, standardizing and

contextualizing the home site data. What holds true for the constant monitoring of vital data on individual bodies is also true for the permanent surveillance of home sites and adjacent areas. The emerging forms of biopolitics in AAL are based on algorithms, but they also seem to require new institutional forms of processing, managing and storing data. This raises the questions of data protection, transparency and legitimate access to personal data. Both the individual embodied persons and their habits and lifestyle get a digital companion with his own history. Who has legitimate access to data, profiles and why?

(6) Last but not least, the question whether the AAL really enhances the autonomy of the aging or disabled citizens should be raised in the context of governmentality (Helbing et al. 2015). Therefore, the focus should be concentrated not only on the flesh-and-blood subject and populations but also on the technological, organizational and institutional infrastructures AAL residents live in. Although AAL is considered to be rooted in the voluntary cooperation of users with service providers and administrative bodies, the new forms of permanent monitoring of life and lifestyles might also create illusions of autonomy. One could ask: to what extent and with which consequence is there a benevolent paternalism at work in algorithmic governmentality and its increasing possibilities of governing at distance? *Algorithmic governmentality* focusses on the remote monitoring of the individual body, the personal or group-related lifestyle and the residential spaces citizens in need of (technological) support live in. Big Data in this respect presumably constitutes a form of soft biopolitics exercising a benevolent control over individuals or populations. This emerging benevolent paternalism is grounded in a "moral economy" invoking the ethos of a good life. However, if the algorithmic governmentality appears to be a benevolent one, why then should one take issue with the concept at all? To put it bluntly: Big Data biopolitics, donning its benevolent digital furs, tends to camouflage the power relations it implies. In this regard, the topic of legitimization and control by non-political authorities and organizational bodies should not be neglected. With respect to the Big Data generated technologies of the self which inform senior citizens about the perils of their individual lifestyle or what they must take notice of in and around their living space, it can at least be doubted whether the individual, as Cheney-Lippold (2011: 176) puts it, is capable of "really experiencing the effect that algorithms have in determining one's life". The naturalization of technological environments in AAL residential sites camouflages the fact that a multitude of (organizational) actors in distant locations contribute to the calculations, cybernetic categorizations and (medical) decision-making. The task is to acknowledge and make subject of discussion the networks of power senior residents and other needy citizens live in. What is therefore necessary is an open ethical reflection on the implications, possible (side-) effects and consequences of the biopolitics of Big Data in AAL projects at this early stage of technology development and implementation.

Conclusion: Towards a technology-based moral economy of life and lifestyle?

We have discussed the possible impact of Big Data in AAL on the individual body as well as on collectivities. The use of Big Data and smart information technologies in homes and wider residential areas constitute a new form of biopolitics which on the one hand actively involves individual embodied persons, and on the other hand constitutes new collectivities to which physicians, health care providers, public managers and political decision-makers attribute information and data which characterize them as deviant from parameters and margins set by statistical normal distribution curves. They consequently become subjects of major interest for health care providers, public services, administrators and bioscience as well. What is more, they also become subjects of (permanent) surveillance and intervention while 'living in a lab' wittingly or unwittingly. We have therefore pointed out that Big Data in AAL can be comprehended as a further step in the scientification and datafication of everyday life. The implementation of in-home and mobile sensor networks allows for a provision of continuous data streams that feed the data banks of individual care providers as well as organizational entities. All in all, this will constitute a vulnerable infrastructure filled up with intimate data which will awaken covetous motives among insurance-companies, employers etc. The generation and storage of Big Data in AAL leads to what we call a digital duplicate, a virtual copy of embodied persons based on sophisticated, standardized communication technologies. In this sense, Big Data and coded computer algorithms do not only allow for a sophisticated modulation of control, but contribute to a further *normalization* (Link 2013) by making use of

seemingly private vital data as well as information on everyday routines and of citizens in private households and larger residential areas. But isn't there more to Big Data / algorithmic biopolitics?

Though datafication of AAL goes hand in hand with normalization, permanent surveillance and the creation of a virtual duplicate, however, this view on biopolitics still underscores the strong moral and ethical implications lying behind the idea of AAL. It seems important to us to raise the question whether the blend of Big Data and AAL can be seen as a digital moral economy of lifestyle. This is, as Fassin (2005) puts it, a subtle way of governing human bodies and lifestyles by making health and bodily integrity the core values of legitimation. The individuals and sub-populations are not only asked to take care of their own body and well-being, they need to do so for the well-being of the community and society. Big Data in AAL could be an interesting case to show that we are moving from a politics of life itself to a sophisticated science-based moral economy of good life based on the health and the promises of permanent autonomy. However, this still is a form of conduct of conducts imposed by the Big Data driven biopolitics, even if this occurs in subtle ways.

References

Cheney-Lippold, John: A New Algorithmic Identity. Soft Biopolitics and the Modulation of Control. In: Theory, Culture & Society 2011, Vol. 28 (6), S. 164-181

Demiris, George: Smart homes and ambient assisted living in an aging society. New opportunities and challenges for biomedical informatics. In: Methods of Information in Medicine 2008, 47 (1), 56-57

Demiris, George/Thompson, Hilaire: Smart homes and ambient assisted living applications: from data to knowledge-empowering or overwhelming older adults? Contribution of the IMIA Smart Homes and Ambient Assisted Living Working Group. IMIA Yearbook of Medical Informatics 2011, 6(1), 51-57

Driscoll, Kevin: From Punched Cards to Big Data: A social History of Database Populism. In: communication +1 2012, Vol. 1, Article 4, http://scholarworks.umass.edu/cpo/vol1/iss1/4/. (accessible 20.01.2016)

Fassin, Didier: Compassion and Repression: The Moral Economy of Immigration Politics in France. In: Cultural Anthropology 2005, Vol. 20, Issue 3, 362-387

Foucault, Michel: Der Wille zum Wissen. Sexualität und Wahrheit 1. Frankfurt am Main, Suhrkamp 1983

Foucault, Michel: Die Politik der Gesundheit im 18. Jahrhundert. In: Österreichische Zeitschrift für Geschichtswissenschaften 1996. 7. Jg. Heft 3, Kulturen der Krankheit, S. 311-326

Foucault, Michel: Analytik der Macht. Frankfurt am Main, Suhrkamp 2005

Foucault, Michel/Martin,Rux/Martin, Luther H./Paden, William E./Rothwell, Kenneth S./Gutman, Huck/Hutton, Patrick H.: Technologien des Selbst. Frankfurt am Main, Fischer 1993

Gietzelt, Matthias/von Bargen, Tobias/Kohlmann, Martin/Marschollek, Michael/Schwartze, Jonas/Song, B./Wagner, M./Wolf, Klaus-Hendrik/Haux, Reinhold: Home-centered Health-enabling Technologies and Regional Health Information Systems – An Integration Approach Based on International Standards. Methods of Information in Medicine 2014, 53(3), 160-166

Groves, Peter/Kayyali, Basel/Knott, David/Van Kuiken, Steve: The Big Data revolution in healthcare. Accelerating value and innovation. Center for US Health System Reform, Buisness Technology Office, New York, McKnisey&Company 2013

Helbing, Dirk/Frey Bruno S./Gigerenzer, Gerd/Hafen, Ernst/Hagner Michael/Hofstetter Yvonne/van den Hoven, Jeroen/Zicari, Roberto V./Zwitter, Andrej, T-Revolution: Digitale Demokratie statt Datendiktatur. Big Data, Nudging, Verhaltenssteuerung: Droht uns die Automatisierung der Gesellschaft durch Algorithmen und künstliche Intelligenz? Ein gemeinsamer Appell zur Sicherung von Freiheit und Demokratie, Spektrum 17.12.2015

Hilbe, Johannes/Schulc, Eva/Lindner, Barbara/Them, Christa: Development and alarm threshold evaluation of a side rail integrated sensor technology for the prevention of falls. In: International Journal for Medical Informatics 2010, 79(3), 173-180

Jakobsen, Anna S./Laursen, Lars C./Ostergaard, Birte/ Rydahl-Hansen, Susan/Phanareth, Klaus V.: Hospital-admitted COPD patients treated at home using telemedicine technology in The Virtual Hospital Trial: methods of randomized effectivness trial. In: Trials 2013, 14(1), 280

Knaup, Petra./Schöpe, Lothar Using Data from Ambient Assisted Living and Smart Homes in Electronic Health Records. In: Methods of Information in Medicine 3/2014, 149-151.

Koselleck, Reinhart: Erfahrungsraum und Erwartungshorizont – zwei historische Kategorien. In: Koselleck, Reinhart. Vergangene Zukunft. Zur Semantik geschichtlicher Zeiten. Frankfurt/Main, Suhrkamp 1989, 349–375

Krohn, Wolfgang/Groß, Matthias/Hoffmann-Riem, Holger: Realexperimente. Ökologische Gestaltungsprozesse in der Wissensgesellschaft. Bielefeld, transcript 2005

Lemke, Thomas: Die Regierung der Risiken. Von der Eugenik zur genetischen Gouvernementalität. In: Ulrich Bröckling/Susanne Krasmann/Thomas Lemke (Hrsg.), Gouvernementalität der Gegenwart. Frankfurt am Main, Suhrkamp 2000, S. 227-264

Lemke, Thomas: Biopolitik zur Einführung. Hamburg: Junius 2007

Leone, Alessandro/Diraco, Giovanni/Siciliano, Pietro: Detecting falls with 3D range camera in ambient assisted living applications: a preliminary study. In: Med. Eng. Phys. 2011, 33 (6), 770-781

Link, Jürgen: Versuch über den Normalismus. Wie Normalität produziert wird. Göttingen, Vandenhoeck & Ruprecht 2013

Lupton, Deborah: Self-tracking Modes: Self-Monitoring and Data Practices. `Imminent Citizenship: Personhood and Identity Politics in the Informatic Age´ workshop, Canberra, Australia, August 27th 2014

Manzeschke, Arne/Weber, Karsten/Rother, Elisabeth/Fangerau, Heiner: Results of the study »Ethical questions in the area of age-appropriate assisting systems«, Berlin 2015

Nitzsche, Thomas/Thiele, S./Häber, Anke/Winter, Alfred: Communication Architecture for AAL – Supporting Patient Care by Health Care Providers in AAL-enhanced Living Quarters. In: Methods Inf Med 2014, 53(3), 167-174

Reeder, Blaine/Chung, J./Le, T./Thompson, Hilaire/Demiris, George: Assessing Older Adults` Perceptions of Sensor Data and Designing Visual Displays for Ambient Environments. In: Methods of Information in Medicine 2014, 53(3), 152-159

Rose, Nicolas: The Politics of Life Itself. Biomedicine, Power, and Subjectivity in the Twenty-First Century. Princeton/Oxford, Princeton University Press 2007

Sarasin, Phillip: Reizbare Maschinen. Eine Geschichte des Körpers 1765-1914. Frankfurt am Main, Suhrkamp 2001

Schmidt, Silke/Rienhoff, Otto: Interdisciplinary Assessment of Permanent Health Monitoring, Amsterdam/Berlin//Tokyo/Washington DC: IOS-Press 2012

Spitalewsky, Katharina/Rochon, J./Ganzinger, Matthias/Knaup, Petra: Potentials and requirements of IT for ambient assisted living technologies. Results of a Delphi study. In: Methods of Information in Medicine 2013, 52(3), 231-238

TAB (Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag): Schwerpunkt: Hope-, Hype- und Fear-Technologien. TAB-Brief 39/2011

van Dijck, José: Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. In: Surveillance & Society 2014, 12(2), 197-208

Vimarlund, Vivian/Wass, S.: Big Data, Smart Homes and Ambient Assisted Living. In: IMIA Yearbook of Medical Informatics 2014, 143-149

Weber, Max: Die protestantische Ethik I. Eine Aufsatzsammlung. Gütersloh, J.C.B. Mohr (Paul Siebeck) 1981

Weingart, Peter: Verwissenschaftlichung der Gesellschaft - Politisierung der Wissenschaft. In: Zeitschrift für Soziologie Jg. 12, Heft 3 (1983), 225-241

Wichert, Reiner/Klausing, Helmut (Hrsg.): Ambient Assisted Living: 7. AAL-Kongress 2014 Berlin, Germany, January 21-22 2014. Heidelberg/New York/Dordrecht/London, Springer 2015

Zwitter, Andrej J./Hadfield, Amelia: Governing Big Data. In: Politics and Governance 2014, Vol. 2, Issue 1, 1-3

Harald Weston:

Data Analytics as Predictor of Character or Virtues, and the Risks to Autonomy

Abstract:

Can we measure and predict character with predictive analytics so a business can better assess, ideally objectively, whether to lend money or extend credit to that person, beyond current objective measures of credit scores (when available) and standard financial metrics like solvency and debt ratios? We and the analysts probably do not know enough about character to try to measure it, though it might be more useful to measure and predict a person's temperance and prudence as virtues, or self-control as psychology, or sense of obligation, particularly a moral commitment or sense of duty to honor a contract and re-pay a loan. The pervasive data surveillance of people that goes with "big data" and predictive analytics is not only an invasion of privacy in general, but an impairment of the aspect of privacy called autonomy that will constrict and alter a person's choices and development of self.

Agenda:

Author:

Harold Weston

- Clinical Associate Professor, Department of Risk Management and Insurance, J. Mack Robinson College of Business, Georgia State University, Atlanta, Georgia
- ☎ + 1 - 404 – 413 – 7480 , ✉ hewston@gsu.edu

ISSN 1614-1687

Introduction

The question of measuring character comes from a *New York Times* article with the catchy title, *Determining Character with Algorithms*, which reported on two companies using data analytics to determine what they loosely call character in assessing a potential borrower. SAT scores, grade points averages, colleges and majors, even longevity of a cell phone number are used. "The idea, validated by data, is that people who did things like double-checking the homework or studying extra in case there was a pop quiz are thorough and likely to honor their debts." (Hardy, 2015). Trying to score for character is a clever idea.

Separately, companies use algorithms for hiring (Miller, 2015a). The contention is that these data points are less subjective measures than the usual methods of hiring, which can often lead to claims of discrimination or favoritism (Lam, 2015). Probably so, though some bias is inevitable because an algorithm is written by a human, as a Microsoft programmer notes (Miller, 2015b). This is not the first time people have tried to quantify what has previously been a subjective exercise. A few years earlier, *The Times* noted that people on dates were inquiring of each other's credit scores early in the relationship to assess the potential financial costs that the other might bring to the relationship (Silver-Greenberg, 2012). This seems to be an earlier use of character scoring with an existing metric.

This raises the question whether character can be scored. This essay contends that character is too broad and flexible to be measured, or is being measured by this predictive analytics. Rather, something more specific like the virtue of prudence or temperance is being measured, or something different like following rules and honoring promises should be the target. Whatever is being predicted, we should expect there will be an effect upon an person's autonomy knowing that all past transactions and data points are being evaluated, either in constraining choices, or trying to game the system for a better score, or foreclosing personal development.

Credit Scores And Non-Credit Measures From Data

Credit scores have been used by lenders, employers and landlords to help assess the risk and reliability of prospective borrowers, employees and tenants. A similar insurance score is used by insurers to help assess risk of prospective insureds (Halon and Boyd, 1996). Credit scores, called FICO scores because based on the proprietary algorithm invented by Fair Isaac & Co in 1950 (thus FICO), use past bill-paying practices and utilization of credit by the individual (Boulard, 2004); the scores are considered good predictors of risk of the individual as borrower, employee, tenant, or for losses as an insured (FTC, 2007; Brockett and Golden, 2007). FICO scores are subject to some problems such as unfair practices by the credit reporting agencies including failure to investigate disputes (Siegal, 2015), and are subject to controversies, including mysteries about how scores are determined, and declines in scores due to potential creditors merely inquiring about the scores or the status of a mortgage (Ritzholz, 2010), and due to uncovered medical expenses, (CFPB, 2014; Rosenthal, 2014), although revisions to the algorithm are supposed to reduce the impact of medical losses (Fair Isaac, 2014; Andriotis, 2014). The collection and use of credit scores is governed by the Fair Credit Reporting Act, and enforced by the United State Federal Trade Commission, and to a lesser extent by individuals affected by the score through lawsuits under the FCRA. The use by insurers of a slightly different "insurance score" is governed by the federal law, and by state law overseen by state insurance commissioners.

Information gleaned from consumers' use of the internet and on-line shopping, and other publicly available information such as voter registration, frequent shopper or loyalty cards, giving e-mails to businesses, restaurant reservations, types of computers, geolocation applications on smart phones and license plate readers[60], and of course criminal records, are all data sets. This is called "data exhaust" (Davenport, 2013). Data brokers compile this information and create their own scores, which are not subject to the FCRA (FTC 2012, FTC 2014a), though these scores have sometimes been offered to and used by employers and landlords

[60] The police are doing this and storing the information, as are repo crews (Atiyeh, 2014, EPIC, Lynch 2014).

in violation of the FCRA (FTC, 2014b, Wyatt, 2012)[61]. The credit reporting agencies – TransUnion, Experian, Equifax – which are subject to the FCRA, also have sideline businesses that are competitive with data brokers that "create and sell 'consumer evaluation,' 'buying power' and 'marketing' scores, which are ingeniously devised to evade the FCRA (a 2011 presentation by FICO and Equifax's IXI Services was titled Enhancing Your Marketing Effectiveness and Decisions With Non-Regulated Data'). The algorithms behind these scores are designed to predict spending and whether prospective customers will be moneymakers or money-losers." (Taylor and Sadowski, 2015; and see Cohen, 2013: 1916). The database marketing company Acxiom (one of many data brokers) reviews 50 trillion data transactions yearly on 500 million consumers worldwide, including 190 million Americans, with 1,500 data points per person based publicly available information like home valuation and voter registration to Internet usage to create precise profiles on consumers, ranking consumers from "high-value prospects, to be offered marketing deals and discounts regularly, while dismissing others as low-value — known in industry slang as 'waste.'" (Singer, 2012).

The ability to predict customer's interests and financial risks is captured in the infamous story in *The New York Times Magazine* that retailer Target used predictive analytics to determine which young female customers were likely pregnant, even before the women knew, to then offer them coupons and hook them into Target for their maternity and baby clothing and care needs (Duhigg, 2012). Other risks and critiques of big data have been addressed in many articles.[62]

Character or Behavioralism as the Metric

If the trait to be measured directly or indirectly with data analytics for risk assessment in lending is character, we need to look a little at what is character. "Character is the whole of which the virtues are some of the components; but a character trait can be a virtue or a vice depending on the circumstances," writes Hartman (1998: 550). He gives the example of self-confidence: when acting despite peer pressure this is virtuous, while when against good arguments it is vice. Hartman describes character as "one's standard pattern of thought and action with respect to one's own and others' well-being and other important concerns and commitments." (Hartman, 2007: 316). Sherman (1989: 1) describes character as "a person's enduring traits; that is, with the attitudes, sensibilities, and beliefs that affect how a person sees, acts, and indeed lives. As permanent states, these will explain not merely why someone acted this way *now*, but why someone can be *counted on* to act in certain ways." Wilson defines character as "empathy and self-control" (Wilson, 1991: 5).

Character is developed from the practice of virtues, which "are traits of character that constitute praiseworthy elements in a person's psychology. To have a virtue is to have a praiseworthy character trait appropriate to pursuing the particular kind of good with respect to which the trait counts as a virtue." (Audi, 2012: 273; similarly Feldman, 2000: 1438). As Philippa Foot wrote, "… virtues are in general beneficial characteristics, and indeed ones that a human being needs to have, for his own sake and that of his fellows." (1978: 107). Aristotle says that doing virtuous acts does not alone make the person virtuous, the acts must be done virtuously. "The agent also must be in a certain condition when he does them; in the first place he must have knowledge, secondly he must choose the acts, and choose them for their own sakes, thirdly his action must proceed form a firm and unchangeable character." (§ 1105a.). Wilson (1985:15) describes virtue as:

> *"habits of moderate action; more specifically, acting with due restraint on one's* impulses*, due regard for the rights of others, and reasonable concern for distant consequences. […] Virtue is not learned by precept, however; it is learned by the regular repetition of right actions. We are induced to do the right thing with*

[61] The collection and distribution of such information outside of the Fair Credit Reporting Act and the standing of a person to sue over that use is a question before the United States Supreme Court in *Spokeo, Inc. v. Robins*, No. 13-3339 (certiorari granted April 27, 2015, on appeal from 742 F.3d, 409 (9th Cir., 2014)).

[62] Michael Lewis (2015) has wryly suggested that some type of analytics be used by universities – Harvard – to predict who among the applicants is likely to be worth a few billion dollars years after graduation, so Harvard doesn't again turn down the next Steven Schwarzman, who is worth $13 billion and whom Harvard turned down as a student.

respect to small matters, and in time we persist in doing the right thing because now we have come to take pleasure in it. By acting rightly with respect to small things, we are more likely to act rightly with respect to large ones."

Although character develops through practice into habits, it allows – even requires – that sometimes the right thing to do might be wrong in most circumstances. This is Aristotle's "practical wisdom." Consider the person who reliably pays her bills, lives within her budget, saves enough for the occasional indulgence. The virtue of temperance is shown here. Yet now our person's daughter needs expensive medical care, and this being the United States the out-of-pocket medical expense will be thousands of dollars.[63] Many news stories report that people with high deductible medical plans in the U.S. forego or delay medical treatment (e.g. Abelson, 2015). Does the person of character tell the daughter, "Too bad, I have bills enough to pay, we're out of money, you're out of luck"? Or does this person say, "My daughter needs medical care to live, everyone else will get paid later"? This is the modern equivalent of the classic ethical problem of whether it is right for the poor person to steal food and medicine to survive. But in this example there is no theft, only unpaid and un-payable bills: a common dilemma for Americans, for whom 33% put off medical care due to the cost in 2014 (Riffkin, 2014) with a higher percentage in 2013 (Commonwealth Fund, 2013[64]), and many are unable to pay their medical expenses (Cohen and Kirzinger, 2014; Kaplan, 2014).

The ethical problem can be extended to the mother proceeding with the necessary medical care knowing she will be unable to pay the medical bill. From the legal perspective, not paying the existing creditors is breach of contract, and entering into a contract for medical care knowing one cannot pay the bill is fraud and deceit. From a business point of view trying to predict a customer's likelihood to pay his or her bills, it is the legal and financial issues to be forecast, not the ethical, although some psychological assessment can be useful here too (see Ding, Chang and Liu, 2006, discussed further below.)

The virtues of temperance and prudence might be the proper category to measure for a borrower's financial risk. A virtue is the mean between the excess and deficiency (Aristotle, §1106A-1108; Gottlieb, 2009: 19-20). The person who is temperate in most things in life probably does not spend money excessively, thus should be able to handle money and repay any loan. The temperate person will not be extravagant, and will control his or her passions, and will likely act with deliberation rather than impulse. Predictive analytics has shown that people who buy felt pads for the bottom of the chair legs, or buy carbon monoxide detectors for their home, or buy premium birdseed, are better financial risks, while people who frequent particular bars or place charges on their credit cards for marriage therapy or pawnshops are worse risks (Duhigg, 2009). Academic studies have found that male business executives who are unmarried or on the third wife pursue more aggressive investment risks (Nicolosi and Yore, 2015), and an executive's prior legal infractions, such as traffic violations, driving under the influence of alcohol, domestic violence, reckless behavior, disturbing the peace, have a positive correlation

[63] The economists and policy makers have previously provided incentives to Americans to control their unnecessary or discretionary medical expenses by creating high deductible medical plans, and then providing tax-free incentives to save for medical expense eventualities through health saving accounts, $6,550 in 2014 and $6,650 in 2015. This assumes Americans earn enough to save this amount, plus save for retirement, a doubtful proposition given that the median income for all families in 2013 was $46,7000 (Fed Reserve Bulletin, 2014) and median retirement savings in 2014 for households age 55-64 was $104,000 (GAO Retirement Survey, 2015).

[64] "More than one-third (37%) of U.S. adults went without recommended care, did not see a doctor when sick, or failed to fill prescriptions because of costs, compared to as few as 4 percent to 6 percent in the U.K. and Sweden, according to the study published today as a Health Affairs Web First article. In addition, nearly a quarter (23%) of U.S. adults either had serious problems paying medical bills or were unable to pay them, compared to less than 13 percent of adults in France, the next-highest country, and 6 percent or fewer in the U.K., Sweden, and Norway. About two of five (41%) U.S. adults spent $1,000 or more out-of-pocket for care in the past year—by far the highest rate of any country surveyed.

"Uninsured adults in the U.S. were the most likely to struggle to afford health care. However, even U.S. adults insured all year were more likely than adults in other countries to forgo care because of costs, to struggle with medical bills, and to face high out-of-pocket costs, with 42 percent paying $1,000 or more out-of-pocket for medical care. According to the study, U.S. health insurance has higher deductibles and higher cost-sharing, and does not place limits on out-of-pocket costs. This potentially explains why even people with health insurance in the U.S. struggle to afford needed health care." (Commonwealth Fund, 2013).

with a propensity to perpetrate fraud (Davidson, Dey and Smith, 2012). Another way to describe this is impulse control, exemplified in the famous marshmallow experiments and their follow up studies (Konnikova, 2014).

As humans have long done, we look to a person's friends to help assess the person. Do you associate with quality people or scoundrels? Facebook extends its predictive analytics to evaluate the credit scores of a person's friends to help assess this persons' creditworthiness (LaFrance, 2015).

In psychological terms, temperance looks much like self-control, which does correlate with risk (e.g. Limerick & Peltier, 2014; Fischer, Kastenmüller & Asal, 2012). A behavioral economics study was able to correlate prudence and temperance with risk (Ebert & Wiesen, 2014).

Impulsiveness and impulse control are psychological traits, which can also be examined as behavioralism. Behavioralism, like character, evinces a person's tendency to act in predictable ways. Behavioralism is a response, perhaps trained through incentives and rewards. Studies on self-control failures explain why people act impulsively and take on too much debt (Limerick & Peltier, 2014; Baumeister, 2002). Credit card companies study such psychology to assess lending risk (Duhigg, 2012; Gathergood, 2012). Thus impulse control as psychology can be behavioralism, and can also be the virtue of temperance or reason. This is an important distinction for data analytics and the measurement of character, because it seems that the data wranglers seek to measure propensity and constancy, but mostly as an aspect of behaviorialism. The person who reliably pays his or her bills on time because of fear of a bad credit score, or the converse reaction of desire to maintain a good credit score, acts under different motivation than the person who pays his or her bills on time because it is the right thing to do as a moral obligation – the obligation of contracts and promises. The whistleblower who acts for the reward shows behavioralism, while the whistleblower who acts to expose wrongdoing regardless of gain or loss shows ethics. Hartman (1998: 548) gives the example of the person who acts from courage because it is the right thing (character), in contrast to the person who does exactly the same thing because the economic incentives guide him without a bit of courage (behavioralism). In virtue ethics, "there is the agent's motivation to act appropriately in that field …. Moral virtue requires not just good deeds, but good motives." Not just having virtue, "but action grounded in virtue" (Audi, 2012: 275).

Data analytics for credit risk seems to be scoring for propensity and self-control, which might equate with the virtues of temperance and prudence, while self-control gets closer to a measure of character. Whether using psychological or ethical terminology to describe the position, the prudent person is likely to avoid taking on too much financial or other type of risk, thus making the person a better customer for a financial lender.

Thus despite catchy news titles and loose talk, it seems unlikely that data analytics is yet scoring for character. With the crucial Aristotelian components of individual flourishing and the common good that compose and result in character, it seems a metric for character would be fore more complex.

Duty, Commitment or Promise as the Metric

Another dimension that predictive analytics could assess is a sense of duty or commitment, or possibly guilt, because this will direct the customer to fulfill the deal to repay the debt. Here we have something closer to ethics – an obligation. Ding, Change and Liu (2006: 819) examined the ethical side of people's decisions not to repay credit card debts, finding that people who have internal senses of control are more like to take responsibility and have an intent to repay credit card debt than people who believe control is external to them and thus incur the debt with little intention to pay it. Further, people with more risk-taking behaviors are likely to have lower ethical standards and expect to benefit from unethical conduct.

The sense of duty or intent creates a bridge to law in the form of contracts as to why contracts are binding. The philosophical basis of contracts as commitments is developed in two lines, not necessarily incompatible. Gordley (1991: 10-13) argues the proper theoretical underpinning of contracts is essentially Aristotelian, as explained by Aquinas, based on the virtues of truth telling in keeping promises, ad commutative justice that recognizes the bargain made and owed. Fried (2014: 4, 137-140) develops a theory of contracts as a Kantian

obligation, where a person undertakes a promise, a moral invention whereby a person creates an obligation arising out of trust and respect for others.

In this way, we might look at predictive analytics as possibly measuring the strength or propensity of persons' sense of moral obligation. That possibility seems for now to be beyond what the analytics is trying to do, but surely the concept could be measured in psychological terms as strength of rectitude or guilt. Ding, Chang and Liu (2006: 828) recommend the use of personality tests for "locus of control" and risk-taking as good predictors for lenders to assess an intention to repay. A lender would prefer a potential borrower with a high sense of control and low risk-taking. But the same measures would work against the merchant who would prefer a customer with a low sense of control and high risk-taking who will succumb to impulse purchases (see Achtziger, 2015). This is the internal conflict of predictive analytics: the consumer score that is bad for the lender is good for the merchant, and thus the incentives and enticements to the consumer are like the devil on one shoulder and the angel on the other each whispering into the consumer's ears to be good or be bad.

The predictive possibility of measuring a person's sense of obligation might be an interesting link to the philosophical and legal questions of what makes a contract binding: how strong a sense of obligation might be necessary to predict contractual fulfillment, how weak might demonstrate fraudulent intent? Further study could compare such predictive ability of commitment with the economic notion of "efficient breach" to renounce one's obligations for economic advantage regardless of commitment because it results in higher overall optimal utility for all parties. Do the economists prefer the efficiency of predictability and commitment, or the efficiency of breach for better options and utility? These questions are beyond the scope of this essay.

Implications for Privacy and Autonomy

A different dimension of contracts and promises leads to the question of autonomy, a subset of privacy.

Predictive analytics invades privacy in many ways, mostly obviously by companies holding and using all sorts of bits of personal information about a person, often without a person knowing their acts have been observed, compiled, traded and turned against the person. Privacy is a broad and multi-dimensional concept. If the concern about privacy in general is ownership and use of information (data), then we are largely stuck with a debate about spying, intellectual property, and rights; these issues have been addressed in many articles about data and privacy. The more interesting implication, which I focus on here, is the subset of privacy called autonomy (Halper, 1996: 133; Kupfer, 1987), because it affects a person's freedom to enter into contracts, and freedom of self-determination (or self-governance) and psychological development.

One aspect of autonomy is the ability to make contracts. Fried contends that the undertaking of the promise is an act of autonomy (2014:14,16, 144). A promise voluntarily (autonomously) made, with other legal factors, forms a contract, but a promise formed involuntarily is not autonomous, and contract law absolves the promise as coerced, using various legal defenses such as duress, fraud, mistake. Fried's focus is on contracts, not the philosophical problem of autonomy. He takes it as given that a person has autonomy to enter into a contract. But with predictive analytics that seek to predict, then instigate, then recommend, then initiate a contract, we should be concerned about the autonomy of the person being analyzed. In some respects, this path of prediction to contract is an algorithmic version of the charming salesman or conman. The consumer can walk away from the sales- or conman. Predictive analytics, however, follows the person around the Internet, as anyone who notices repeat advertisements knows.

Autonomy itself has many dimensions. Raz (1986:369) calls it a purposeful, self-determining, responsible agent, able to make plans and decide for oneself what to do by projection into possible futures, to control one's destiny through successive decisions. Fallon says "autonomy is largely a *descriptive* concept, which refers to people's actual condition and signifies the extent to which they are meaningfully 'self-governed' in a universe shaped by causal forces. … To be autonomous, one must be able to form a conception of the good, deliberate rationally, and act consistently with one's goals" (Fallon, 1994: 877). This sounds right, but research shows that the long-term planning to act consistently with one's goals can be a problem for people with impulsiveness. (Baumeister, 2002; Benhabib and Bisin, 2005). Christman (1988) surveys many dimensions of autonomy: the psychological

condition of self-governance; the right to be free of interference with one's choices and desires; a set of preferences; a choice of a person's life and freedom of action; the ability to have values and preferences and to choose them and approve them without manipulation, and to make such choices as a matter of utilitarian-type utility or welfare; moral agency; acting with reason and rationality (the Kantian model) or at least some cognitive awareness of ones choices. (See also Fallon, 1994: 877. This should lead to an examination of the differences among law, philosophy and economics as to the meanings of reason and rationality.) These are important distinctions that Christman addresses at length, but as concerns predictive analytics, the distinctions are mostly shades of the problem when predictive analytics seeks to sway a person's actions and choices. Thus a system that seeks to make people act on impulses or temptations, and preempt the control process people can develop to resist impulses, interferes with self-governance. "The power of big data is thus the power to use information to nudge, to persuade, to influence, and even to restrict our identities" (Richards and King, 2013: 44). (See Benhabib and Bisin, 2005, examining agents' automatic impulses against agents' control processes for consumption-saving decisions, and Baumeister, 2002 on the weak self-control showed by people who act impulsively.)

Cohen (2013: 1908) has a slightly different conception of autonomy based on a "liberal" model of precultural determination, which she rejects because of her important point that autonomy is necessarily culturally-situated.

Predictive analytics looks at a person and regardless of that person's possible choices, treats the person as having nearly made or inevitably will make certain choices. Cohen (2013: 1917) contends that such predictions are "designed to produce … a particular kind of subject … whose preferred modes of self-determination play out along predictable and profit-generating trajectories." If the prediction is correct, all should be well. But if the prediction is incorrect, or the person creates a new choice not in line with past conduct, then that rational choice (if it is a rational choice) is constrained or preempted by the business user's model. Kerr and Earle (2013) call this a preemption prediction; with a preemption prediction, predictive analytics might accurately predict the desire, or it might constrain or prevent that choice, at least insofar as the business user and the person have occasion to possibly intersect. Of course, the business faces positive and negative risks in constraining the person's autonomy: a good prospect may make a bad choice that adversely affects the business, and a bad prospect may now make a good choice that could be profitable to the business, but whom the algorithm says to avoid.

The autonomous person, in developing character, will be aware of choices and their lasting impact on his or her life, as Raz notes (1986: 371). That impact might only be realized long afterwards, when time, other events, and judgment allows one to assess impact. This is defeated with the enduring compilation of data about people. In a simple way, credit scores – an early predictive analytic -- already constrain and guide people's choices, as people try to improve their credit score. This guidance is not necessarily bad, it might even be good, but it means that the scoring system is altering choices.[65] We should expect that individuals who are aware of other predictive analytics scores on the choices offered to them may seek to improve the predictive scores by doing things that the analytics favor to improve the score. Where the individual is engaged in gaming the system, such as by doing Internet searches for things that look good for predictive scoring, or buying felt pads for the bottoms of chair legs and buying bags of premium birdseed as a previously discussed study showed, we might call it fair play, even if it creates another level of distraction and manipulation. Nevertheless, where the individual engages in conduct, or refrains from other conduct, to improve the score, this constrains autonomy and molds the individual in a way desired by the business interest. Raz (1986: 373, 378) says in general that the "choice must be free from coercion and manipulation by others," although he allows that some coercion to protect a person against harm is not necessarily bad. Cohen (2013: 1920) says, specific to big data, "it is modulation, not privacy, that poses the greater threat to innovative practice. Regimes of pervasively distributed surveillance and modulation seek to mold individual preferences and behavior in ways that reduce the

[65] People are also admonished to build their personal brands, (Peters, 1997), for which "authenticity is the key" (Hyder, 2014).

serendipity and the freedom to tinker on which innovation thrives." (The point is developed in other ways in Christen, et al., 2013.)

If such modulation improves the person, intentionally or not, we should be wary that our technology is up to the task of building character or "'normalized soul training of the Orwellian nightmare." (Cohen, 2013: 1916). More likely such feedback modulation is behaviorist, fostering internalization of new norms and constructed subjectivity (Cohen, 2013: 1917, 1924), which diminishes autonomy and individual development.

Selective Disclosure, Data Collection And The Right to Be Forgotten

A person cannot function autonomously without some control over the information about that person (Kupfer, 1987: 81-82). Beardsley (1971) calls this selective disclosure. This is a crucial problem and foundational basis of big data, collected from data exhaust: that the selective disclosure a person makes in one place becomes an aggregated enduring disclosure available to the entire commercial world. Predictive analytics runs afoul of autonomy, because autonomy requires "that others affirm the social boundaries of this self. They must grant the individual control over this movements and information about himself. They must also permit the individual to have some say in who can experience him and when." (Kupfer, 1987: 82). Further, we need privacy to try out options and rehearse our thinking without being convicted by public opinion. (Kupfer, 1987: 83). This cannot be done when our every inquiry, search and message is scanned, recorded, tagged and compiled[66] with public information and business exchanges into a predictive model of ourselves. The data is many bits of information gathered from almost anywhere on the Internet, whether put there by the consumer or not (usually not), and stored and indexed forever (Tsesis, 2014), by over 100 companies that track a person's moves through the Internet (Madrigal, 2012). This surveillance is the digital rebuke to Beardsley's (1971: 70) "selective disclosure," "the conceptual core of the norm of privacy," that set "the conditions one will be communicated *about*, much less to determine what will be *known* about one" (See also Wall, 2011). The person who gives up, fully or partially informed, some information for some commercial advantage, (Beardsley, 1971: 67) may never have expected that the information would then be sold and merged with other information to create a model and category of that person. Beardsley wrote decades before our current concern of data exhaust, digital breadcrumbs, and ersatz and misleading corporate privacy policies. Thus bits of information that might have been public but too hard to find, and thus remained essentially private, are now public and easy to find, and are gathered and distributed or at least searchable. Cohen labels this "informational capitalism" (2014: 1912). This enduring retention of previously hard-to-find discrete bits of information is the focus of the right to be forgotten, or right of erasure, enacted in the European Union (European Commission, Right to be Forgotten). The E.U. Data Protection Directive, 96/46EC, states in its preamble that the object of the data protection laws is to protect privacy and freedoms of natural person:

> *"(2) Whereas data-processing systems are designed to serve man; whereas they must, whatever the nationality or residence of natural persons, respect their fundamental rights and freedoms, notably the right to privacy, and contribute to economic and social progress, trade expansion and the well-being of individuals;*
>
> *[…]*
>
> *(10) Whereas the object of the national laws on the processing of personal data is to protect fundamental rights and freedoms, notably the right to privacy, which is recognized both in Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms and in the general principles of Community law; whereas, for that reason, the approximation of those laws must not result in any lessening of the protection they afford but must, on the contrary, seek to ensure a high level of protection in the Community;"*

[66] See for example, Herold, 2014, on Google "reading" e-mails.

Citing to these and other provisions of the Directive, and the right of access by persons to erase or block data that does not comply with the Directive, the European Court of Justice ordered Google Spain to remove links to particular information as violating the Directive (Google Spain v. Agencia, 2014[67]). The French Commission nationale de l'informatique et des libertés made a similar order in 2015 (CNIL 2015). Thus if our every transgression is indexed and retrievable, then our future autonomy of a developed self in the commercial "free" world will be constrained, as it was in the Eastern European countries under Soviet- dominated police-state surveillance.

Conclusion

Predictive analytics is its own ideology: "Big Data is the intellectual engine of the modulated society. Its techniques are techniques for locating and extracting consumer surplus and for managing, allocating, and pricing risk, and it takes data sets at face value. But the values of predictive rationality and risk management are values, and they are the values with which serious critics of Big Data need to contend." (Cohen 2014: 1924). Its impact will expand choices for some people, constrain choices for others, and alter people's own behaviors and thus autonomy, all while seeming to be value free. To be sure, an excess of choice is not always better, as studies have shown. (I may want a glass of wine with my lunch, though 100 choices is overwhelming to a decision, yet I may be tempted by a favorite wine on the list.) A system that facilitates choices by anticipating our desires and offering our preferred options faster is useful. But a system that preempts our reason and rationality to act on those impulses and options (I should not have the wine today), subverts our self-governance, and interferes with our future development by ever reminding us of our past impulses. Commercial firms that use the same information and models at the same time both to tease one's impulses to spend money, and scold one's self-control to save or repay money, will grind away a person's autonomy, judgment and character.

References

Abelson, R., 2015, "Health Insurance Deductibles Rising Faster Than Wages, a Study Finds," New York Times (Sept. 23), B1.

Achtziger, A., Hubert, M., Kenning, P., Raab, G., Reisch, L., 2015. "Debt Out of Control: The Links Between Self-Control, Compulsive Buying and Real Debts," Journal of Economic Psychology, 49: 141-149.

Andriotis, A., 2014, "FICO Recalibrates Its Credit Scores," Wall Street Journal, (Aug. 7) http://www.wsj.com/articles/fico-recalibrates-its-credit-scores-1407443549.

Aristotle, The Nicomachean Ethics, Lesley Brown, (trans.) (2009). (Oxford: Oxford University Press).

Atiyeh, C., 2014, "Screen-Plate Club: How License-Plate Scanning Compromises Your Privacy," Car and Driver, (Oct.), http://www.caranddriver.com/features/screen-plate-club-how-license-plate-scanning-compromises-your-privacy-feature.

Audi, R., 2012, "Virtue Ethics as a Resource in Business," Business Ethics Quarterly vol. 22 (2) (April) pp. 273-291.

Baumeister, R., 2002. "Yielding To Temptation: Self-Control Failure, Impulsive Purchasing, And Consumer Behavior," Journal of Consumer Research, 28(4): 670–676.

[67] And see King and Raja, 2013, and Rosen, 2012, and Tsesis, 2014 for a discussion of the differences between the E.U. and U.S approaches to privacy and the right to be forgotten.

Benhabib, J. and Bisin, A., 2005. "Modeling Internal Commitment Mechanisms And Self-Control: A Neuroeconomics Approach To Consumption–Saving Decisions," Games and Economic Behavior, 52: 460-492.

Boulard, Garry, 20014, "The Roar Over Credit Scores," State Legislatures, 30(9): p. 19-21 (October).

Brockett, P and Golden, L., 2007, "Biological and Psychobehavioral Correlates of Credit Scores and Automobile Insurance Losses: Toward an Explication of Why Credit Scoring Works," Journal of Risk and Insurance, vol. 74 (1), p. 23.

Christen, M., Alfano, M., Bangerter, Ena dn Lapsley, D., (2013), "Ethical Issues of 'Morality Mining'," in Rahman, Hakikur; Ramos, Isabel. Ethical data mining applications for socio-economic development. Hershey, PA: Information Science Reference, http://dx.doi.org/10.5167/uzh-81466.

Christman, J., (1988), "Constructing the Inner Citadel: Recent Work on the Concept of Autonomy," Ethics, 99 (1): 109-124.

CFPB, 2014, "Data Point: Medical Debt and Credit Scores," Consumer Financial Protection Bureau (May).

CNIL, 2015. "Right to Delisting: Google Information Appeal ejected," CNIL (Sept. 21), http://www.cnil.fr/english/news-and-events/news/article/right-to-delisting-google-informal-appeal-rejected/.

Cohen, J., 2014, "What Privacy Is For," Harvard Law Review, 126: 1904-1933.

Cohen, R. and Kirzinger, W., 2014, "Financial Burden of Medical Care: A Family Perspective," NCHS Data Brief, No., 142, Centers for Disease Control (Jan.), available at http://www.cdc.gov/nchs/data/databriefs/db142.htm.

Commonwealth Fund, 2013, "New 11-Country Health Care Survey: U.S. Adults Spend Most; Forgo Care Due to Costs, Struggle to Pay Medical Bills, and Contend with Insurance Complexity at Highest Rates," Commonwealth Fund (Nov. 13), http://www.commonwealthfund.org/publications/press-releases/2013/nov/new-11-country-health-care-survey.

Davenport, T., 2013, "Who Owns Your Data Exhaust?," Wall Street Journal, (Nov., 20), http://blogs.wsj.com/cio/2013/11/20/who-owns-your-data-exhaust/.

Davidsons, R., Dey, A., and Smith, A., 2012, "Executives' 'Off-the-Job' Behavior, Corporate Culture and Financial Reporting Risk," NBER Working Paper 18001 (2012).

Ding, C., Chang, K. and Liu, N., 2009. "The Roles of personality and General Ethical Judgments in Intention to Not Repay Credit Card Expenses," Service Industries Journal, 29(6): 813-834.

Duhigg, C., 2009, "What Does Your Credit-Card Company Know About You?," New York Times Magazine, (May 12), p MM40.

Duhigg, C., 2012, "Psst, You in Aisle 5," New York Times (Feb. 16), p. MM30.

Ebert, S., & Wiesen, D., 2014, "Joint Measurement Of Risk Aversion, Prudence, And Temperance," Journal Of Risk & Uncertainty, 48 (3), 231-252.

EPIC, "License Plate Recognition Systems," Electronic Privacy Information Center, https://www.epic.org/privacylicenseplates.

European Commission, "Fact Sheet on the Right to be Forgotten Rule (C131/12), http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf

Fair Isaac, 2014, "FICO Score 9 Introduces Refined Analysis of Medical Collections," Fair Isaac Corp., (Aug. 7), http://www.fico.com/en/newsroom/fico-score-9-introduces-refined-analysis-of-medical-collections

Fallon, R., A. Jr., 1994, "Two Senses of Autonomy," Stanford Law Review, 46: 875-905.

Federal Reserve Bulletin, 2014, "Changes in U.S. Family Finance from 2010 to 2013: Evidence from the Survey of Consumer Finances," Federal Reserve Bulletin (September) 100(4).

Feldman, H., 2000, "Prudence, Benevolence, And Negligence: Virtue Ethics And Tort Law," Chicago-Kent Law Review, 74: 1431-466.

Fischer, P., Kastenmüller, A., & Asal, K. (2012), "Ego Depletion Increases Risk-Taking," Journal Of Social Psychology, 152 (5), 623-638.

Foot, P., 1978, Virtues and Vices in Darwall, Stephen (editor) (2003), Virtue Ethics (Malden, MA: Blackwell Publishers Ltd.).

Fried, C., 2015, Contract As Promise: A Theory of Contractual Obligation (Second edition.)

FTC, 2007, "Credit-Based Insurance Scores: Impacts on Consumers of Automobile Insurance: A Report to Congress By the Federal Trade Commission," Federal Trade Commission (July).

FTC, 2012, "Protecting Consumer Privacy In An Era of Rapid Change," Federal Trade Commission Report (March).

FTC, 2014a, "Data Brokers, A Call for Transparency and Accountability," Federal Trade Commission Report (May)

FTC, 2014b, "Two Data Brokers Settle FTC Charges That They Sold Consumer Data Without Complying With Protections Required Under the Fair Credit Reporting Act," Federal Trade Commission Press Release (April 9).

GAO Retirement Security, 2015. "Retirement Security. Most Households Approaching Retirement Have Low Savings, United States Government Accountability Office (May).

Gathergood, J., 2012. "Self-control, financial literacy and consumer over-indebtedness," Journal of Economic Psychology, 33 (3): 590-606.

Google Spain SL, Google Inc. v. Agencia Española de Protección de Datos (AEPD), Mario Costeja González, 2014, Case C-131/12, Court of Justice, http://curia.europa.eu/juris/document/document.jsf?text=&docid=152065&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=305802

Gordley, J., 1991, The Philosophical Origins of Modern Contract Doctrine, Oxford: Clarendon Press.

Gottlieb, P., 2009, The Virtue of Aristotle's Ethics, Cambridge: Cambridge University Press.

Halon, D and Boyd, L., 1996, "The Role of Credit History Scoring in Personal Lines Insurance Underwriting, CPCU Journal, vol. 49 (1), p. 40-44.

Halper, T., 1996, "Privacy and Autonomy: From Warren and Brandeis to Roe and Cruzan," Journal of Medicine and Philosophy, 21: 121-135.

Hardy, Q., 2015, "Determining Character with Algorithms," New York Times (July 27), p. B5.

Hartman, E. M., 1998, "The Role of Character in Business Ethics," Business Ethics Quarterly, vol. 8 (3): pp. 546-559.

Herold, B., 2014, "Google Under Fire for Data-Mining Student E-mail Messages," Education Week (March 13, updated June 23, 2015), http://www.edweek.org/ew/articles/2014/03/13/26google.h33.html.

Hyder, S., 2014, "7 Things you Can Do To Build An Awesome personal brand," Forbes, (Aug. 18), http://www.forbes.com/sites/shamahyder/2014/08/18/7-things-you-can-do-to-build-an-awesome-personal-brand/.

Kaplan, K, 2014, "Medical Bills Are A Burden For More Than 1 In 4 Families, CDC Says," Los Angeles Times, (Jan. 28).

Kerr, I., and Earle, J. (2013), "Prediction, Preemption, Presumption: How Big Data Threatens Big-Picture Privacy," Stanford Law Review Online, 66: 65-72.

King, N. T. and Raja, V.T., (2013), "What Do They Really Know About Me in the Cloud? A Comparative Law perspective on Protecting Privacy and Security of Sensitive Consumer Data," American Business Law Journal, 50: 413-380.

Konnikova, M., 2014, "The Struggles of a Psychologist Studying Self-Control," The New Yorker (Oct. 9), http://www.newyorker.com/science/maria-konnikova/struggles-psychologist-studying-self-control.

Kupfer, J., 1987, "Privacy, Autonomy, and Self-Concept," American Philosophical Quarterly, 24 (1): 81-89.

LaFrance, A., 2015. "Not Even the People Who Write Algorithms Really Know How They Work," The Atlantic, (Sept. 18) http://www.theatlantic.com/technology/archive/2015/09/not-even-the-people-who-write-algorithms-really-know-how-they-work/406099/.

Lam, B, 2015, "For More Workplace Diversity, Should Algorithms Make Hiring Decisions?," The Atlantic (June 22), http://www.theatlantic.com/business/archive/2015/06/algorithm-hiring-diversity-HR/396374/.

Lewis, M., 2015, "Harvard's 'Moneyball for Life," New York Times, (June 21), p. SR2.

Limerick, L., & Peltier, J. W. (2014). "The Effects Of Self-Control Failures On Risky Credit Card Usage," Marketing Management Journal, 24(2): 149-161.

Lynch, J., 2013, "Automated License Plate Readers Threaten Our Privacy," Electronic Frontier Foundation (May 6), https://www.eff.org/deeplinks/2013/05/alpr.

Madrigal, A., 2012. "I'm Being Followed: How Google—and 104 Other Companies— Are Tracking Me on the Web," The Atlantic (Feb. 29), http://www.theatlantic.com/technology/archive/2012/02/im-being-followed-how-google-151-and-104-other-companies-151-are-tracking-me-on-the-web/253758/.

Miller, C., 2015a, "Can an Algorithm Hire Better Than a Human?," New York Times, (June 28), p. SR4.

Miller, C., 2015b, "Teaching Algorithms Not to Discriminate," New York Times (Aug. 11), p. A3.

Nicolosi, G. and Yore, A., 2015, "'I Do': Does Marital Status Affect How Much CEOs Do?," Financial Review, 50: p. 57–88.

Peters, T., 1997, "The Brand Called You," Fast Company, (Aug. 31), http://www.fastcompany.com/28905/brand-called-you.

Raz, J., The Morality of Freedom (Oxford: Clarendon Press, 1986).

Richards, N. and King, J., 2013, "Three Paradoxes of Big Data," Stanford Law Review Online, 66: 41-46.

Richards, N. and King, J., 2014, "Bit Data Ethics," Wake Forest Law Review, 49: 393-492.

Riffkin, R., 2014, "Cost Still a Barrier Between Americans and Medical Care," Gallup, (Nov. 28), http://www.gallup.com/poll/179774/cost-barrier-americans-medical-care.aspx.

Ritholtz, B., 2010, Where's the Note? Leads BAC to Ding Credit Score, The Big Picture" (Dec. 14), http:// www.ritholtz.com/blog/2010/12/note-bac-credit-score/.

Rosenthal, Elisabeth, 2014, "When Health Costs Harm Your Credit," New York Times, (March 8).

Rosen, J., 2012, "The Right to Be Forgotten," Stanford Law Review Online, 64: 88-91 (Feb. 13).

Siegal, T., 2015, "TransUnion, Equifax and Experian Agree to Overhaul Credit Reporting Practices," New York Times, p. B3 (March 10).

Silver-Greenberg, J., 2012, "Perfect 10? Never Mind Tha. Ask Her for Her Credit Score," New York Times (Dec. 26), p. A1.

Singer, N., 2012, "You, For Sale: Mapping, and Sharing, the Consumer Genome," The New York Times, (June 16), BU1.

Taylor, A and Sadowski, J., 2015, "How companies Turn Your Facebook Activity Into a Credit Score," The Nation (May 27), http://www.thenation.com/article/how-companies-turn-your-facebook-activity-credit-score/.

Treiger-Bar-Am, K, 2008, "In Defense of Autonomy: An Ethic of Caring," New York University Journal of Law & Liberty, 3: 548-598.

Tsesis, A., 2014, "The Right to Erasure: Privacy, Data Brokers, and the Indefinite Retention of Data," Wake Forest Law Review, 49: 433-480.

Wall, E., 2011, "Privacy and the Moral Right to personal Autonomy," International Journal of Applied Philosophy, 25 (1): 69-85.

Wilson, J. Q. (1999). On Character, (Washington, D.C.: The AEI Press).

Wilson, J. Q. (1985). "The Rediscovery of Character: Private Virtue and Public Policy," The Public Interest, 81: 3-18.

Wyatt, E., 2012, "U.S. Penalizes Online Company in Sale of Personal Data," New York Times (June 13), p. B2.

Philipp Richter and Andreas Kaminski:

„Before you even know …" – Big Data und die Erkennbarkeit des Selbst

Abstract:

Der Big Data-Technologie wird das Potenzial zugeschrieben, durch Mustererkennung in aggregierten Daten Verhaltensweisen von Personen zu prognostizieren, noch bevor diese intendiert und reflektiert würden. Zumeist widmet sich die Big Data-Debatte daher den Befürchtungen möglicher Einbußen von Privatheit und Freiheit. In unserem Beitrag wählen wir jedoch einen anderen Zugriff und fragen, inwiefern die Big Data-Visionen das Selbstsein betreffen, also das Konzept davon, wer ich selbst eigentlich bin. Wenn Selbstsein, Martin Heidegger zufolge, eigentlich bedeutet, in kritischer Distanz zu vorgegebenen Möglichkeiten zu leben, stellt sich die Frage, welches Kriterium hier eine klare Unterscheidbarkeit gewährleistet: Wie kann ich wissen, echt ich selbst zu sein, ohne nur wiederum andere Üblichkeiten nachzuahmen? Wir fragen daher, ob die Big-Data-Technologie subjektive Verzerrungen und Täuschungen über das Selbst ausschalten könnte. Hierfür wird der Begriff „Selbstsein" bei Heidegger erarbeitet. Dabei lässt sich allerdings zeigen, dass dieser letztlich ohne Trennschärfe bleibt, da die nur negative Charakterisierung von eigentlichem Selbstsein, Heidegger zufolge, echt oder unecht ausgestaltet sein kann – der aus der Möglichkeit echter oder unechter Selbsterzählung resultierende Relativismus wurde bisher nicht hinreichend beachtet. Big Data kann zwar uneigentliche Erzählmuster vermeiden, nicht aber uneigentliches oder eigentliches Verhalten differenzieren; ob ich z.B. manches – unecht eigentlich – aus Trotz oder Kalkül gegen mein eigenes Selbstverständnis tat, bleibt verborgen. Fazit: Es muss entweder ein Kriterium für echtes Eigentlichsein entwickelt oder der Anspruch, darüber Auskunft zu geben, verworfen werden – bei Heidegger und Big Data. Verzichtet man jedoch auf die Idee einer (echten) Eigentlichkeit verzichtet man auf den Begriff des Selbst.

Agenda:

Authors:

Dr. Andreas Kaminski

- High Performance Computing Center Stuttgart (HLRS), Department of Philosophy of Science & Technology, Nobelstr. 19 D, D-70569 Stuttgart
- ☎ ++49 (0)711 685-64288, Fax: ++49 (0)711 685-87209, ✉ kaminski@hlrs.de

Dr. Philipp Richter

- TU Darmstadt, Institut für Philosophie, FG: Philosophie der wissenschaftlich-technischen Kultur, Marktplatz 15 / Schloss, Wallhaus, D-64283 Darmstadt
- ☎ ++49 (0)6151 16-24785, Fax: ++49 (0)6151 16-24785, ✉ richter@phil.tu-darmstadt.de

ISSN 1614-1687

Einer üblichen Meinung nach kann keiner besser als ich selbst wissen, was mich eigentümlich ausmacht und wer ich eigentlich bin. Vor allem durch das introspektive Wissen davon, was ich eigentlich will, scheine ich einen privilegierten Status in der Frage zu haben, wer ich eigentlich bin. Doch kann von einer solchen Autorität[68] der 1. Person gesprochen werden? So dass der Person ein Vorrang (des Anspruchs, des Zugangs, der Relevanz) im Wissen über sich selbst zukommt? In der Moderne ist ein solches Privileg auf vielfache Weise bestritten worden. An die Stelle einer Autorität des Ich sind zahlreiche Kulturtechniken objektiver Selbsterkenntnis getreten: von der Psychoanalyse über Psychometrie[69] zum Coaching, von den kybernetischen Feedbackschleifen über Apps zur Beobachtung des eigenen Verhaltens sind Formen entwickelt worden, welche die Möglichkeiten unmittelbarer Selbsterkenntnis als begrenzt betrachten und daher auf einen technisch vermittelten Zugang zum eigenen Selbst setzen. Anlass hierfür könnten Erfahrungen der Intransparenz des eigenen Selbst sein, wie sie exemplarisch Peter Bieri veranschaulicht:

> *„Gibt es einen Willen, den ich haben kann, ohne davon zu wissen? Oder ist der Spielraum meines Wollens deckungsgleich mit dem Spielraum des mir bekannten Wollens? Wenn ich von jedem Willen, der mir möglich ist, automatisch auch wüßte, so läge darin eine vollkommene Nähe zu mir selbst. Ich wäre über mich selbst stets auf dem laufenden und mir ganz und gar durchsichtig. Doch in Wirklichkeit ist die Intimität zwischen mir als Wissendem und mir als Wollendem längst nicht so groß. Ich kann, was meinen Willen betrifft, Überraschungen erleben. […] Das sind Erfahrungen, die eine Ahnung davon vermitteln, daß man über seine Wünsche und ihre mögliche Handlungswirksamkeit längst nicht alles weiß. Und natürlich gibt es auch längerfristige Beobachtungen, die mich darüber belehren können, daß es außer dem vertrauten noch einen verborgenen Willen in mir gibt. Wenn ich auf mein Tun zurückblicke, so mag ich darin ein Muster entdecken, das mich verblüfft sagen läßt: Es sieht ganz so aus, als hätte ich die ganze Zeit etwas anderes gewollt, als ich dachte."*[70]

Was Bieri hier über das eigene Wollen ausführt, das erst rückblickend einem nicht intendierten Muster zu folgen scheint, bildet den Ausgangspunkt für eine Technologie, der auch das Potenzial zugeschrieben wird, das Selbst einer Person besser als diese selbst zu erkennen: Big Data. Ein wesentlicher Zug von Big Data-Technologien ist, in großen Datenmengen Muster zu erkennen. Dabei ist es zunächst gleichgültig, woher die Daten stammen und worüber sie handeln. Entscheidend ist vielmehr, dass durch Algorithmen Strukturen in den Daten erkennbar werden, die zur Prognose von Krankheiten, kriminellem Verhalten, Konsumpräferenzen und vielen weiteren Verhaltensweisen verwendet werden könnten. Big Data-Technologien können eingesetzt werden, um exakt die Muster, von denen Bieri spricht, freizulegen.

Bekannt geworden ist der Erfolg des US-Discounters Target mit seinem „pregnancy prediction model" eine Schwangerschaft und sogar den Geburtsmonat vorherzusagen. Untern anderem wurde der Kauf größerer Mengen von 25 verschiedenen Produkten (wie unparfümierte Seifen oder Nahrungsergänzungsmittel wie Magnesium, Calcium und Zink) als ein Muster erkannt, das mit Schwangerschaften korreliert. Target interessiert sich für Schwangerschaften, um daraufhin mit personalisierten Angeboten die werdenden Eltern stärker an sich zu binden. Als ein Vater sich beschwerte, dass seine minderjährige Tochter Werbung für werdende Eltern erhielt, war Target zunächst in Erklärungsnot. Es stellte sich jedoch heraus, dass die Tochter tatsächlich schwanger war, nur der Vater davon nichts wusste. Dass unter anderem ihr Konsumverhalten eine Schwangerschaft indizierte, war der Tochter nicht bekannt, obgleich ihr klar war, schwanger zu sein. Andrew Pole, ein Statistiker und Ökonom, der bei Target an Vorhersagemodellen für Kunden arbeitet, führt Big Data genau im Sinne Bieris als Möglichkeit an, das Unbekannte des eigenen Willens zu erfassen: „Just wait. We'll be sending you coupons for things you want before you even know you want them."[71]

---

[68] Vgl. dazu Moran, Richard: Authority and Estrangement. An Essay on Self-Knowledge.

[69] Vgl. dazu die Geschichte der Psychometrie, etwa: Gelhard, Andreas: Entgrenzung der Psychotechnik: Der neue Geist des Kapitalismus und das Problem der Prüfungstechniken; oder: Kaminski, Andreas: Maßstäbe der Macht. Psychometrische Prüfungen als Techniken.

[70] Bieri, Peter: Das Handwerk der Freiheit, S. 39f.

[71] Vgl. den Artikel in der New York Times: Duhigg, Charles, How Companies Learn Your Secrets.

Vor dem Hintergrund dieser technologischen Entwicklung entsteht die Frage, ob Big Data das Selbst tatsächlich erfassen kann. Und wenn ja, in welchem Sinne? Viele der um Big Data geführten Debatten werden primär unter den Gesichtspunkten eines möglichen Verlustes von Freiheit und Privatheit geführt.[72] In Artikeln zu Big Data werden aber wiederholt Verhaltens- und Persönlichkeitsmuster thematisiert. Dies legt den Gedanken nahe, dass es sich auch um eine Technik handelt, die unser Selbst in spezifischer Weise erfassen könnte. Allerdings muss hier zunächst geprüft werden, was mit „Selbst" überhaupt gemeint sein kann. In einem 1. Schritt werden wir daher zunächst grundlegende Prinzipien von Big Data darstellen. In einem 2. Schritt wollen wir Martin Heideggers Überlegungen zum Selbstsein als eine philosophische Theorie einführen, um die Rede vom Selbst zu klären und zu differenzieren. Dabei, so die Hoffnung, soll nicht nur der Anspruch von Big Data an einem ausgearbeiteten philosophischen Begriff vom Selbst geprüft werden, sondern umgekehrt soll die Big Data-Vision auch als Impuls zu einer Klärung und kritischen Auseinandersetzung mit diesem Begriff beitragen. Dies wird Aufgabe des dritten Abschnitts sein.

Die Idee der Big Data Analytics

In der Moderne sind verschiedene Kulturtechniken der Selbsterkenntnis und -formung entstanden. Einige orientieren sich stark an Realtechniken wie etwa psychometrische Tests, andere setzen vielmehr auf Formen spezifischer Intellektualtechniken wie die Psychoanalyse. Worin besteht aber die Besonderheit von Big Data im Vergleich zu anderen Ansätzen? Big Data gilt primär als eine mathematisch-informationstechnische Methode, um Muster in großen Datenmengen zu entdecken. Diese Beschreibung ist sicherlich zutreffend, nur verdeckt sie das eigentlich relevante. Denn die Eigenheit von Big Data tritt erst dann hervor, wenn man die Idee der Reduktion auf eine solche Methode erfasst: Durch einen Verbindung von mathematisch-statistischer und informationstechnischer Methodik sollen fachlich-theoretische Annahmen über den Gegenstandsbereich minimiert werden. Im Kontext unserer Problematik wird Big Data häufig als neue Verhaltenswissenschaft (‚behavioral research') beschrieben. Dabei ist der entscheidende Punkt, dass Mathematiker und Informatiker die führende Rolle in dieser Verhaltenswissenschaft spielen. Man die Idee daher so formulieren, dass es sich um eine Verhaltenswissenschaft ohne Verhaltenswissenschaftler handelt.[73] Warum diese Forschungsmethodik einen gegenüber klassischer soziologischer und psychologischer Forschung veränderten Anspruch aufweist, ist vermutlich durch drei Annahmen motiviert, die im Bild von Big Data konvergieren:

(1) Ubiquitäre und informelle Datenerzeugung: Big Data schaffe Möglichkeiten der Beobachtung von Personen, die neuartig sind; die zugleich informelle als auch ubiquitäre Datenerzeugung schließe nämlich Täuschungen, die beim Beobachten von Personen durch Personen ein methodisches Problem darstellen, weitgehend aus, und sie biete eine Beobachtungstiefe und -genauigkeit, welche unerreicht sei.

(2) Theorielose Mustererkennung: Dabei soll Big Data-Analysen auf Theorien verzichten können.[74] Gefordert seien allein mathematische Methoden der Musterkennung. Die Muster fänden sich in den Daten, der (lernende) Algorithmus legt diese lediglich frei. Auf (subjektive) Interpretationen, auf Modelle, auf ein qualitatives bzw. intersubjektives Verstehen des zugrundeliegenden Zusammenhanges, wie in den Sozialwissenschaften notwendig, könne daher verzichtet werden.

---

[72] Vgl. z.B. Roßnagel, Alexander, Big Data - Small Privacy, S. 562, 566; Zwitter, Andrej, Big Data ethics, S. 3f. Vgl. dazu auch die auf der Jahrestagung des deutschen Ethikrates gehaltenen Vorträge: Marquardt, Wolfgang, Was ist Big Data? Versuch einer Positionsbestimmung; Dabrock, Peter, Freiheit, Selbstbestimmung, Selbstkonstruktion – zur Ethik von Big Data.

[73] Wie gesagt der Idee nach: In vielen Projekten, aber keineswegs in allen, wirken Psychologen und Neurowissenschaftler mit. Gleichwohl ist die führende Rolle an Mathematiker und Informatiker übergegangen.

[74] Vgl. Chris Anderson, der in seinem Essay "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete" das Ende der Theorie gekommen sieht. Dort schreibt er: "Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves."

(3) Type und Token: Die Muster, welche (theorielos) in den ubiquitär und informell gewonnenen Daten erkannt werden, sollen Auskunft über die Person geben, in dem diese in einen Sozialzusammenhang[75] eingeordnet, als Token eines Typs bestimmt werden.[76] Der Zusammenhang zwischen einem Token und seinem Typ wird hier ausschließlich als probabilistische Ähnlichkeit der zugrundeliegenden Muster bestimmt.

Soweit das Bild von Big Data – es ist jedoch durchaus fraglich, ob reale Umsetzungen dieses erfüllen. So spielen Vorüberlegungen von Verhaltenswissenschaftlern zum Teil durchaus eine wichtige Rolle. Ferner dürfte es von Person zu Person und Situation zu Situation sehr unterschiedlich ausfallen, in welchem Maße die informelle Datenerzeugung nicht doch durch bewusstes Verbergen und verändertes Verhalten verzerrt wird. Gleichwohl ist es ungemein schwer, keine Daten zu erzeugen oder Muster zu verbergen; auch weil Algorithmen inzwischen daraufhin angepasst sind. Zunehmend bedeutet der Datenverzicht zugleich den Verzicht auf Teilnahme am Alltag. Schließlich: Wenn Muster den Personen selbst nicht bekannt sind, ist es fraglich, inwiefern sie verborgen werden können.[77]

Eine Kritik der Umsetzung der methodischen Ansprüche ist sicherlich von großer Wichtigkeit. Im Rahmen unseres Aufsatzes wählen wir aber einen anderen Weg. Wir vergleichen nicht methodischen Anspruch und Umsetzung, Idee und Realität von Big Data. Wir nehmen eine Einklammerung vor und gehen davon aus, die Idee würde erfüllt: Kann selbst unter dieser Annahme davon ausgegangen werden, dass dann das Ziel erreichbar ist, nämlich das Selbstsein einer Person zu enthüllen? Also über deren Selbst und Eigenheiten präziser, verlässlicher und tiefgreifender Auskunft zu geben als es ihr selbst möglich wäre? Im Folgenden prüfen wir dies, indem wir die Idee von Big Data und den Begriff des Selbst bei Heidegger miteinander ins Verhältnis setzen. Dies soll aber nicht einseitig geschehen, so dass Big Data nur an einem vorgegebenen Begriff des Selbst gemessen wird; vielmehr soll die die Idee dieser Technologie auch dazu genutzt werden, den Begriff des Selbst zu klären und philosophische Forschungsdesiderate aufzuweisen.

Der Begriff „Selbstsein“ bei Heidegger

Die beiden zentralen mit dem Begriff des Selbstsein verbundenen Fragen lauten: „Wer bin ich eigentlich?“ Und „wie kann ich eigentlich ‚ich selbst‘ sein?“[78] Auf den ersten Blick scheint es, als bin ich in besonderem Maße „ich selbst“, wenn ich mich kritisch vom alltäglichen Verhalten distanziere und sichtbar unangepasst lebe: Wer also trotz Alltag in wachen Momenten in der Lage ist, seine „Eigenheit“[79] zu retten und gegen Gruppendruck und Üblichkeiten, das zu tun, was er eigentlich will, scheint dem eigenen Selbstsein besonders nahe zu sein. Mit dieser Auffassung sind allerdings theoretische Probleme verbunden, die wir zunächst an einem Beispiel verdeutlichen und dann mit Heideggers Ausführungen in *Sein und Zeit* (1927) vertiefen wollen.

Nehmen wir an, jemand arbeite im Finanzamt einer Kleinstadt; unauffällig und sachlich korrekt versieht er seit vielen Jahren den Innendienst. Von außen betrachtet scheint es sich um ein sehr angepasstes Leben ohne besondere Eigenheiten und Überraschungen zu handeln. Wenn dieser Beamte plötzlich ‚zu sich kommt‘ und feststellt, dass er bisher sein „Leben nur mitmacht, statt es selbständig zu führen“[80], muss er, um die Frage zu beantworten, wer er eigentlich ist und sein will, zunächst eine Geschichte von sich selbst erzählen. Wie wurde

[75] Dies klingt nach einer Wiederkehr der von Anderson verabschiedeten Taxonomie – und also nach einem Widerspruch? Für unser Argument müssen wir indes nicht zeigen, dass sich *dieses Bildes* von Big Data und Selbst bruchlos in allen zentralen Positionen aufweisen lässt – das Bild leitet unseres Erachtens die verschiedenen Positionen an und orientiert sie. Daher brächte eine Spannung zwischen Bild und Position unser Argument nicht in eine Bredouille. Allein, es liegt hier auch keine Spannung oder ein Widerspruch vor. Type und Token sollen nämlich nicht durch ein Verstehen und also theoretisch gewonnen, sondern allein mathematisch als Muster in den Daten aufgewiesen werden.

[76] Vgl. zur Rolle anonymer Sozialtaxonomien auch Wiegerling, Klaus et al.: Ubiquitärer Computer – Singulärer Mensch, S. 71-84.

[77] Vgl. den Fall der entdeckten Schwangerschaft durch die Modelle von Target.

[78] Luckner, Andreas: Klugheit, S. 20.

[79] Luckner, Andreas: Klugheit, S. 18.

[80] Luckner, Andreas: Klugheit, S. 21.

er zu dem, der er ist und sich nun in einer existenziellen Krise befindet? Ausgehend von der Erzählung seines bisherigen Selbstseins kann dann eine Neuorientierung, was er „wirklich und eigentlich will"[81], stattfinden. Doch was ist das Wesentliche in dieser Selbsterzählung und welcher Schluss ist daraus für das Selbstsein zu ziehen? Der Beamte könnte sich sagen, um ich selbst zu sein, sollte ich sofort kündigen, denn bisher habe ich mich immer nur untergeordnet und mir selbst selten Raum gegeben. Es wäre jedoch ebenfalls denkbar, dass die Geschichte anders erzählt wird: Bisher habe ich, könnte der Beamte sagen, eigentlich nie etwas durchgehalten, außer meiner Arbeit. Z.B. habe ich alle Versuche, ein Hobby oder eine erfüllende Freizeitbeschäftigung zu finden immer abgebrochen. Ich sollte also nicht kündigen, sondern durchhalten: Vielleicht kann ich durch das Bewusstmachen meines Durchhaltevermögens auch etwas für die Gestaltung anderer Lebensbereiche gewinnen? So wird deutlich, dass sich je ein anderes Selbstsein ergibt, je nachdem, wie die bisherige Geschichte erzählt wird: Ist unser Finanzbeamter (eigentlich) einer, der kaum etwas durchgehalten hat, oder einer, der sich immer untergeordnet hat? Welche Geschichte ist die richtige? Und wie lässt sich also entscheiden, welche Neuorientierung dem eigentlichen Selbst des Beamten entspricht und welche nicht?

Die Frage eines *eigentlichen* Selbstsein diskutiert Martin Heidegger in seinem Hauptwerk „Sein und Zeit".[82] Die Problemstellung ist, wie Selbstsein gedacht werden kann, wenn die menschliche Existenz als etwas individuell Erlebtes nicht auf eine allgemeingültige Wesensbestimmung reduzierbar ist. Wesensbestimmungen erklären, was etwas oder jemand – in einer bestimmten Hinsicht – eigentlich ist und somit auch, was es bzw. jemand eigentlich sein sollte. Nimmt man z.B. an, dass der Mensch wesentlich ein vernünftiges Lebewesen ist, dann sind Phantasien oder Gefühlsregungen nur beiläufige und das eigentliche Selbstsein störende Einflüsse.[83] Im Rahmen *dieser* Wesensbestimmung gilt, dass das Menschsein desto besser realisiert wird, desto weniger lebensbestimmend Gefühle sind. Eine Wesensbestimmung der menschlichen Existenz könnte jedoch auch von der Empfindsamkeit her gedacht werden, sodass ganz andere Verhaltensweisen als wesentlich erscheinen würden. Weitere Wesensdefinitionen sind denkbar, nur treffen sie, so Heidegger, nie vollständig das, was es ausmacht, als ein Selbst zu existieren. Basaler als jede Wesensbestimmung sei das Erlebnis der eigenen Existenz, das Heidegger terminologisch als „Dasein" fasst. Das Dasein „ist ein Wer, kein Was" (§9, 45), d.h., dieses Seiende weist eine ursprüngliche Individualität auf. Eine erste formale Charakterisierung dieses Dasein ist, das es sich „in seinem Sein verstehend zu diesem Sein verhält" (§12, 52f.). Selbstsein wird von Heidegger als ein praktisches „Selbstverhältnis" konzipiert:[84] „In seinem Sein [geht es dem Dasein] um dieses Sein selbst" (§4, 12). Dasein verhält sich also immer zu dem, was es war, ist und sein kann. Heidegger erklärt diesen Sachverhalt im Vokabular von Möglichkeiten. Dasein sei „je seine Möglichkeiten selbst" (§39, 181). D.h., es ist ein Verhalten zu Möglichkeiten zu sein, womit jedoch auch gewählte, negierte, ignorierte, erhoffte, zukünftige etc. Möglichkeiten des Selbstseins gemeint sind.

Ist das aber nicht auch nur wiederum eine – freilich sehr formelle – weitere Wesensdefinition? Erklärt Heidegger hier nicht erneut, was die menschliche Existenz im Allgemeinen ausmacht, und widerspricht damit der eigenen Behauptung einer ursprünglichen Wesenslosigkeit bzw. Individualität des Selbst? Bei genauerer Betrachtung wird deutlich, dass hier keine Inkonsistenz vorliegt, da Dasein als prinzipiell frei von einer *absoluten* wesentlichen Bestimmung gedacht wird. Allerdings sind die potenziell unendlichen Möglichkeiten, wer ich selbst sein will, durch meinen Umgang mit Seinsmöglichkeiten begrenzt, d.h., gewissermaßen durch ‚meine' eigenen bisherigen Wesensdefinitionen (vgl. §6, §12). So ist z.B. die Aussage „Der Mensch ist ein vernünftiges Lebewesen" eine der Möglichkeiten, was als das Wesentliche meiner Existenz bezeichnet werden kann. Diese Festlegung kann jedoch nicht absolut und endgültig sein, da sich jedes Dasein hierzu wiederum affirmierend, indifferent oder ablehnend o.ä. verhalten kann. Das Möglichsein des Daseins wird also durch die Aussage nicht vollständig erschlossen. Heidegger geht es hier um eine gleichsam ‚transzendentale Figur' und deshalb ist die Definition von Dasein keine weitere Wesensbestimmung neben anderen. Diese transzendentale Figur lässt sich

[81] Luckner, Andreas: Klugheit, S. 22.

[82] Im Folgenden zitieren wir „Sein und Zeit" durch Angabe des Paragraphen und der Seitenzahl der 18. Auflage.

[83] Vgl. Sein und Zeit, §10, in dem Heidegger die Daseinsanalytik von verwandten Untersuchungen des menschlichen Seins (Anthropologie, Psychologie, Biologie) abgrenzt. Die drei genannten Einzelwissenschaften vom Menschen erforschen diesen stets selektiv als Seiendes mit einem bestimmten Wesen (z.B. der Mensch ist ein besonderes Lebewesen bzw. Organismus).

[84] Figal, Günter: Phänomenologie der Freiheit, S. 98, 404f. Vgl. Herrmann, Friedrich-Wilhelm von: Subjekt und Dasein, 30f.

analog zum Cogito-Argument bei Descartes erläutern: Was gedacht wird, kann, so Descartes, wahr oder falsch sein, somit ist jedoch das Gedachtwerden des Gedankens als eine Bedingung für sein Wahr- oder Falschsein unbezweifelbar. Analog dazu kann das Dasein sein Wesen in dieser oder jener Weise inhaltlich bestimmen (wie in der Analogie Gedanken wahr oder falsch sein mögen), woran als eine Bedingung ersichtlich wird, dass das Dasein eine solche Bestimmung aus einer Vielzahl von Möglichkeiten vornimmt und sich dabei notwendig zu dieser Vielzahl verhält (wie das „Ich denke" alle wahren *und* falschen Gedanken begleiten können muss). Manche der Möglichkeiten erscheinen für das Dasein passend und werden affirmiert, andere negiert, obwohl sie auch denkbar wären etc. Diesen Sachverhalt, dass Dasein sich notwendig zu den in Aussagen geäußerten Auffassungen des Selbstseins wiederum verhält, bezeichnet Heidegger als „Jemeinigkeit" (§9, 42f.): Die Möglichkeiten des Daseins (sich selbst zu verstehen und praktisch tätig zu werden) sind also im Sinne der transzendentalen Figur notwendig immer auch je seine. Wie im Falle des Arguments bei Descartes iteriert diese gedankliche Struktur: Die vermeintlich absolute Wesensbestimmung, „Ich bin wesentlich Dasein und sonst nichts" wird als eine der vielen Auslegungen meiner Existenz durch mein Verhalten zu ihr begleitet und erscheint somit nicht mehr absolut (weil es z.B. „meines Erachtens zutrifft, was Heidegger schreibt und ich mich daher als derartiges Dasein verstehe" oder ähnliche ‚jemeinige' Zusatzformulierungen sind denkbar). Heidegger charakterisiert das Dasein daher in zweifacher Hinsicht als ein Möglichsein. „Dasein ist je das, was es sein kann und wie es seine Möglichkeit ist" (§31, 143). Zur Vereinfachung wollen wir von Möglichkeiten$_1$ (vorliegende Möglichkeiten, „was es sein kann") und Möglichkeiten$_2$ (Verhalten zu den Möglichkeiten$_1$, „wie es seine Möglichkeiten ist") sprechen. An Heideggers Formulierung wird deutlich, dass es sich nicht um einen Klassenunterschied verschiedener Arten von Möglichkeiten handelt (z.B. bloß abstrakte Wünsche vs. echte Handlungsmöglichkeiten), sondern um Verhalten zu Aspekten dieser Möglichkeiten.

Was Selbstsein bei Heidegger ausmacht, kann nun ausgehend von Husserls Begriff der Intentionalität verdeutlicht werden. Erstens ist das Selbstsein unausweichlich ein Verhalten zu ..., es ist ein Wahrnehmen, Fühlen, Denken von etwas etc.; zweitens – und hier überschreitet Heidegger Husserls Überlegungen – handelt es sich stets um ein praktisches Verhalten zu ..., das nicht notwendig reflexiv-kognitiv sein muss. D.h., dass das Selbstsein nicht nur in ‚wachen Momenten' stattfindet, sondern auch implizit in jeglichen nichtthematischen Verhaltensweisen geschieht. Damit dieses vorbewusste oder bewusste Selbstsein praktisch stattfinden kann, muss es Entscheidungsoptionen – Möglichkeiten$_1$ – geben. Schließlich ist Dasein drittens – und das ist mit Blick auf das Beispiel des Finanzbeamten der entscheidende Punkt – stets als ein Verhalten zu Entscheidungsoptionen (Möglichkeit$_1$) zu denken. Der Finanzbeamte findet sich in der Entscheidungssituation vor, zwischen Kündigung und Nicht-Kündigung zu wählen (Möglichkeit$_1$). Wie er sich entscheidet, hängt in dieser existenziellen Situation davon ab, wie er sich versteht und wer er sein will. Hier eröffnen sich Möglichkeiten$_2$, die sein Selbstsein betreffen: Als ein Verhalten zu den Möglichkeiten$_1$ kann er tun, was man in diesen Situationen so tut (z.B. Verdrängen, zum Psychiater gehen etc.), oder er versucht, ‚ganz er selbst zu sein' o.ä. Letzteres hängt jedoch entscheidend davon ab, welche Geschichte er bei der Vergewisserung über sich selbst vorfindet bzw. rekonstruieren kann. Die Erzählung seiner Selbst ist nicht willkürlich, denn es hat sich „immer schon irgendwie entschieden, in welcher Weise Dasein je meines ist" (§9, 42). Die Möglichkeiten seines Seins hat „das Dasein entweder selbst gewählt oder es ist in sie hineingeraten oder je schon darin aufgewachsen" (§4, 22). Zusammengefasst meint Dasein also – dem „formale[n] Begriff" nach (§12, 53) – Möglichsein in zweifacher Hinsicht, ohne Verstellung durch *absolute* Wesensdefinitionen. Jedoch ist Dasein material betrachtet jemeinig, d.h., es findet sich in einer mehr oder weniger aktiv gestalteten Selbsterzählung mit zahlreichen, *nicht-absoluten* Wesensbestimmungen vor.

Mit dem Finanzbeamten haben wir hier ein Beispiel gewählt, in dem sich eine Person ausgehend von einer Krise explizit über das, was sie ist, vergewissern will. Dieser explizite Zweifel und die Selbstvergewisserung sind jedoch nicht notwendig der *eigentliche* Fall von Selbstsein. Es stellte sich ja die Frage, welche (materiale) Selbsterzählung des Beamten nun die richtige sei und welche weitere Verhaltensweise seinem eigentlichen Selbstsein stärker entspräche. Hier sind Heideggers weitere Ausführungen zu einem uneigentlichen oder eigentlichen Selbstsein relevant: Ob eigentliches oder uneigentliches Selbstsein vorliegt, lässt sich an der Wahl bestimmter Verhaltensweisen oder Lebenswege nicht ablesen. Die Möglichkeiten weisen keine *intrinsischen* Eigenschaften auf, die sie zu eigentlichen oder uneigentlichen machen. Dafür steht die aspektuelle Unterscheidung von Möglichkeit$_1$ und Möglichkeit$_2$; die verschiedenen Optionen (M_1) sind nicht dadurch eigentlich oder uneigentlich, weil sie bestimmte intrinsische Eigenschaften dazu machten, sondern durch mein Verhalten zu ihnen; und es gibt zwei Möglichkeiten (M_2) sich zu ihnen zu verhalten. Daher ist auch nicht-

konformes Verhalten nicht notwendig deckungsgleich mit dem eigentlichen Selbstsein.[85] Denn es bliebe die Frage, welche per Konvention vorgegebenen Möglichkeiten auf welche Weise negiert und welche dagegen als die eigentlich richtigen beibehalten werden sollten? Wie am Beispiel des Beamten deutlich wurde, könnte sowohl die radikale Lebensänderung als auch ein Durchhalten in bisherigen Rollenmustern als eigentliches Selbstsein beschrieben werden – jeweils geht es um Möglichkeiten$_1$, die sowohl eigentlich als auch uneigentlich gelebt werden könnten.

Heidegger würde zudem hinter seine Kritik an verfehlt absoluten Wesensbestimmungen zurückfallen, würde er ein eigentliches Selbstsein in einem inhaltlichen Sinne als die einzig richtige Lebensweise empfehlen – die Jemeinigkeit der Existenz und die Möglichkeiten$_2$ würden negiert. Dass Heidegger diese Fehler nicht macht, wird deutlich, wenn er erklärt, die beiden Weisen des Selbstseins könnten „wiederum echt oder unecht sein"[86]. So wird deutlich, dass die explizite Reflexion auf das bisherige Leben, eine kritische Distanz zum alltäglichen Sein und das aktiv gewählte Ganz-Anders-Sein o.ä. nicht hinreichend sind, um von einem eigentlichen Selbstsein zu sprechen. Denn die explizit reflektierte und erzählte Geschichte des eigenen Selbst, kann unecht sein in dem Sinne, dass sie ausgehend von vorgegebenen Möglichkeiten des alltäglichen Man erzählt wird (§27), sich also wiederum vorgegebenen Strukturen anpasst: Sie kann z.B. in den typischen Erzählmustern des Aufsteigers oder eines Pechvogels erzählt werden („Manche – so auch ich – haben einfach immer Pech!"). Somit ergibt sich eine Kreuzklassifikation, die in der Literatur kaum Beachtung gefunden hat:[87]

Selbstsein	echt	unecht
Eigentlichkeit	**1.**	**2.**
Uneigentlichkeit	**4.**	**3.**

Zunächst scheint es in der Tat, als ließen sich die Gegenbegriffspaare Freiheit bzw. Authentizität und Fremdbestimmung bzw. Angepasstheit auf die Unterscheidung des eigentlichen und uneigentlichen Selbstseins abbilden. Auch Heideggers Wortwahl legt dies nahe, denn das uneigentliche Selbstsein sei an die es „umgebende Welt verfallen" und es existiere in „durchschnittlicher Alltäglichkeit" etc. Dagegen tritt die Rede vom eigentlichen Selbstsein stets in aktivischer Konnotation auf (wählen, entscheiden, die Möglichkeiten ergreifen, entschlossen sein etc.). Wie aber verhalte ich mich, wenn ich (1.) im echten Sinne eigentlich ich selbst bin, wenn sich doch kein wesentliches Ideal des Existierens als Grundlage der Beurteilung angeben lässt und sich unterschiedliche Geschichten des Selbstseins erzählen lassen? Das explizite Verhalten zum „Möglichen als solchem" (Möglichkeit$_2$), wie Heidegger Eigentlichkeit charakterisiert, kann ja wiederum echt oder unecht sein. Was aber heißt dann *echt* Eigentlichsein? Wann habe ich meine eigenen Möglichkeiten echt „eigens ergriffen" (§27, 129)? In Abgrenzung zu den drei anderen Seinsweisen kann eine zunächst indirekte Erläuterung erfolgen.

Die unechte Eigentlichkeit (2.) wird ihrem Begriff nicht gerecht, denn sie entspricht nicht der Jemeinigkeit, sondern postuliert ein absolutes Wesen des Existierens. Es liegt das sog. „existenzialistische Missverständnis" des Selbstseins vor.[88] Aus Naivität oder strategischem Kalkül wird eine Lebensform (eine bestimmte Möglichkeit$_1$) als die ausnahmslos richtige postuliert: Wer selbst sein wolle, müsse sich entsprechend dieses Ideals verhalten und sich somit von dem, „was man im Allgemeinen so tut" und „was die Anderen für richtig

[85] Andreas Luckner hat dargelegt, dass das eigentliche oder uneigentliche Selbstsein keine exklusiven Lebensformen sind, die wie vorliegende Optionen gewählt und gelebt werden könnten, vgl. Luckner, Andreas: Wie ist es, selbst zu sein, S. 155.

[86] Heidegger, Martin: Sein und Zeit, § 31, S. 146.

[87] So wird diese Stelle beispielsweise im Kommentar von Andreas Luckner gar nicht eigens behandelt, vgl. die Kommentierung des § 31 in: Luckner, Andreas: Martin Heidegger „Sein und Zeit", S. 66-69.

[88] Luckner, Andreas: Wie ist es, selbst zu sein, S. 155.

halten" abgrenzen. Die vermeintlich unangepasste Lebensform stellt sich jedoch bei genauerer Betrachtung, aufgrund der nicht hintergehbaren Institutionalität jeglichen Verhaltens, wiederum als allgemeiner Typus mit spezifischen Rollenerwartungen und vorgegebenen Möglichkeiten heraus.[89] Eigentlichsein kann nicht im Wählen einer bestimmten Lebensweise (z.B. mit ‚existenzialistischem Pathos' die Verachtung für jegliche Sinnsuche zur Schau stellen und sich als Nihilist inszenieren) oder in einem ständigen ‚aus der Rolle fallen' (also ein direkt nicht-angepasstes Verhalten zeigen) bestehen. Denn zu diesen Seinsweisen ist ja wiederum ein Verhalten im Modus der Möglichkeiten$_2$ denkbar, dann erscheinen sie als vorgegebene und nicht „ganz eigene" Möglichkeiten.

Dagegen könnte mit einer unechten Uneigentlichkeit (3.) eine strategisch-manipulative Maskerade gemeint sein: Um bestimmte Ziele zu erreichen, werden vorgegebene Möglichkeiten$_1$, z.B. die Üblichkeiten einer Profession, einer Branche oder eine bestimmte Rolle, besonders stark im Sinne der Erwartungen anderer gelebt. Die Instrumentalisierung zeigt allerdings, dass die Rolle nicht als alternativlos begriffen wird, hier ließe sich weiter fragen, ob nicht im Modus der Möglichkeiten$_2$ ein Selbstsein im Sinne von 1, 2 oder 4 vorliegt.

Dem echten Uneigentlichsein (4.) fehlt mit Blick auf die eigene Existenz etwas, dass das echte Eigentlichsein aufweist. Im Gegensatz zum echten Eigentlichsein verhält sich das echte Uneigentlichsein nicht zum „Möglichen als solchem", sondern ‚verbleibt' bei den Möglichkeiten$_1$. Echtes Uneigentlichsein wäre allerdings nicht die unreflektierte, totale Identifikation mit Rollenmustern und Institutionen (es wäre verfehlt, hier das Gegenstück zur unechten Eigentlichkeit zu sehen), sondern ließe, wie Heidegger schreibt (§41, 194f.), ein Wünschen alternativer Rollen etc. durchaus zu, z.B. „ich wünschte, ich hätte einen anderen Beruf ..." oder „Ach, würde es sich doch nur so fügen, dass ...". D.h., auch eine Person, die womöglich vom Standpunkt einer unechten Eigentlichkeit als bieder oder angepasst denunziert wird, muss nicht notwendig an sich uneigentlich sein. An Entscheidungen und Verhaltensweisen (Möglichkeiten$_1$) lassen sich eigentliches und uneigentliches Sein nicht ablesen, da es hier um die Möglichkeiten (M_2) sich zu Optionen (M_1) zu verhalten geht; aber dieses Verhalten zu Möglichkeiten kann eben nicht einfach *abgelesen* werden. Dies stellt in nuce die Problemsituation zu entscheiden, ob jemand eigentlich oder uneigentlich ist.

Die Charakterisierung einer echten Eigentlichkeit (1) muss – aufgrund des Auslöseerlebnisses der Angst vor dem Tode (§45, 233) – ausgehend von einem Erwachen aus den alltäglichen Handlungsroutinen und den zweifelhaft gewordenen Üblichkeiten erfolgen. Das echte Eigentlichsein ist also zunächst nur negativ als ein „nicht uneigentlich sein" bestimmt. Wie aber lässt sich nun – in einem zweiten Schritt – der Fehler einer echten Uneigentlichkeit und einer unechten Eigentlichkeit vermeiden? Heidegger verbleibt hier zumeist in einem aktivischen Vokabular, begrifflich strenge Ausführungen zur Kriterien-Problematik finden sich wenige – trifft also Adornos Kritik an Eigentlichkeit und Uneigentlichkeit als inhaltsleeren „überhöhten Edelsubstantiven" zu?[90] Auch in der Literatur, z.B. bei Andreas Luckner, fehlt ein positives Kriterium für das echte Eigentlichsein, das einen Relativismus vermeiden könnte. In einem Moment der Desorientierung, wenn (echtes) Eigentlichsein fehlt, werde deutlich, dass ich mir die Handlungsmöglichkeiten, vor denen ich stehe, zu fast „gar keinem Teil überhaupt, als die meinigen zurechnen kann."[91] Nun müsse man „sich selbst ins Spiel bringen", dabei komme es „ganz auf die jeweilige Person selbst an", nun sie selbst zu sein. „Erst wenn die betreffende Person ein Kriterium der Entscheidung darin finden kann, was sie wirklich und eigentlich will", ist sie eigentlich sie selbst.[92] Doch, wann genau ist das der Fall? Da eine Selbsterzählung echt oder unecht erfolgen kann, ist fraglich, wie ich überhaupt wissen kann, was denn nun *ich* eigentlich will und wer *ich* echt eigentlich bin.

Es liegt eine ähnliche Problematik wie angesichts des Begriffs der Freiheit vor: Immanuel Kant hat gezeigt, dass ein Freiheitsbegriff nach David Hume oder neuerdings Harry Frankfurt als ein Wollen 2. Ordnung nicht hinreichend ist, um von Freiheit in einem echten Sinne sprechen zu können. Denn ich kann mich darüber täuschen, irren oder später korrigieren, dass ich gerade dies eigentlich wollte. Dies führt in die Problematik eines Relativismus des Selbstseins: Ganz gleich, ob das echte Eigentlichsein ein bloßes Moment des Aufwachens

[89] Vgl. Andreas Luckners Kritik an Odo Marquarts Philosophie der Üblichkeiten, Luckner, Andreas: Klugheit, S. 157ff.

[90] Adorno, Theodor W., Jargon der Eigentlichkeit, S. 2; S. 9.

[91] Luckner, Andreas: Klugheit, S. 20ff.

[92] Luckner, Andreas: Klugheit, S. 20ff.

aus dem Alltag oder eine Lebensweise im Ganzen ausmacht, jeweils bleibt die Problematik bestehen, wie der Relativismus der Selbsterzählung vermieden werden kann, ohne ein absolutes und endgültiges Existenzideal (*unechte Eigentlichkeit*) zu postulieren. Denn wenn jede Selbsterzählung gleichermaßen richtig ist, dann ist auch keine falsch – der Finanzbeamte ist also stets echt eigentlich, ganz egal, wie er sich zu den Möglichkeiten$_1$ verhält und wie er sich angesichts der Optionen entscheidet. Das wäre freilich ein äußerst missliches Ergebnis für den Begriff eines eigentlichen Selbstseins, da dieser aufgrund des Relativismus hinfällig wäre: Die Aussage „Das muss eben jeder selbst wissen, wer er eigentlich ist" überzeugt nicht und ist nicht korrekt formuliert, vielmehr müsste es heißen: „Es ist nicht möglich zu wissen, wer man eigentlich selbst ist." Kann die Big Data-Technologie, die bewusste und unbewusste Verhaltensweisen objektiv erfasst und auf Muster zurückführt, den Relativismus der Selbsterzählung vermeiden?

Big Data und Schluss

Das Problem in Heideggers Konzeption des Selbstsein ist also: An den gewählten oder negierten Optionen (M_1) lässt sich nicht ablesen, was eigentliche und uneigentliche Möglichkeiten sind. Eigentlichkeit erfordert, dass ich mich zu Möglichkeiten als meinen Möglichkeiten verhalte – sie erfordert also eine Bezugsweise auf Möglichkeiten. Aber es gibt kein Kriterium, das sicherstellt, dass ich mich zu Möglichkeiten als meinen verhalte. Das Problem könnte auch so verstanden werden, dass Heidegger einen formalen Begriff des Selbst bildet, aber dieser auf einen Übergang zu materiellen Möglichkeiten verweist: denn was *meine* Möglichkeiten sind, lässt sich formal nicht mehr durch Kriterien bestimmen. Ohne Kriterium lässt sich aber gar nicht mehr sagen, was ich bin und was ich nur sein zu glaube. Eigentlichsein und sich Für-Eigentlich-Halten sind ununterscheidbar. Somit ist auch die Möglichkeit eines introspektiven, unmittelbaren Zugangs zum eigentlichen Selbst hinfällig und muss als Illusion abgetan werden.

Big Data-Technologien scheinen der Idee nach eine Lösung aus dieser misslichen Lage anzubieten. Wie gesagt, akzeptieren wir dazu die Annahmen, dass die Technologie – ohne Verzerrung durch einen Beobachter – in der Lage ist, die Muster des eigenen Verhaltens freizulegen. In diesen Mustern findet sich demnach die Spur des eigenen Verhaltens – die materialisierte Geschichte meines Selbst; unverstellt durch Erzählungen, die immer dem Verdacht ausgesetzt sind, dass der Erzähler täuscht oder sich täuscht, in dem die Erzählung den Erzählmustern folgt, die *man* so in Erzählung verwendet und – damit diejenigen vermeidet, die sich etwa nicht ‚schicken'. Man denke an die Erzählmuster von demjenigen, der immer Pech hat, der gegen Widerstände kämpft oder dem unaufhaltsamen Aufsteiger, oder einen, dem durch Zufall alles glückt usw. Solche vorgegebenen Erzählmuster werden durch Big Data vermieden. Eigentlichkeit besteht im Verhalten zu meinen Möglichkeiten – diese zeigen sich doch im Muster meiner bisherigen Geschichte, die unverzerrt in den Daten meines bisherigen Verhaltens verborgen ist und nun durch entsprechende Suchalgorithmen freigelegt werden können.

Auch wenn wir die bislang genannten Voraussetzungen akzeptieren, gibt es eine weitere Voraussetzung, die wir nicht akzeptieren können, wenn wir nicht die von Heidegger freigelegte Struktur des Begriffs Selbstsein verfehlen wollen: die bisherige Geschichte meiner Verhaltensweisen, so unverzerrt und unverfälscht sie auch wiedergegeben werden mag, verbürgt ja nicht, dass ich mich in dieser Geschichte immer eigentlich verhalten habe. Nun mag es zwar aus guten Gründen widersinnig erscheinen, dass ich mich in dieser Geschichte immer nur uneigentlich verhalten haben mag – und sich nicht etwa zumindest Kritik, Regungen, Impulse gegen dieses Verhalten als zumindest Spur meines eigentlichen Selbstseins zeigen.[93] Das Problem bleibt: Woher kann ich wissen, wann ich mich eigentlich und wann uneigentlich verhalten habe. Auch ein Impuls mag aus bloßem Trotz vorkommen – er ist per se nicht eigentlicher als anderes verhalten. Und von diesem anderen Verhalten gilt, was wir weiter oben festgestellt haben: Es gibt kein Kriterium, das zwischen eigentlichem und uneigentlichem Verhalten zu unterscheiden erlaubt. Mein tatsächliches Verhalten daher ungebrochen für mein eigentliches Selbst zu nehmen, führt zu einem Fehlschluss (es wäre vielleicht zu suggestiv ihn als den ‚personalistischen Fehlschluss' zu bezeichnen). Mit Blick auf unser Eingangsbeispiel: Dass der Finanzbeamte

[93] Widersinnig, da ohne Spur das eigentliche Selbst eine Entität ohne Wirklichkeit wäre.

bislang immer alle seine Strebungen nicht durchgehalten hat (außer seiner Tätigkeit als Finanzbeamte), heißt weder, dass er eigentlich eine Art „Quitter" sei, noch dass er eigentlich wäre, wenn er nicht vorzeitig aufgebe.

Ziehen wir ein Fazit: Wir haben die Idee von Big Data und Heideggers Begriff des Selbstsein miteinander konfrontiert und dabei ein zweifaches Defizit freigelegt. Heideggers formale Unterscheidung zwischen Eigentlichkeit und Uneigentlichkeit bleibt ohne Kriterium. Vor allem die bisher in der Literatur nicht hinreichend beachtete Kreuzklassifikation unter Einbezug der Prädikate „echt" und „unecht" konnte dies verdeutlichen. Es ist fraglich, wie ein solches Kriterium ohne Bezug auf materielle Aspekte der jeweiligen Möglichkeiten gegeben werden könnte: Schließlich geht es um *meine* Möglichkeiten. Und dieser Bezug lässt sich ohne Beziehung auf mich und damit meine Geschichte nicht einlösen. Die Suche nach materiellen Mustern des Selbst steht aber vor dem Problem, eigentliche von uneigentliche Mustern zu unterscheiden – und führt damit wieder zu Heidegger zurück. In diesem Sinne profitieren beide Diskurse voneinander, das Ergebnis ist aber wie gesagt, ein zweifaches Defizit. Wir sehen es aber als Vorzug, dieses freizulegen, da der Big Data-Diskurs häufig unterkomplex im Begriff des Selbst bleibt. Umgekehrt treiben viele Interpreten Heideggers Theorie nicht so weit, dass sie an die Probleme herankommen; sie bleiben gleichsam in Distanz, wo die Übersicht gesichert scheint.

Doch wie von hier aus weiter verfahren? Ließe sich nicht vielleicht, wenn nicht auf den Begriff des Selbstseins, so doch auf die Unterscheidung von Eigentlichkeit und Uneigentlichkeit, die so viel Kritik erfahren hat, verzichten?[94] Unserem Verständnis nach hieße das letztlich, doch auf den Begriff des Selbst gänzlich zu verzichten oder dogmatisch eine Behauptung über das Selbst kritiklos als wahr zu betrachten – wobei sicherlich beides als äußerst unbefriedigend erscheint.

References

Adorno, Theodor W. (1964): Jargon der Eigentlichkeit. Zur deutschen Ideologie, Frankfurt a. M., Suhrkamp.

Anderson Chris (2008): The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory, aufgerufen am 24.11.2015.

Bieri, Peter (2012): Das Handwerk der Freiheit: Über die Entdeckung des eigenen Willens, München, Hanser.

Dabrock, Peter (2015): Freiheit, Selbstbestimmung, Selbstkonstruktion – zur Ethik von Big Data. Vortrag auf der Jahrestagung des deutschen Ethikrates. http://www.ethikrat.org/dateien/pdf/jt-21-05-2015-dabrock.pdf. Zuletzt abgerufen am 16.1.2016.

Duhigg, Charles (2012): 'How Companies Learn Your Secrets', The New York Times, 16 February. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all&_r=1.Zuletzt aufgerufen am 28.12.2015.

Figal, Günter (2009): Verstehensfragen. Studien zur phänomenologisch-hermeneutischen Philosophie, Tübingen, Mohr Siebeck.

Gelhard, Andreas (2014): Entgrenzung der Psychotechnik: Der neue Geist des Kapitalismus und das Problem der Prüfungstechniken. In: Kaminski, Andreas; Gelhard, Andreas (Hg.): Zur Philosophie informeller Technisierung, Darmstadt, Wissenschaftliche Buchgesellschaft, S. 185-203.

Heidegger, Martin (2001): Sein und Zeit. 18. Auflage, Tübingen, Max Niemeyer Verlag.

Herrmann, Friedrich-Wilhelm von (2004): Subjekt und Dasein. Grundbegriffe von „Sein und Zeit", Frankfurt a. M., Klostermann.

Kaminski, Andreas (2014): Maßstäbe der Macht. Psychometrische Prüfungen als Techniken. In: Kaminski, Andreas; Gelhard, Andreas (Hg.): Zur Philosophie informeller Technisierung, Darmstadt, Wissenschaftliche Buchgesellschaft, S. 171-185.

[94] Ökonomisch vielleicht, obgleich die Werbung teilweise von der Suggestion der Eigentlichkeit zehrt.

Luckner, Andreas (2001a): Martin Heidegger „Sein und Zeit". Ein einführender Kommentar. Paderborn/München u.a., Schöningh/UTB.

Luckner, Andreas (2001b): Wie ist es, selbst zu sein. Zum Begriff der Eigentlichkeit (§§ 54-60). In: Rentsch, Thomas (Hg.): Martin Heidegger. Sein und Zeit. (Reihe: Klassiker Auslegen), Berlin, Akademie Verlag, S. 149-168.

Luckner, Andreas (2005): Klugheit. New York/Berlin, De Gruyter.

Marquardt, Wolfgang (2015): Was ist Big Data? Versuch einer Positionsbestimmung. Vortrag auf der Jahrestagung des deutschen Ethikrates. http://www.ethikrat.org/dateien/pdf/jt-21-05-2015-marquardt.pdf. Zuletzt abgerufen am 16.1.2016.

Moran, Richard (2011): Authority and Estrangement. An Essay on Self-Knowledge, Princeton, Princeton University Press.

Roßnagel, Alexander (2013): Big Data - Small Privacy? In: Zeitschrift für Datenschutz 3/11 (2013), 562-567.

Wiegerling, Klaus; Heesen, Jessica; Siemoneit, Oliver; Hubig, Christoph (2008): Ubiquitärer Computer – Singulärer Mensch. In: Klumpp, Dieter; Kubicek, Herbert ; Roßnagel, Alexander; Schulz, Wolfgang (Hg.): Informationelles Vertrauen für die Informationsgesellschaft, Berlin, Springer, S. 71-84.

Zwitter, Andrej (2014): Big Data ethics. In: Big Data & Society 1 (2).

Bruno Gransche:

The Oracle of Big Data – Prophecies without Prophets

Abstract:

The need for foreknowledge intensifies and a prophetic promise of today's palm readers causes us wet palms: *letting the world speak for itself*. Big Data comes with the promise of enabling people to listen to that speaking world and of gaining accurate foreknowledge by Big Data predictions. The uncertainty of our modern, complex world overstrains our present coping capabilities, causing a feeling of slipping off a slippery slope, which in turn causes a need for increasing our own foreknowledge. Part of the Big Data promise is to grant better foreknowledge by overcoming the wrongness of scientific theory or causation assumptions. But thus, people have no other option than to *believe* in these results and perform their actions in *good faith*. This makes Big Data based outcomes *a matter of faith*. This article argues that Big Data based outcomes can be seen as today's oracle, as prophecies without prophets and reflects on the consequences of that perspective.

Agenda:

Author:

Dr. Bruno Gransche:

- Fraunhofer Institute for Systems and Innovation Research ISI, Breslauer Str. 48, 76139 Karlsruhe, Germany
- ☎ + 49 - 721 - 68 09 424 , ✉ bruno.gransche@isi.fraunhofer.de, 💻 www.brunogran.de
- Relevant publications:
  - Vorausschauendes Denken: Philosophie und Zukunftsforschung jenseits von Statistik und Kalkül. Bielefeld: transcript 2015, 453 p.
  - Philosophie und Zukunftsforschung. Aktuelle Beiträge zum Begriff der Zukunft. Philosophische Rundschau, 61/3, 2014: pp. 206–227.
  - Wandel von Autonomie und Kontrolle durch neue Mensch-Technik-Interaktionen: Grundsatzfragen autonomieorientierter Mensch-Technik-Verhältnisse. With Shala, E.; Hubig, Ch. et al., Stuttgart: Fraunhofer Verlag 2014, 231 p.

ISSN 1614-1687

Seeking foreknowledge – The perfect conjecture

The future is today's hot topic. Our world is apparently always ahead in time and widely focused on future events and developments. The new is at large considered better than the old, the time to come more important than the past, innovation beats tradition, trend researchers and prediction specialists earn much more money, attention, and appreciation than historians and archaeologists. When tradition and ancient custom were generally held in high regard, they provided a liable orientation and good foreseeability for everyday decision making. Metaphorically speaking: In hardly and slowly changing cities, a ten-year-old street map did just fine – in highly dynamic and rapidly changing environments, old maps turn useless ever faster. With the same pace, our world structures are liquefying, static orientation approaches get obsolete, the need for constantly updated information, predictive efforts and anticipation rises. In that situation of a "liquid modernity"[95], the son of a blacksmith is no longer automatically becoming a blacksmith and he is no longer sure to be needed as such his entire lifetime. Today's decisions – as the choice of occupation – need a great deal of anticipation. Will welders be needed in 2050 or welding robot operators instead? Will more red or yellow shirts be sold next summer? Will male insurance clients keep on causing more severe car accidents than female ones?

The future is considered more important than ever and knowledge about the future seems to be the oil of the 21st century. The problem is: There is no such thing as 'knowledge about the future' in a complex, dynamic and non-deterministic world. Prometheus, the Greek Titan and epitome of science, alone had a perfect foreknowledge of the one single and only possible future in the deterministic Greek mythological cosmos. But – because of that – he could not do a thing to make a difference, to change this future because that would have meant to instantiate a second future different from the foreseen one, which in turn would make that perfect foreknowledge impossible. Prometheus was famous and envied for that foreknowledge and even tortured by Zeus for it.[96] This ancient myth already understood that you cannot have both: Either you can, in principle, gain knowledge of the future (being a Titan, an Oracle or a prophet) and this at the cost of not being able to perform different actions than those that inevitably lead to that one and only deterministic future, or you are free to make a difference, to influence the future, to manipulate or create different alternative futures which comes with the impossibility of foreseeing them.[97]

The future is not the realm of facts but of objectives and ambitions, there is nothing true or false about the sentence 'I will buy a Richter painting.' One can believe in the so communicated plan, maybe even based on whether it is considered probable or not. But believing in a stated ambition, judging future alternatives according to alleged intuitive or scientifically calculated probabilities is far from positive knowledge about the future. The best grasp we can get about the future for epistemological reasons are more or less educated guesses, better or worse underpinned assumptions – that is: conjectures.

All we can do to approximately satisfy our rising need for foreknowledge is to further educate our guessing capabilities and develop our *art of conjecture*.[98] Our ability to perfect this kind of artistry seems to lose – despite remarkable progress – the arms race with the world's increasing complexity. This feeling of slipping off a slippery slope causes a call for new arms in increasing our own foreknowledge, which, in principle, can never

---

[95] See: Bauman 2000

[96] There is one other Greek mythological figure being famous for her accurate foreknowledge of some parts of the coming: Cassandra. So why did Zeus not interrogate this mortal woman instead of meddling with a Titan – who, of course, knew he will be tortured but could not help it? The fact that knowing the future and being free to change it exclude each other holds for Cassandra as well. Here coming with the curse that no one ever believed her prophecies. So Zeus could have extorted the foreknowledge from Cassandra, but in turn he would not have believed her anyway.

[97] According to Kant, foreseeing the future of freely acting people is impossible, or as he puts it: If actions could be foreseen, there would be no freedom. "[U]nd wenn wir alle Erscheinungen seiner [des Menschen, BG] Willkür bis auf den Grund erforschen könnten, so würde es keine einzige menschliche Handlung geben, die wir nicht mit Gewißheit vorhersagen und aus ihren vorhergehenden Bedingungen als nothwendig erkennen könnten. In Ansehung dieses empirischen Charakters giebt es also keine Freiheit" (Kant 1998, 634-635 [577-578]).

[98] Jouvenel 1964.

exceed its conjectural nature but extols itself as predictive knowledge. One approach to ease the disturbing uncertainty of an open future was stochastics and calculating probabilities. Nonetheless, this 'statistical foreknowledge' is still conjectures expressed in numbers.[99] So the need for foreknowledge remains and intensifies. It is causing us – as we are addicted to anticipation – wet palms when we encounter the prophetic promise of today's palm readers.

The promise of Big Data – Listen to the world itself

So far, all efforts to anticipate future developments have been somehow limited by the cognitive capabilities of the anticipator. The ever limited conjecturing ability, even at the level of artistry, falls behind the open future's uncertainty. Models and theories are at the very core of the efforts to deal with uncertainty and to anticipate possible futures. Causality, for instance, insight into causal connections, is probably the dominant way of anticipating future events. Causal connections allow to predict the effect of a certain cause given similar enough circumstances. Models and theories (e.g., probability theory) are what enables the above mentioned 'statistical foreknowledge' but also all sorts of explanation of what might happen based on what happened. A theory based guess – a hypothesis – is considered improved in contrast to a mere wild guess; often its conjectural character is hidden and then called prognosis, forecast, or prediction. On the slippery slope of today's dynamic world, even the most advanced anticipation efforts, even those based on highly elaborate scientific theories, are witnessed to fail epically as seen at the financial crisis 2008. Obviously, our best anticipation capabilities are not good enough for our immense need for foreknowledge and improved theories have not brought a breakthrough so far which leads some to the suspicion that the theory foundation itself might be a shortcoming.

Big Data is now claimed to lessen the need for theories and it comes with the promise of enabling people to listen to 'the world itself'.

> *"The promise is that, with high levels of data generation and developments in computational analysis, the world (coded through datafication) can begin to speak for itself without its (more than) fallible human interpreter."*[100]

Or as put in the much cited article "The End of Theory": "With enough data, the numbers speak for themselves. […] Data without a model is just noise. But faced with massive data, this approach to science — hypothesize, model, test — is becoming obsolete."[101] Even the best models are flawed and "a caricature of a more complex underlying reality."[102] This leads to the promise of a Big Data enabled 'better way':

> *"There is now a better way. Petabytes allow us to say: 'Correlation is enough.' We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can […] let statistical algorithms find patterns where science cannot. […] Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all."*[103]

Reality in its vibrant abundance – so the luring promise – could be accessed through their authentic data, thus circumventing the anemic and essentially curtailed scientific models and theories. Understanding the datafied

[99] All calculations, even the most sophisticated, that distribute probabilities to alternative developments or events still suffer from the flaw that it is a mere guessing how many percent were to be distributed. It has become a habit to distribute 100% making three equally probable events each 33% probable. But given a fourth unknown possible event, maybe only 75%, had to be distributed on the three known events in the first place. Stochastics can provide quite sophisticated information on the known futures, but the number of known futures taken into account is restricted by one's conjecturing abilities. And are not the unknown futures much more in number, thus in likeliness, and – being unknown – causing much more uncertainty?

[100] Chandler 2015, 837–838.

[101] Anderson 2008.

[102] ibid.

[103] ibid.

language of our IT system pervaded world in its alleged original richness with the help of nowadays computational 'superpowers' – such as Big Data algorithms – seems to let the proverbial dream of the emperor who wanted a map of his empire being as detailed as the reality come true.[104] A map provides more orientation than the actual reality because it omits all unimportant details. Concerning these omitted details, the map is wrong, but that is just how it can provide orientation. "Essentially, all models are wrong, but some are useful."[105] – This famous aphorism is about to be outdated as expressed by Google's research director Peter Norvig: "All models are wrong, and increasingly you can succeed without them."[106]

According to that promise of 'Correlation supersedes causation', Big Data algorithms mapping the 'data empire' could lead to such a 'perfectly accurate' map of the reality because they would overcome the constitutive difference between map and empire, between model and world. Therefore, Big Data is becoming notorious for its "unreasonable effectiveness"[107], the 'end of theory', and thus being responsible for the "death of the theorist"[108]. If theory is the base of our best conjecturing abilities and if theory itself is the shortcoming of our anticipatory efforts, does the alleged death of theory then imply the death of conjecture, thus giving room to flawless since theoryless predictions? Does this scientific deicide committed by Big Data finally offer us direct foreknowledge?

Data directly derived from our very movements, actions, communications, interactions, body functions, etc. would allegedly not be distorted by any theory of causation imposed by the people trying to make sense of it. Brave new world, where Big Data systems are used to find correlations that could not have been even searched for. The wrongness of the models does no longer matter if there are no model-based hypotheses guiding the questions and defining what counts as an answer. For these algorithms, there is no such thing as unimportant details because the purpose that it has to prove useful for (such as orientation for a map) is no longer predefined. Big Data is so delightfully longed for because it is expected to give us answers we did not even know the question for, which is to bring digital serendipity to a whole new level. This is just the kind of uncertainty about our futures we are confronted with in our complex world and that stochastics failed to tackle: We need answers even if the questions were already too complex to ask, we need to approach the 'unknown unknowns', the things we do not even know that we do not know them.[109]

The problem with the promise – A matter of faith

Just as the scientific method, the use of theory and models was not just invented as some sort of elitist brain jogging, the 'death of theory' would come with some major problems.

The first problem refers to the misunderstanding that mistakenly identifies the 'datafied world' with the 'world itself', meaning that already the promise of listening to the world itself via Big Data technologies is a modern myth. To state the obvious: Any set of data – no matter how incomprehensibly gigantic – is selective. The promise clearly disregards the fact that data is no pre-social phenomena but always already socially constructed or socially determined in its condition of formation. Data is influenced by people with certain interests and mind-sets and the data producing, collecting, storing, and processing technologies are so as well, thus selecting only data within their sensing capabilities and their scopes, that people with certain objectives and with theories about the means by which these objectives are possibly obtainable designed. The datafied world is distinguished

[104] Lyotard 1984, 55.

[105] Box; Draper 1987, 424.

[106] Anderson 2008.

[107] Halevy et al. 2009.

[108] Steadman 2013.

[109] At this point, the said aspects mainly concerning science reach into political, governance, and resilience debates. See: Chandler 2014. For a quite famous use of the concept of 'unknown unknowns' see: Rumsfeld 2011.

from the world itself, at least by its inscribed traces of theory and models[110]; so claiming the complete death of theory by Big Data analytics is techno-deterministically biased and myopically dealing with the illusion of pre-social objective data.

Big Data systems do not 'kill' the theoretical inheritance of data itself. They do whatsoever circumvent much theoretical wrongness in data collection and pattern recognition, what might be enough for Norvig and others to hold on to that promise. But theory comes in not only in data formation but also at the point where information meets human actors. The problem in having an answer to an unknown question is that you never know how to make sense of it. If the answer is 42, for instance, you are in trouble figuring whether kilo, percent, or years, etc. If this information should make sense and be used to motivate actions, then theory and causality have to be invested by human actors inevitably all along with the allegedly overcome wrongness again. So, at the very moment the algorithmic findings are perceived by human actors, they get subjected to some sort of causal or theoretic assumption – consciously or not, be it in a careful methodologically structured scientific or an intuitive emotional prejudicial superstitious way. For example: If Big Data systems would find a strong correlation between being depressed and being a teacher, and given a will to change that situation, people have to come up with some cause-effect assumption whether the job might depress people working in it or people predisposed for depression choose to be a teacher. In short: Do teachers get depressed or do depressed get teachers? The mere correlation cannot guide any action to solve this problem – theory can. Correlation does not supersede causation if you wish to change something and you need to know how.

In our complex world, human actors are no longer the only entities performing actions or action-like processes. Artificial agents sell stocks, filter and channel information flows, and perform all sorts of actions human actors come to deal with as mere results or as participants in all forms of human-technology interactions or co-actions.[111] If assistive systems give recommendations on how to act according to found correlations (or nudges or forces people in a certain direction by modifying interfaces, contents, systems behavior, etc.), it is crucial to be able to deduce the system's behavior and its underlying processes in order to understand and evaluate the recommendations. If this theory-based validation by people is still possible, then the whole human-technology interaction is still as 'defective' as the theories are. In order to unleash its promised potential to deal with unknown unknowns and to overcome theoretical deficiency, the systems and algorithms have to deal with a data quantity and heterogeneity being impossible for humans to grasp even with much time and effort – which is one definiens of Big Data. Delivering insights – or predictions based on them – that people without algorithmic help could never have found is the alleged potential of Big Data systems and it is at the same time the exclusion of scientific validation because accountability, verification (for the time being), and falsification are essential for science.

When people get confronted with information, processes, part-actions and actions, or results based on Big Data algorithms they have no chance to retrace how these outcomes were generated, what they were based on, and if they are 'true, right, or correct' (if any of these concept applies at all). Thus, in a datafied world widely pervaded with Big Data technology and artificial agents acting on this basis, people have no other option than to *believe* in these results and perform their actions in *good faith*. Within these systems that are claimed to render scientific method obsolete, there is no space for scientific falsification. This makes Big Data based outcomes *a matter of faith*. It is information (or hybrid actions based on this information) coming from a source that is principally obscure to human actors. And at the very moment it enters the human sphere, it becomes an orienting force, guiding people's and agents' actions no matter of their original correctness. As for the claim of 'death of theory', this is where its potential validity ends: Algorithms, systems, artificial agents may be able to perform beyond theory only on the ground of abundant data[112], but human beings are not. When

[110] GPS data, for instance, with which movement profiles can be created inherit assumptions of both the special and the general theory of relativity and, thus, of course, their theoretical correctness and wrongness.

[111] Gransche et al. 2014

[112] NB: This 'beyond theory' refers only to their performance. As well as data, IT systems and artificial agents are no pre-social phenomena but underlie a theory-compromised formation process.

encountering human actors, the (if so ever at all) flawless since theoryless information is 'corrupted' by more or less theory-based interpretations and validations of people before being transformed into actions.

Beyond scientific validation possibilities, Big Data findings are indistinguishable from Big Data creations or data noise artifacts. Given a large enough search room, there are always correlations.[113] Big Data findings and creations have the same potential impact on human behavior: How could people distinguish them in the first place? Those who do not know their difference are forced to believe in both or none equally. If only enough people believe in this guiding character of Big Data based outcomes – and the current hype is strongly suggesting that this is absolutely the case –, then these outcomes develop a self-fulfilling and self-defeating power as known from the respective prophecies.[114] So, does the end of theory correlate with a renaissance of prophecy?

The Oracle of Big Data – Prophecies without Prophets

Wherever people are ignorant, there will be prophets.[115] Scientific prognoses – in terms of probabilities including their range of uncertainty, their limitations of validity, and their condition of formation like transparency about the set of information and hypotheses they are based on – can be used to inform decisions; they lessen people's ignorance if not mistaken in its conjectural nature. Post-theory Big Data predictions, on the other hand, lack this self-referring information. People, nevertheless, using them to base their actions on are ignorant about their range of uncertainty and validity, their formation circumstances, etc. Big Data based outcomes, being a matter of faith, can be seen as today's prophecies. As they are not claimed by deficient mortal beings but by some sort of pseudo omniscient algorithmic deity, they are the paradox of prophecies without prophets. Thus, Big Data becomes some sort of today's oracle, a voice revealing insights and predictions from an abundant yet obscure source that is claimed to be the world itself – or at least as close to it as we can hope to get. And just like the ancient oracles, its power does not derive from any correctness of the content of any single prophecy but from the people believing in it. In contrast to scientific prognosis, which is a matter of doubt, those Big Data prophecies being a matter of faith are immune to critique or falsification. Both the oracles of ancient times and those of Big Data have this immunity in common; the former because they were seen as an authentic direct message from the Gods in a deterministic cosmos, the latter because it is broadly believed to be the world speaking for itself. The actual events either prove the correctness of their prediction or the wrongness of the fallible interpreter.

Prognosis and prophecy are two ways of dealing with future unknowns.[116] The modern approach of prognosis accepts the existence of indispensable unknowns along with the notion of an open future. The ancient prophecies placed all the uncertainty in the impartial human knowledge and misunderstandings of a principally knowable future. If Big Data prophecies take the fallible interpreter out of the equation providing prophecies without prophets, this would not only mean that positive foreknowledge would after all be possible but also even directly accessible. Prognosis and prophecies have similar power as socially effective speech-acts. Prophecies, in addition, have two advantages as powerful speech-acts of which the first is the said bonus of infallibility. The second one is a strong awareness of its circularity, which primarily holds for ancient prophecies. While prognoses inherit the scientific tendency to see themselves as uninvolved observers, as mere describing objective entities, prophecies always included their effect in the prophesied future (Oedipus for instance). That is why we know self-fulfilling and suicidal or self-defeating prophecies but not such prognoses. This valuable

[113] On a global scale, there is a good chance that, every time I breathe in and out, one human being dies and another one is born at the same time what obviously does not make my breath lethal or life giving; nonetheless this correlation could get 'recognised' by Big Data pattern recognition. It is human causal common sense that instantly classifies this correlation as absurd. Who knows how many artificial agents already sold, filtered, channelled masses of stocks, information, services, and wares on that kind of correlation? All we might see is a changed price in the end with no chance to check which correlations lead to it.

[114] Merton 1948.

[115] "Partout où les hommes seront ignorants, il y aura des prophètes", d'Holbach, Paul Henri Thiry, 123.

[116] Esposito 2013.

awareness of circularity is one lesson to learn from the ancient prophecies and should be transferred to scientific prognoses[117] and to today's Big Data predictions.

Conclusion

Big Data systems do not bring the end of theory, but – apart from the theory inheritance within data itself – they postpone theoretical interpretation within the information-action chain to a point where it might cause less wrongness on the one hand but also less possibility to evaluate and correct previous parts in that chain on the other hand. This might lead to problems concerning accountability of co-actions to which a hybrid variety of human an artificial actors contribute. Model wrongness is not overcome but relocated and in disguise, thus withdrawn from scientific critique and improvement processes. Predictions – shifting from prognoses to modern prophecies – change their nature from being a matter of doubt to a matter of faith. As decreased fallibility of prophetic foreknowledge comes with decreased freedom of action (Prometheus) and as the appearance of prophets is connected with increased ignorance (d'Holbach), the renaissance of prophecies should alert a progressive democratic society but yet not lead to defensive overreaction as there are insights to be learned from prophetic future anticipation such as a strong awareness of prediction circularity.

Big Data services are indubitably playing an increasing role not only in science but also in politics and economy as well and, therefore, many questions are to be dealt with. How should a society reintroducing the concepts of oracles and prophecies (even if not under these names) at the expense of scientific methods deal with that kind of strategy shift in approaching complex and open futures? What do powerful oracles and prophecies mean in terms of responsibility, accountability, democracy, resilience, governmental influence, and (self-)governance capabilities? Who and where are the new prophets staging themselves as 'out of the equation' and staging the objectivity of 'the world speaking for itself' while strategically acting from behind the curtain? What do they win with this disguise? Are Google's and other Big Data Titans' imperatives actually a surprisingly honest totalitarian rule – "So, follow the data." – and are they a revealing witness of their potentially hazardous approach on (not) shaping the future – "Now go out and gather some data, and see what it can do."[118]?

References

Anderson, Chris: The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, 2008. Available from <http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory>. Accessed 27 September 2015.

Bauman, Zygmunt: Liquid modernity. Cambridge, UK, Malden, MA: Polity Press; Blackwell, 2000.

Box, George E. P.; Draper, Norman Richard: Empirical model-building and response surfaces. New York: Wiley 1987.

Chandler, David: Beyond neoliberalism: resilience, the new art of governing complexity. Resilience 2 (1), 2014, pp. 47–63.

Chandler, David: A World without Causation: Big Data and the Coming of Age of Posthumanism. Millennium - Journal of International Studies 43 (3), 2015, pp. 833–851.

d'Holbach, Paul Henri Thiry: Le Christianisme dévoilé: In Premières Oeuvres, pp. 94–137. Paris: 1797.

[117] This is widely the case in stock market prognosis, bets, and futures but still rare in scientific prognosis.

[118] Halevy et al. 2009, 12.

Esposito, Elena: Formen der Zirkularität in der Konstruktion der Zukunft. In Prophetie und Prognostik: Verfügungen über Zukunft in Wissenschaften, Religionen und Künsten, edited by Daniel Weidner and Stefan Willer, München: Wilhelm Fink 2013, pp. 325–340.

Gransche, Bruno; Shala, Erduana; Hubig, Christoph et al.: Wandel von Autonomie und Kontrolle durch neue Mensch-Technik-Interaktionen: Grundsatzfragen autonomieorientierter Mensch-Technik-Verhältnisse. Stuttgart: Fraunhofer Verlag 2014.

Halevy, Alon; Norvig, Peter;Pereira, Fernando: The Unreasonable Effectiveness of Data. IEEE Computer Society, 2009, pp. 8–12.

Jouvenel, Bertrand de: L'Art de la conjecture. Monaco: Éditions du Rocher 1964.

Kant, Immanuel: Kritik der reinen Vernunft. Hamburg: Felix Meiner 1998.

Lyotard, Jean-François: The postmodern condition: A report on knowledge. Manchester: Manchester University Press 1984.

Merton, Robert K: The Self-Fulfilling Prophecy. The Antioch Review 8 (2), 1948, pp. 193–210.

Rumsfeld, Donald: Known and unknown: A memoir. New York: Sentinel 2011.

Steadman, Ian: Big data and the death of the theorist, 2013. Available from <http://www.wired.co.uk/news/archive/2013-01/25/big-data-end-of-theory>. Accessed 27. Sep. 2015.