Article

Beyond data as representation: The performativity of Big Data in surveillance

Tobias Matzner

University of Tübingen, Germany.
tobias.matzner@uni-tuebingen.de

Abstract

The article discusses problems of representational views of data and elaborates a concept of the performativity of data. It shows how data used for surveillance contributes to the creation of suspect subjectivities. In particular, the article focuses on the inductive or explorative processing of data and on the decoupling of data generation and analysis that characterize current uses of data for surveillance. It outlines several challenges this poses to established accounts of surveillance, including David Lyon's concept of "surveillance as social sorting" and Haggerty and Ericson's "surveillant assemblage." These problems are attributed to a representationalist view, which focuses on the veracity of data. This can lead to ignoring problematic consequences of surveillance procedures and the full scope of affected persons. An alternative account of data as performative is proposed, using Judith Butler's concept of "citationality." This account shows how surveillance is entangled with the production of subjects through data in general. Surveillance is reformulated as a particular way in which subjects are produced that is parasitical to other forms of subjectivation.

1. Introduction

Early in 2014, a kind of supergroup of scholars of Surveillance Studies and related fields published an article rethinking the impact of surveillance after Snowden (Bauman et al. 2014). Among many other important topics, they focused on the data-based surveillance by the NSA and its allied secret services. Regarding so called "meta-data," the authors argued:

Given the magnitude of the data thereby accumulated, analysts do not read all the content, but rather visualize the graph of the relations that are identified and focus on what seem to be the most significant sections showing specific nodes of connections between data. This is far from a full reading of the contents of such data. It is also far from a scientific procedure which might warrant claims to certainty and precision about the results obtained. It is, rather, part of a process of intuition and interpretation that may vary considerably from one analyst to another. *Fears about Big Brother are thus largely irrelevant.* (Baumann et al. 2014: 125; emphasis mine)

This statement is exemplary for a certain strand of debates about surveillance based on the technologies and practices subsumed under the label "Big Data": The authors analyze the certainty and precision of the data and of claims derived from it. These insights are used to judge the potential for surveillance the data provide. Most of the time, the uncertainty of knowledge based on data is directly linked to diminished capabilities of surveillance. While such detailed scrutiny of data analysis in surveillance contexts is

Matzner, Tobias. 2016. Beyond data as representation: The performativity of Big Data in surveillance. *Surveillance & Society* 14(2): 197-210.

<http://library.queensu.ca/ojs/index.php/surveillance-and-society/index> | ISSN: 1477-7487

© The author(s), 2016 | Licensed to the Surveillance Studies Network under a [Creative Commons Attribution Non-Commercial No Derivatives license](#).

important work, a critical account of Big Data must not stop there. The authors continue their reflection by arguing: “What is at stake in this respect is less a marriage between technology and a science of society, more one between technology and a speculative faith in systems designed to ‘read’ big data” (Baumann et al. 2014: 125). It is necessary to emphasize that particularly the NSA and other secret services can have a strong negative influence on the lives of individuals based on this “speculative faith.” Thus, these technologies do have effects that cannot be countered by just undoing claims to the veracity of the data or the adequacy of the methods of analysis.

Other scholars in Surveillance Studies evoke the image of Big Brother to argue that such a powerful—but centralized—figure is insufficient for grasping the possibilities of data-based surveillance. For example, Haggerty and Ericson in what would become one of the elementary texts for research on data-based surveillance write:

In the intervening decades, however, the abilities of surveillance technologies have surpassed even his [Orwell’s] dystopic vision. Writing at the cusp of the development of computing machines, he could not have envisioned the remarkable marriage of computers and optics which we see today. (Haggerty and Ericson 2000: 606)

We have to add that “today” was the year 2000 when people started to become familiar with a search engine called Google, and Facebook had not even been invented. Yet, the authors’ relating of emerging surveillance capabilities to the progress and “development of computing machines”¹ points to another important problem when assessing surveillance and Big Data. “Computing machines” did not only enable surveillance but brought forth many aspects of what today is central to quality of life (at least in Western countries). Many applications we might not want to miss yield data that create potentials for surveillance. Often, it is impossible to disentangle the functions of a service based on data and the potential for surveillance it yields (Matzner 2014). This problem extends far beyond the quality of data analysis. It concerns questions of how technological developments with far-reaching social consequences are entangled with potentials for surveillance. I will show that these problems persist even if the data yield at best dubitable insights about persons and the world. The effects of surveillance can only be scrutinized when taking into account what the practices involving data actually *do* and how they are related, regardless of whether the quality of the data and its analysis withstands scholars’ scrutiny or not.

Those concerns are the points of departure for the following inquiry towards a suitable theoretical approach concerning the implications of surveillance enabled by Big Data—based on a performative concept of data. The article proceeds as follows: In section 2, I will briefly discuss a definition and important features of Big Data, pointing out two characteristics that are particularly challenging for established theories of surveillance. I will discuss this with regard to David Lyon’s concept of surveillance as well as Haggerty and Ericson’s “surveillant assemblage.” I will then go on to discuss the epistemic issues of Big Data, which usually form the background for arguments on Big Data and surveillance, in section 3. I show why these concerns do not suffice to grasp the problems of surveillance, since they adhere to what I call a *representationalist* view of data. After discussing the problems of this view in more detail in section 4, I propose an alternative view of data as *performative* in section 5. This enables the analysis of the parasitical relation surveillance has to other forms of subjectivity in section 6.

¹ Although this is a rather old-fashioned term in engineering and computer science I like its application in this context since we are no longer only dealing with desktop PCs, portable computers, and smartphones but literally with computing machines of all sizes and capacities by now.

2. Big data and new challenges for theories of surveillance

“Big Data” is used in many contexts, ranging from research to politics and marketing, which has led to a variety of definitions. Many of these refer to data with “high volume,” usually in a scale that requires sophisticated data centers to be processed; “high velocity,” regarding the fast and almost real-time production of the data; and “high variety,” concerning both the content and the formal structure. Kitchin (2014: 1-2) synthesizes the diverse definitions that can be found in the literature by adding “exhaustive in scope,” “fine-grained in resolution and uniquely indexical in identification,” “relational in nature,” “flexible,” and “scaleable.” These features of data are interdependent with changes in how data are generated and analyzed. For the following argument I want to single out two particular characteristics that are usually related to Big Data: First, the inductive or explorative processing of data; and second, the decoupling of data generation and analysis.

Concerning the first point, I will discuss practices and algorithms that do not look for fixed categories or properties in the data, but which are *meant to* derive sensible categories and new interesting relations between the data. I highlight the expression “meant to” in this sentence, since many of these claims concerning data analysis are controversial—see for example the opening quote—and a lot of important work is done to ascertain what a data-based epistemology actually means (Floridi 2012; Kitchin 2014). However, in this article I discuss which role this criticism plays for inquiries in terms of surveillance and I will show that it is important to take consequences of surveillance into account that do not only depend on the actual information provided by the data, but by the larger social and political setting. This setting is not only influenced by the current state of technology but also by its projected functions.

Big Data is often discussed in terms of finding correlations (usually with a quip on mistaking them for causation). Yet, I want to take into account a lot of serious research in statistics as well as computer science that tries to establish sophisticated methods for the analysis of data which engage not just correlations but more complex patterns, regularities, and associations. All of these methods are inductive or explorative in a broad sense: Rather than looking for something in the data, the data tell us what to look for. This pertains to methods in machine learning, knowledge discovery, pattern recognition, and many more disciplines.

Concerning the second point: The high variety, relationality, and flexibility mentioned in the definition of Big Data fit in with the decoupling of data generation and analysis. Many methods related to Big Data try to tie in all kinds of sources that might be useful for the desired outcome. This can be data that has been generated at different times, different places, and in different contexts compared to where the analysis takes place. Consequently, a rule of thumb to store all available data has emerged, since it might be useful in the future. Sometimes this is data that is necessary for the operation of the respective service, sometimes it is data that is just generated for potential future applications. Concerning surveillance, this leads to a phenomenon which could be called “prospective surveillance”: huge databases that are just stored for the time being—but with the possibility to be used for purposes of surveillance at any time in the future. For example, data-based border controls do not only store data about the eligibility of persons to enter a country. Instead, they access a wide range of data and patterns, such as booked flights and corresponding payment details or “alerts for firearms and criminal records, missing and wanted persons, stolen vehicles and witnesses” (Adey 2012: 196-7). Only upon the request of a person to cross a border, a query is formed that analyzes this data pertaining to the subject in question. So rather than just checking a piece of knowledge—eligibility to cross—the data-based border is recreated ad hoc for the particular individual: “[T]he sovereign decision is repeatedly performed and re-performed in the remaking of the border by distributed border agents and personnel [...]” (Adey 2012: 196) As I will show below, this does not only mean the re-performance of the border, but also of the subject that wishes to cross.

Describing Big Data as high velocity and high variety also points to the fact that the data can constantly change, e.g. social networking sites or location tracking. The decoupling of the generation and analysis of data does not just mean that data from one context are analyzed in another context. It also entails that data become data for a program only in the very moment a request is made to that program. In this moment, the current state of the data is accessed and influences the analytics. But that is a state that may well have been different in earlier requests and change already for the next one. In this context, data is not just raw material that can be stored and copied until it is analyzed, which then extracts information from it. As Lisa Gitelman (2013) concisely formulates: “raw data is an oxymoron.” Data are what they are only in practices of data creation and use. And when two of these practices are put in touch, this influences both of them, because this is not just using information from one context in the other. Rather, in a sense, the practice of data use is enlarged, influencing the way its meaning is established. So it matters if data-based practices become entangled with surveillance.

Data used for surveillance often accrue in contexts that relate to activities and possibilities many people enjoy or even consider a basic service, e.g. social networking, cell phones (especially, but not only the “smart” type), having electricity at home, driving cars, getting health services, working, and all the other things enabled by “computing machines” today. Surveillance is not a separate application of Big Data but a potential that emerges from all uses of data analysis (Amoore 2011), which lately has been voiced in the catchphrase of “surveillance as the business model of the internet” (Rashid 2014).

Broad concepts of surveillance like the “business model of the internet” or surveillance as social sorting are deliberate and have their reasons. David Lyon defines surveillance as the “systematic, routine, and focused attention to personal details for a given purpose (such as management, influence, or entitlement [...])” (Lyon 2014: 2). Degli Esposti expands this definition to dataveillance: “the systematic monitoring of people or groups, by means of personal data systems in order to regulate or govern their behavior” (Degli Esposti 2014: 210). Such processes play an increasing role in almost any part of our lives. The way we network, apply for jobs, find partners and friends, choose consumer products, dress, eat, etc. is increasingly structured by the collection and analysis of data. Systematic, routine, and focused attention to personal data discerns suspects, but also friends, lovers, job candidates, or valuable customers. A critical perspective on Big Data has to enable distinctions of these modes of subjectivity and the new relations between them, which are brought about by Big Data. Thus, for the rest of this paper I want use the term surveillance for the discerning of *suspects*.² Whether or not to also call the systematic, routine, and focused attention to personal data in order to differentiate partners, friends, consumers, etc. surveillance, but surveillance for a different purpose, is maybe a more political and rhetorical question. I think both approaches are conceptually sound. I choose to focus on surveillance as suspecting, since this concerns a negative attribution.³ In contrast, many other forms of subjectivity for which Big Data plays a role are actually endorsed. This is important when analyzing the effects that legitimize or make tolerable the uses of quite different sources of data for surveillance (as suspecting). Furthermore, fusing all these processes under a general term of surveillance (like surveillance as the business model of the internet) might lead to an impression of hopelessness or over-alertness. In a quotidian perspective, countering surveillance connotes *evading* to be recorded, filmed, evading the production of any data. Contrary to this focus on *what is represented* by data, Surveillance Studies has shown *how* data is used and *how* categorization takes place. I think that this theoretical movement has to be pursued further by differentiating surveillance as suspecting as a particular way of data use from other uses of data and data-based subjectivization. This perspective necessitates the reconsideration of a particular aspect in accounts on how data is used:

² Or, more precisely in the terms I develop below, the performative creation of suspect subjectivities.

³ This distinction between positive and negative connotations concerning those *under* surveillance (e.g. suspect of a crime vs. valuable customer) importantly differs from attributing a positive or negative value to surveillance *itself*, as e.g. in Fuchs (2011).

overcoming what I call representationalist accounts of data. This will be explained in detail in the following section.

3. Epistemic concerns and the representationalist view

Generally, three ways of how information is gathered can be discerned:

1. Hidden generation of information;
2. Open, perceptible generation of information;
3. Information actively provided.

The first pertains for example to the surveillance programs by secret services that are intended to remain hidden. But targeted advertising also falls in this category. Even though with enough time and resources a lot can be learned about these practices, they remain in the background of another activity, e.g. surfing the web for news. That is, an active initiative is required to know about this kind of data collection, together with resources and the competence to do so. In this sense it is hidden.

The second category pertains to all means of surveillance that are meant to employ the psychological effects of being watched. Here, beyond the information gathered by surveillance, creating the consciousness of being under surveillance is an important factor. Fake surveillance cameras meant to deter criminals are an extreme instance of this aspect. The various airport security procedures are another example of this category.

The last category has two sub-categories: information provided voluntarily and information provided yielding to soliciting or forcing disclosure. The first pertains for example to information in social networking sites. Of course, not all of this information is given voluntarily but for example in response to all kinds of social pressures. Incentives like using “free” services or just the effects of advertisement and PR might succeed in soliciting the provision of information. Examples for stronger means of solicitation in this sub-category are subpoenas or searches.

Regarding Big Data, usually the first and the last categories apply. Most probably, the revelations about data collection by states and big companies alike have deterring effects for using some kinds of information and communication technologies. But they are rarely publicized to achieve this very effect by the surveilling institutions—which would fall under the second category. Usually they have to be unveiled by whistleblowers or extracted from the small print of end user licenses by persons equipped with the appropriate knowledge.

Here I want to focus on problems that arise when combining data generated in the first mode with data generated in the third mode for purposes of surveillance. What I reveal about myself, e.g. on a social networking site or in an e-mail, is directed at a more or less clear-cut idea of one’s audience and usually has a more or less conscious aim. Such aims might be as diverse as looking good in an application process, flirting, trying to conform to the social expectations of one’s peers in terms of look or language, making a political statement, telling a joke, and many more. Information gathered covertly, to the contrary, usually tracks a device like a mobile phone or a piece of software such as a particular web browser. Only by implication is this related to a person: my shopping habits, for example, might show regular purchases of books I would never read but my friends like as presents. If information from both modes is combined, a more or less conscious self-perception is contrasted with a more or less precisely induced external perception. Here, an important aspect of all uses of Big Data is particularly salient. To understand the data, one needs circumstantial knowledge of the social norms and practices that structure the context in which the data has been generated. This has led to serious challenges concerning the epistemic status of Big Data (boyd and Crawford 2012; Floridi 2012; Kitchin 2014; Manovich 2012).

Very often, they show that claims to the information to be gained from data is not warranted, or they are wrongly interpreted. I want to emphasize again that this is an important task of critical discussions of Big Data. Yet, concerning surveillance, it is not enough, for two reasons: one pragmatic, one systematic.

The pragmatic reason is that such a critique actually might implicitly legitimate surveillance. If a practice of surveillance is criticized because it is based on wrong judgments of data, this can mean (or be read as): the surveillance would be all right if these failures did not happen. Stories so much loved by the popular media of data based criminal investigations leading to the persecution of children, animals, dead persons or otherwise completely uninvolved persons are especially prone to this effect. Furthermore, such a critique might lead to requests for even more data to better grasp the social context and enable a more fine-grained analysis.

The systematic reason regards the question: What does it mean that surveillance is based on shaky epistemic grounds? Further questions suggest themselves: Is surveillance that does not produce certain knowledge about those under surveillance actually surveillance? But this line of inquiry has the wrong focus, for the simple reason that surveillance practices based on Big Data are out there and running full force on all scales. On a daily basis people are denied access to countries, are refused asylum, cannot board airplanes, do not get insurance, are arrested, etc. based on Big Data. So rather than asking for the knowledge such systems produce I want to focus on the subjectivizing effects of surveillance processes. And part of the claims to justify or legitimize such subjecting practices are based on the data such systems can *access* (rather than the way they are analyzed): personal, self-provided data, i.e. data collected in the third mode. Yet, it is judged in combination with hidden data gathering.

On a different scale, many systems that engage Big Data for surveillance do not aim at producing knowledge in the sense of certain facts about a person. They single out suspects as potential or possible future wrongdoers. Often, such systems produce visual incentives or even instructions to act, but in the absence of information how this verdict came to be (Leese 2016). So rather than just aiming at “actionable intelligence,” which Gandy (2012: 125) sees as a result of data-based surveillance, action and intelligence (viz. knowledge) are fused into machinic judgments that engage both security personnel and subjects of surveillance to act (Leese 2016).

To consider these factors, what I call the representationalist view of data does not suffice. Most applications of Big Data are developed, marketed, and used on the premise that the data represent a certain aspect of the world in computable form. The very fascination (and market value) of Big Data stems from the claims that using these technologies we might gain even more detailed and comprehensive knowledge of the world than ever before. Many critics of these epistemic claims implicitly acknowledge that the aim or purpose of Big Data is to represent the world—they just show that the data do not represent what it is meant to. This is the representationalist view.

But from a political or critical stance the question should be: What happens when databases and algorithms are in place that are meant to judge risk and suspicion? What do these practices do to subjects, beyond the question whether the judgments are derived from an appropriate analysis of data? What do automatic, data-based incentives to act mean for subjects under surveillance?

4. Challenging the representationalist view

A first approach to put data-based surveillance in a larger context is based on Deleuzian theory. In this vein, Haggerty and Ericson describe the “surveillant assemblage,” which “operates by abstracting human bodies from their territorial settings and separating them into a series of discrete flows. These flows are then reassembled into distinct ‘data doubles’ which can be scrutinized and targeted for intervention” (Haggerty and Ericson 2000: 606).

An important difference from the representationalist view is a clear grasp of the consequences of the decoupling of producing data—described as abstraction from context or in Deleuzian terms “territorial settings”—and the uses made of these data. The use of the data is no longer seen as *discovering* or *extracting* something from the disparate variety of data but of *assembling* a so called “data double.” These data doubles are often created on-the-fly and tailored to various purposes of scrutiny, attention but also intervention (Amoore 2011). This entails that many and different data doubles pertaining to one person can exist. Deleuze captures this by coining the term “dividual” in his “Postscript on the Societies of Control” (Deleuze 1992: 5). Dividuals are created from the diverse flows emanating or abstracted from one individual, thus producing an entity that is “made up” only of particular aspects of an individual. In this text, Deleuze analyzes an important shift from the disciplinary societies as described by Foucault (1975), which brought about the modern individual in institutions like schools, hospitals, prisons, and the military. These rather centralized normalizing disciplinary processes have been replaced by a much more fragmented and localized requirement to function, which is established and controlled by information and communication technologies (among other factors). Rather than being subjected to discipline an imperative to conform to changing requirements and control mechanisms governs, where a coherent “normal” subject is no longer important. Amoore (2011), with recourse to Deleuze and Massumi (2007), coins the notion “data derivative,” similar to the data double, but with an emphasis on temporality. Like the derivatives in finance, “data derivatives” are a bet on a potential future, they are assembled to judge the possibility that something will happen. So data-based verdicts do not only concern who someone is, but who this person will (potentially) be. This underlines that current, data-based modes of surveillance inherently deal in suspicion. They assemble resources of data that will be used to judge the suspects to come (after a crime has happened) but also to already identify those who potentially will commit a crime in advance.

Such Deleuzian theories are an important step away from representationalism. But with the salience of abstraction on the one hand and dividuals or data doubles on the other hand in Deleuzian texts, there is a risk of falling back into a kind of normative representationalism. Talking about the many data doubles or dividuals allows positioning the “original” individual against these constructs. Even if they are no longer criticized as wrong representations but as constructions, they can be seen as an imposition on the “original” individual. Thus, not so much the descriptive analysis but the normative core of such arguments follow a representationalist logic. An instance of this can be seen in David Lyon’s study of the implications of Big Data for surveillance. Referring to Haggerty and Ericson as well as Deleuze, he writes:

Surveillance in the era of Big Data [...] does not focus only on the body or on a population but on definitions to which we may contribute as part of our daily online interactions. It “makes up” the data double, Deleuze’s “dividual” and that entity then *acts back on those with whom the data are associated, informing us who we are, what we should desire or hope for, including who we should become.* (Lyon 2014: 7, emphasis mine)

Here, a somehow independent “we” or “us” is opposed to the data double, which is an imposition when it “acts back on us.” A similar, but already more ambivalent motive appears in Amoore’s text, when she discusses the use of data that accrue in our daily lives for surveillance: “The derivative risk form acts through us and the prosaic, intimate, banal traces of our lives, but yet it forgets us” (Amoore 2011: 39). Similarly she describes using data as “digitized dissection” of a person into “degrees of risk” (Amoore 2011: 35). Such a focus on the subject as origin of data is a problematic normative stance for two reasons.

First, as I pointed out in the characterization of Big Data in section 2, data are combined, related, linked, correlated, with a lot of other data from various sources. Often, this is highly abstracted or depersonalized data used to identify the patterns, links, and regularities rather than information about concrete individuals. *Such* data, however, are the base of the problematic consequences of data analysis mentioned above: being

denied access, entrance, being held and searched, etc. (Amoore 2011; Leese 2016; Adey 2012). Often these applications concern persons who are not at the “origins” of the flows of data. As described above, generation and analysis of data are decoupled. Data that is collected at one place and from a particular set of persons, contributes to assessing and judging different persons in different places. This has two important consequences. First, the focus on the “origins” of data often misses those that suffer most from a data based surveillance infrastructure. For example, data collected from “orderly” citizens can be used to form patterns that allow identifying those that do not fit this verdict (Matzner 2014). Second, surveillance does not always entail the reassembling of a personal representation or data double which then—wrongly or rightly—imposes suspicion on a person. Often, there is no such representation in the first place but just a vague connection and pattern which—sometimes ad hoc—leads to data based judgments. So we are dealing with a mixture of data derived from several persons (not just “our” data doubles), but also patterns, and possibilities that do not deal in individual subjects at all. But in the moment of algorithmic judgment and according action, e.g. the electronic border control or airport security discussed in section 2, they are producing strong subjectivizing effects. So the data subject in question is not a product of splitting a person into data doubles but an instantaneous combination from various sources which I describe below as “citation.”

Second, the increasing permeation of our lives with information and communication technologies leads to the fact that individuals do not only figure in surveillance apparatuses. The collection and analysis of data increasingly structures the way we network, apply for jobs, find partners and friends, choose consumer products, dress, eat, etc. Data doubles are not only assembled to produce suspects but also friends, lovers, job candidates, or customers. With a grain of salt: the original individual against which a data double could be compared is—to an increasing extent—also a data double. So, rather than discussing how data doubles act back on subjects, I propose to focus on how different moments of subjectivation through data relate. Concerning surveillance I identify a particular relation I call “parasitical” below. To achieve a better grasp of this predicament I propose a different way of describing these situations: a performative rather than a representationalist account that analyzes subjectivation by data.

Before outlining this account in detail, I want to comment on another strand of Deleuzian theory. The normative representationalism I have discussed derives mainly from reflections on the “Postscript on the Societies of Control.” But in Deleuzian theory more generally, there would be no subject to compare the data double to. In fact, many authors who refer to Deleuze endorse his thought in particular for overcoming the notion of a subject and the associated “spurious sense of oneness” (Thrift 2007: 7). This supports positions on surveillance from different theoretical backgrounds that, however, express the similar view that governance through data is no longer focused on subjectivation at all. For example, Rouvroy (2013: 157) argues: “Algorithmic governmentality does not allow the process of subjectivation to happen, because it does not confront ‘subjects’ as moral agents [...] but attunes their future informational and physical environment according to the predictions contained in the statistical body.”

Yet, I think this dismissal of the subject is too quick. In a sense, with Deleuze, the subject is gone, so either one can criticize its loss as origin or one needs to find a different basis for critical inquiries. Contrary to this alternative, I think that it is important to note that data-based surveillance and other practices involving data produce subjectivizing moments. But these are ad hoc, heterogeneous moments of subjectivation, not aimed at the coherence or oneness which the authors from Deleuzian strands of thinking rightly suspect. So while Thrift and other Deleuzians provide a valuable framework (although that sounds too rigid to do justice to their theories) for the social or socio-technical environment of surveillance, I think it is equally important to look at the effects that data-based surveillance has on subjectivation.

Consequently, the following account provides two veins of analysis: making sense of heterogeneous, non-coherent modes of subjectivation and comparing the relations of different modes of data-based subjectivity.

5. Data as performative

The conceptual inspiration for this account stems from Rita Raley's analysis of *Dataveillance and Counterveillance*. Elaborating on both Haggarty and Ericson's theory and the history of the concept "data" she arrives at this conclusion:

Our data bodies [...] are repeatedly enacted as a consequence of search procedures. Data is in this respect *performative*: the composition of flecks and bits of data into a profile of a terror suspect, the re-grounding of abstract data in the targeting of an actual life, will have the effect of producing that life, that body, as a terror suspect. (Raley 2013: 128, emphasis mine)

The backdrop of Raley's account is the vast amount of data that accrues when browsing the web or using smartphones and the capacities for "dataveillance" which arise from this data. Yet, such an inquiry must not be restricted to surveillance. Otherwise we might still arrive at a picture where a surveillance apparatus produces deviant lives and suspect bodies based on data while "we" would again be opposed to or resisting this production. The consequences of the pervasiveness of computing machines, which I emphasized in section 2, now become salient. Many of our social interactions yield data and they are increasingly structured by data: who to be friends with, what to do together, which holiday pictures of our acquaintances to watch—we get suggestions for such things and many more based on the analysis of data. In this context, many of us consciously manage the distribution, appearance, and context of data—that is, we are staging data performances. Data does not only produce suspects, but also potential partners, employees and customers. Claims to truthfulness are part of this game—not its precondition as the representationalist account would have it. Data can produce impostures and transparency, show-offs and "open books," facts and lies. Here, not only the data itself is important. The very fact whether or not we appear in a particular way in digital communication—on a social networking site, in a texting service—contributes to who we are.

In the remainder of this paper I want to spell out in greater detail how data can be conceptualized as performative, using a well-known concept of the performative production of subjects and their bodies. In her seminal books *Gender Trouble* (1990) and *Bodies that Matter* (1993), Judith Butler develops a theory of performativity that accounts for the production of subjects regarding both gender and sex, thus social and bodily features. I will apply this account to data. In particular, I focus on one aspect of Butler's rich theory, namely the concept of "citationality."

Butler develops her theory based on J.L. Austin's (1975) concept of performative utterances. Such utterances produce what they name (Butler 1993: 224). Well-known examples are christenings, formal appointments, marriages, or legal sentences. Butler extends this performativity to all speech acts, in particular those that "only" describe things. To quote a famous example, when a physician looks at a newborn child and declares: "It's a girl," this is no simple diagnosis but the start of a process of "girling," i.e. a series of speech acts that produce a gendered and sexed subject (Butler 1993: 232). For this extension of performative force to all kinds of speech acts, Butler takes recourse to Derrida's critical discussion of Austin in an essay called *Signature, Event, Context*. The text shows that the performative power of an utterance neither stems from the intention (Derrida 1977: 14-15) nor the authority of the speaker as Austin has claimed (Butler 1993: 226-7). In contrast, the meaning and the performative power of the speech act stem from its relation to similar acts, where particular structures of power and authority are already established. Butler and Derrida discuss this as "citation" (Derrida 1977: 17; Butler 1993: 227).

This is no simple linguistic citation but re-instantiates power structures. A performative act succeeds because it “accumulates the force of authority through the repetition or citation of a prior, authoritative set of practices” (Butler 1993: 226). For example, there is no powerful authority or ideology that determines what it means to behave like a woman. But the many quotidian gendered acts produce a regularity which then enables both the clear appearance of norm and deviance as well as the power to request according behaviour—at least for some. Importantly, in this perspective the performative, subjectivizing effects of utterances are not limited to subjects performing themselves. To the contrary, many important moments of subjectivation that Butler analyses are “interpellations”: others call a person a woman, treat her as a woman, demand reactions expected from a woman. All those utterances by others have subjectivizing force, which they acquire from “the repetition or citation of a prior, authoritative set of practices.”

Since the legitimacy of claims brought about by a speech act is enabled by other situations and contexts, which are related to the current act, Butler can fuse this linguistic perspective with a Foucauldian account of power:

This view of performativity implies that discourse has a history that not only precedes but conditions its contemporary usages, and that this history effectively decenters the presentist view of the subject as the exclusive origin or owner of what is said. (Butler 1993: 227)

This history of the performative is structured by power relations, which—via citation—infuse a speech act with the very power to create a subject. This structural power, however, has an important consequence: “citations” are not arbitrarily available to everyone. They require the accessibility of that history, the contexts from which to cite. Here, it is important to bear in mind the productive force of power in Foucauldian theories. There is no *pre-existing* subject, which is then told who to be. The subject is produced by power. This is the very point where Butler departs from the structure that I call representationalist: “There is no ‘one’ who takes on a gender norm. On the contrary, this citation of the gender norm is necessary in order to qualify as a ‘one,’ to become viable as a ‘one’” (Butler 1994: 232). This performative effect is based on the citation of other acts. And the subject has to cite these norms as well, in order to count as “one” in a given context. But such a self-citation is closely intertwined with “interpellations” that already cite the norm as well—and by that very force may elicit the subjectivizing acts.

Using Butler’s theory, the performativity of data and its effects on subjectivity can be analyzed with the necessary scope. My observation from above, that data do not only produce suspects, but also friends and lovers, colleagues and acquaintances, directors, and employees can now be detailed with Butler’s terminology. For example, many academics have long internalized the need to cite certain norms to become a particular subject: an aspiring or established researcher. This is increasingly a performativity of data like CVs, publication list, impact factors, citation ranks, etc. Moreover, it is to a growing extent structured by algorithms that, for example, watch research networks and suggest who to follow or what might be an interesting reading. Social networking sites, or dating apps like “Tinder,” are part of other performatives of data, which yield subjects as interesting friend or prospective sexual partner. And of course gender itself is part of the subjectivities produced by data. So what I described as self-provided data above in section 3 can now be understood as the engagement of many kinds of data in performances of subjectivity. We cite all kinds of data in becoming who we are.

Regarding surveillance, the interpellative side of the performativity of data becomes salient. The recombining, relating, and moving to different contexts of data, which happens in data-based surveillance, does not primarily mean a problematic distancing from an originary subject. To the contrary, this process assembles the authority to produce a new subject—in the case of surveillance by “calling” it a suspect. So, similar to Butler’s description of “girling,” surveillance is “suspecting.” But neither of these practices

invent subject positions but “cite” and thus re-install them. Somebody who is stopped at a border, denied a visa, or excluded from boarding a plane based on Big Data becomes a subject for the respective authorities in the very moment these verdicts happen. It is quite probable that there has been no prior relation of these authorities and the subject. A lot of the data-processing goes on in the background and does not engage with subjects at all, nor does it concern data about subjects. Above I have discussed Rouvroy (2013) and others who focus on this aspect of Big Data. But such data-processing assembles the resources for ad hoc, heterogeneous acts with strong subjectivizing force. The promise of Big Data is to have enough data ready—enough resources to cite—that allows to “judge” everyone. A pertinent example of this promise is IBM’s claim to be able to automatically assess the numerous refugees trying to enter Europe after the crisis in the Middle East (Levinson-Waldman 2016).

This shows that the power of suspecting does not stem from “diagnosing” a particular feature of a person—which then could be criticized for not correctly representing this feature in data. To the contrary, data-based surveillance is legitimized by the circumspect collection of data which are then put into new relations. It is based on the power relations that emanate by the possibility to cite a wide range of data in contexts of effective authority and governance. This is illustrated quite well by the fact that surveilling agencies usually do not only ask for more sophisticated algorithms or better analysts but primarily want more data.

This adds an important new perspective to the research on data-based surveillance. An important line of enquiry has analyzed the impact of algorithms on surveillance, or even “algorithmic surveillance” as a subject of its own (Introna and Wood 2004). This focus on algorithms has also been taken up in research on preemptive digital surveillance which is based on risk calculus (Amoore 2011, 2009). In this context, Butler’s and Derrida’s reflections show that no intentional, authoritative actor is necessary to engage the force of citationality. Thus, they are well applicable to the automatic, algorithmic analysis of Big Data, even by quite simple algorithms that are far removed from any “smart” or (artificially) intelligent actor. Butler’s account shows that such an agential position itself is established by power relations that—among others—manifest themselves in the possibility to access a history of performativity and to cite it. It also manifests itself in the necessity to cite a particular norm to count as a subject in the first place.

6. Surveillance as parasitical production of suspects

Since many of the ways we appear as subjects in our daily lives are structured by the performativity of data, they become readily citable in processes of surveillance as suspecting. In particular because surveillance agencies access data about our intimate lives—what we eat, who our friends are, where we pray—a suspecting act that cites these data carries the authority of authenticity. The constant availability, wide scope, and intimate quality of the data that can be cited in surveillance processes establishes the legitimacy and authority of judgments based on that data. That means that the authority of the subjectivizing force in data-based surveillance is parasitical to the many other forms of subjectivity performed by data.

Thus, Amoore’s analysis that the data-derivative “risk form acts through us and the *prosaic, intimate, banal traces of our lives*, but yet it forgets us” (Amoore 2011: 39, my emphasis) is very important. But we have to shift the (normative) emphasis from the “forgetting” of the subjects to the performativity that produces subjects. This performativity gains authority from the possibility to cite this very “prosaic, intimate, banal traces of our lives.” It engages the ways in which we become subjects for creating suspects. Importantly, Amoore’s inquiry already shows that the process of suspecting links data from many persons and many subjects (Amoore 2011: 29). Elsewhere I have discussed approaches which are meant to derive information about a particular person just based on data that others provide (Matzner 2014). The concept of citationality captures these observations: It foregrounds that a judgment about one person is always already a citation of other situations involving other persons. Data-based surveillance

with its sweeping collection processes appropriates all kinds of subjectivizing processes from many context for the purpose of suspecting. But this parasitical relation of data-based surveillance to our daily lives is not limited to engaging data that is provided by others while using data based services—what I called the third mode of collecting data in section 3.

John Cheney-Lippold presents an impressive example for these parasitic entanglements: He outlines a “soft” biopower “operating through conditional access to circuits of consumption and civility” (Cheney-Lippold 2011: 166). In particular, he analyzes the data processing technologies of a data broker that attempts to guess the gender of persons surfing the web and to provide differently designed websites, buying incentives, and products based on this assumption. This is not just marketing to a pre-existing target population. Gender studies have long established that the availability of commodities contributes to gender identity: starting from the pink and blue color codes of children’s toys, to the selling of dress (and undress) for grown-ups to the gendered offerings of all kinds of technology to ameliorate the “problems” of growing old. The decision, however, to whom these offers and incentives are presented, and consequently who is thus gendered, is now the result of data analysis and automation. Data based “access to circuits of consumption and civility” becomes part of gender identity.

This example shows the importance of taking into account the relation of commodities and surveillance—without conflating it to “surveillance as business model.” The important point is that particular modes of subjectivity presuppose the availability and consumption of certain commodities. This shows that Big Data analytics can have more subtle subjectivizing effects than direct data-based judgments. Rather than citing all kinds of data in a subjectivizing process, the analytics just control which elements are available to cite for particular subjects in particular contexts. In this case, this concerns consumer products the choice of which contributes to gender. But we can also think about the timelines or feeds in social media which are algorithmically filtered and which in turn provide resources for us to cite and link in creating the appearance of a subject. However, Cheney-Lippold’s example is illustrating because it concerns non-digital commodities. But both their availability and consumption is closely linked with data and its analysis. Those, in turn, can be cited in surveillance. Increasingly, the possibility to live certain subjectivities—also those not directly concerned with digital technology, is tied to becoming the subject of surveillance as well. Thus, quite material and somehow “old-fashioned” political problems of resources and commodification structure surveillance based on Big Data understood as production of subjects.

Subjectivity produced or co-produced by data appears in many cases that pertain to many aspects of our lives. Thus, my discussion of data as performative does not aim at an elementary critique of the subjectivizing force of data. Just as Butler does not see a way to evade the interpellation of subjects but argues for a performativity that yields different ways of subjectivity (Butler 1993: 223), I do not want to look at evading the performativity of data but to point out consequences of certain subjectivities that emerge from it, suspects in particular. And one of the important consequences of the production of suspect subjectivities is its parasitical relation to other, quotidian performatives of data.

This shows the extensive implications the performativity of data has for critical positions—in particular those that focus on the uniqueness of a subject that is too generalized in the data or dissected into data doubles. Butler discusses a similar problem concerning the attribution of social identity categories (by other persons, not using data):

One might be tempted to say that identity categories are insufficient because every subject position is the site of converging relations of power that are not univocal. But such a formulation underestimates the radical challenge to the subject that such converging relations imply. For there is no self-identical subject who houses or bears these relations, no site at which such relations converge. This converging and interarticulation is the contemporary fate of the subject. (Butler 1993: 229)

This means that the generalizing category of the suspect cannot be opposed by an originary subject. By the very way data-based surveillance subsumes every bit of data under the performative of suspicion, it eradicates or better appropriates the resources to appear as a single, particular subject that could contest this suspecting. Eventually all ways in which we can become a subject are also ways to become a suspect.

7. Conclusion

I have shown that a representationalist account of data does not suffice to address the challenges of Big Data. This is a problem that pertains particularly to theories of surveillance, running the risk of missing important and problematical implications of Big Data. The performative account of data I introduce is better suited to address these problems. From this perspective, surveillance based on Big Data is the production of bodies and lives as suspects. This is related to the plurality of performatives enabled by Big Data: being a subject is increasingly being a data subject in work and leisure, with family and friends. The performativity of data in surveillance is parasitical to such other performatives. Thus, no data remains innocent: All ways in which we can become a subject also are ways to become a suspect.

To maintain the critical stance in theorizing Big Data as surveillance, we have to question further how the “performatives of surveillance” relate to different identities and productions of subjectivity. How does the interplay of algorithmic features and social uses allow these performatives? How can they be structured differently? How does the production of suspects and deviants depend on the parallel production of a normal and behaving citizen? How do claims to veracity, but also space for openness and vagueness figure in these performatives?

A few hints at further research in this direction can be derived from the sources of this paper. Read from a performative perspective, Deleuzian accounts of surveillance can help in their analysis of surveillance as a “state form” creating boundaries and capturing flows (Haggerty and Ericson 2000: 608) opposed to different possible or imaginable configurations. Raley points at subversive and artistic approaches to the relation of surveillant and other performatives, which build on the premise that a “spectatorial position on the (putatively neutral) outside” is no option (Raley 2013: 137). In her opinion this means to play a part in the processes one tries to criticize, albeit a subversive or critical one. Judith Butler, whose theory of performativity I have used to flesh out Raley’s suggestion, also considers subversive subject positions as a promising way of critical engagement (Butler 1990: Ch. 3). The ensuing discussions in gender studies and related fields have shown that a little more caution might be apt (Brown 1993). Albeit, possible commonalities of subversive strategies in either area are a promising line of inquiry to pursue.

The many important studies on the epistemological status of Big Data can be extended with a performative account of data. These performatives of course depend on the epistemology attributed and realized by Big Data but cannot be fully grasped by this perspective. A performative account of the consequences that applications of Big Data have in real life complement this view for a suitable and critical account of surveillance.

Acknowledgments

An early version of this paper has been presented at the EASST conference in Toruń 2014. I would like to thank the participants for their comments and suggestions. I am particularly grateful to Eva von Redecker for her advice. Part of the research that led to this article has been funded by the German Federal Ministry of Education and Research.

References

- Adey, Peter. 2012. “Borders, identification and surveillance.” In *Routledge Handbook of Surveillance Studies*, edited by David Lyon, Kirstie Ball, and Kevin D. Haggerty, 193-201. London: Routledge.
- Amoore, Louise. 2011. “Data Derivatives: On the Emergence of a Security Risk Calculus for our Times.” *Theory, Culture & Society* 28(6): 24-43.

- Amoore, Louise. 2009. "Algorithmic War: Everyday Geographies of the War on Terror." *Antipode* 41(1): 49-69.
- Austin, John Langshaw. 1975. *How to do things with words*. Oxford: Oxford University Press.
- Bauman, Zygmunt, Didier Bigo, Paulo Esteves, Elspeth Guild, Vivienne Jabri, David Lyon, and R.B.J. Walker. 2014. "After Snowden: Rethinking the Impact of Surveillance." *International Political Sociology* 8(2): 121-144.
- boyd, Danah, and Kate Crawford. 2012. "Critical questions for big data." *Information, Communication and Society* 15(5): 662-679.
- Brown, Wendy. 1993. "Wounded attachments." *Political Theory* 21(3): 390-410.
- Butler, Judith. 1993. *Bodies that matter: on the discursive limits of "sex."* New York: Routledge.
- Butler, Judith. 1990. *Gender trouble: feminism and the subversion of identity*. New York: Routledge.
- Cheney-Lippold, John. 2011. "A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control." *Theory, Culture & Society* 28(6): 164-181.
- Deleuze, Gilles. 1992. "Postscript on the Societies of Control." *October* 59: 3-7.
- Degli Esposti, Sara. 2014. "When big data meets dataveillance: The hidden side of analytics." *Surveillance & Society* 12(2):209-225.
- Derrida, Jaques. 1977. "Signature, Event, Context." In *Limited Inc*, 1-25. Evanston: Northwestern University Press.
- Floridi, Luciano. 2012. "Big data and their epistemological challenge." *Philosophy & Technology* 25(4): 435-437.
- Foucault, Michel. 1975. *Discipline and Punish: the Birth of the Prison*. New York: Random House.
- Fuchs, Christian. 2011. "How Can Surveillance Be Defined?" *MATRIZes* 5 (1): 109-133.
- Gandy, Oscar H. 2012. "Statistical Surveillance: Remote Sensing in the Digital Age." In *Handbook of Surveillance Studies*, edited by Kirstie S. Ball, Kevin D. Haggerty and David Lyon, 125-132. New York: Routledge.
- Gitelman, Lisa (Ed.) 2013. *Raw Data is an Oxymoron*. Cambridge, MA: MIT Press.
- Haggerty, Kevin D., and Richard V. Ericson. 2000. "The surveillant assemblage." *The British Journal of Sociology* 51(4): 605-622.
- Introna, Lucas, and David Wood. 2004. "Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems." *Surveillance & Society* 2(2/3): 177-198.
- Kitchin, Rob. 2014. "Big Data, new epistemologies and paradigm shifts." *Big Data & Society* 1(1). Doi: 10.1177/2053951714528481.
- Leese, Matthias. 2016. "'Seeing Futures'—Politics of Visuality and Affect." In *Algorithmic Life: Calculative Devices in the Age of Big Data*, edited by Louise Amoore and Volha Piotukh, 148-164. Milton Park/New York: Routledge.
- Levinson-Waldman, Rachel. 2016. IBM's Terrorist-Hunting Software Raises Troubling Questions. Available Online: <https://www.justsecurity.org/29131/ibms-terrorist-hunting-software-raises-troubling-questions/> (accessed 26 February 2016).
- Lyon, David. 2014. "Surveillance, Snowden, and Big Data: Capacities, consequences, critique." *Big Data & Society* 1(2). Doi: 10.1177/2053951714541861.
- Manovich, Lev. 2012. "Trending: the promises and the challenges of big social data." In *Debates in the Digital Humanities*, edited by Matthew K. Gold, 460-475. Minneapolis, MN: The University of Minnesota Press.
- Massumi, Brian. 2007. "Potential politics and the primacy of pre-emption." *Theory & Event* 10 (2). Available online: <http://muse.jhu.edu/article/218091>.
- Matzner, Tobias. 2014. "Why Privacy is not Enough Privacy in the Context of 'Ubiquitous Computing' and 'Big Data.'" *Journal of Information, Communication & Ethics in Society* 12(2): 93-106.
- Raley, Rita. 2013. "Dataveillance and Counterveillance." In *Raw Data is an Oxymoron*, edited by Lisa Gitelman, 121-146. Cambridge, MA: MIT Press.
- Rashid, Fahmida Y. 2014. Surveillance is the Business Model of the Internet: Bruce Schneier. *Security Week*. Available online: <http://www.securityweek.com/surveillance-business-model-internet-bruce-schneier> (22 July 2015).
- Rouvroy, Antoinette. 2013. "The End(s) of Critique: Data-behaviourism vs. Due-process". In *Privacy, Due Process and the Computational Turn. The Philosophy of Law Meets the Philosophy of Technology*, edited by Mireille Hildebrandt and Ekaterina de Vries, 143-168. Milton Park/New York: Routledge.
- Thrift, Nigel. 2007. *Non-representational theory: Space, politics, affect*. New York: Routledge.