Big Data, New Connections

An Investigation into Big Data and the Production of Subjectivity

Master's Thesis
MSc in Business Administration and Philosophy
Copenhagen Business School, 15 May 2018

Supervisor: Christian Garmann Johnsen,
Department of Management, Politics and Philosophy

107 pages
207,010 characters

Mads Engmann Hildorf - 75137
Adam Khelifi – 71395

Abstract

Big data is becoming more and more popular. Companies use it to get new insights and improve their operations. Consumers engage with Big Data through the internet on social media. This thesis answers the question of how Big Data produces subjectivity and how Big Data in the recent data scandal involving Facebook and Cambridge Analytica produces subjectivity. This investigation is done through utilizing Deleuze and Guattari's concept of the machine as well as Lazzarato's notions on machinic enslavement and social subjection. Big data produces subjectivity through two machines: an expansion-machine and a data-trust-machine. The former produces subjectivity in how it enables the ones interacting with it to engage in expanding activities such as enlarging datasets, finding new datasets, finding new uses for Big Data, etc. The data-trust-machine produces subjectivity in how it enables actions of trust towards big datasets and Big Data solutions which can be hard to comprehend. The thesis also shows that the expansion-machine works in the domain of machinic enslavement where people find themselves in an automatic state of constant expansion. Furthermore, the data-trust-machine works in the domain of social subjection as it produces subjectivity by establishing the binary options of being someone who trusts Big Data or not. In the case of Facebook and Cambridge Analytica, this thesis explored how subjectivity was created through the expansion-machine in the domain of machinic enslavement and the data-trust-machine in the domain of social subjection. Prior to the scandal, the parties involved in the case (Facebook, Aleksandr Kogan, Facebook users, and Cambridge Analytica) all operated in machinic enslavement connected to the expansion-machine. Their subjectivity was produced as they engaged in expanding activities such as enlarging datasets, expanding personal online profiles, converting Facebook profiles into usable data, and using data to influence the US presidential election in 2016 through new Big Data practices. The thesis then examines what happened when consumer trust towards Big Data broke down, as the whole incident want public. This examination focused on Mark Zuckerberg's congressional hearing where he made a binary distinction of blaming him for the incident or not. This establishment will re-

connect the data-trust-machine as data will be overlooked in the event. This binary distinction set up by Zuckerberg produced subjectivity in the domain of social subjection by enabling the audience to make a choice between blaming him or blaming Big Data.

Table of contents

Chapter 1 - Introduction

Big data is here! Going through top themes of consulting management companies we are flooded with Big data articles (Boston Consulting Group, n.d.b), (McKinsey & Company, n.d.) (KPMG, n.d.). Online thought hubs such as TED Talk are sprouting with Big Data talks (TED, n.d.) and more and more universities are beginning to offer degrees in data science (Suffolk University, n.d.) (Allen, 2015). The world has come across a new crave, a new concept and a new era as Viktor Mayer-Schönberger and Kenneth Cukier address it (2017). The examples are plentiful underpinning this new thing's capabilities:

Boston Consulting Group, in collaboration with Amazon, estimated the benefits of using Big Data as means to lead a company. Through 167 companies, in five different industries, their survey proved that leaders making use of Big Data generated on average 12% more revenue. The survey also proved that Big Data leaders have faster growth rates (Boston Consulting Group, n.d.a).

In New York City the citizens were victim to a few hundred manholes blowing up as they caught on fire underneath. The iron covers would shoot several stories up in the air before falling to the ground. By using Big Data analytics, data scientists were able to distinguish the manholes likely to catch fire before it occurred, simply from the datasets on the manholes' age and the time they were last serviced. By renovating the manholes that the analyst had distinguished, the number of occurrences dropped to a bare minimum (Mayer-Schönberger & Cukier, 2017, p. 68-72).

Prematurely born babies stand a high chance of suffering an infection that may cause fatal consequences. Dr. Carolyn McGregor, health informatics researcher, had success with collecting 1,260 different data points per second per baby and using it for the better. She found that prior to an infection, the babies' vital health signs would actually stabilize. This shocked doctors as they had always seen such stabilization as a sign of betterment. This will help doctors to know when the infection takes place and ultimately save a number of premature babies' lives (Mayer-Schönberger & Cukier, 2017, p. 59-61).

Big data is saving companies from bankruptcy, stopping New York City from blowing up from underneath, and will be saving premature babies' lives. Big data is here to collect the data, analyze the content, and work out our problems. But what is this new phenomenon of
Big Data actually?

Big data is a matter of data, large amounts of data. So large amounts that usual data-analytic tools fall short to process it due to lack of processing power. This was the first definition of Big Data, attributed to Roger Magoulas (Halevi, 2012). And this data flows all around us from every engagement with smartphones, navigations systems, social networks, and the internet in general. Everything produces data and close to everything we interact with online, collects this data. The data becomes impenetrable to process by mere analytic tools, but only computing power tuned in with a set of algorithms can work out anything from the data and here the results surfaces. Just as with the preemies example where Big Data led to an insight that will help hundreds of prematurely born babies.

Now the attention towards this phenomenon's significance has long been underway, but in the light of the recent scandal about data privacy involving Facebook and Cambridge Analytica the Big Data attention has reached a new peak. This scandal prevailed as Facebook data had been collected by a third party, a company named Cambridge Analytica. They obtained 87 million users' personal Facebook data, and through it they sought to work out ways to influence American voters for the US presidential election of 2016. It is in this relation we will find our motivation as to wonder how Big Data and human subjectivity might be connected.

How then can we assess this matter? How are we able to talk of subjectivity in general? An approach to this account can be found through Deleuze and Guattari's terminology of machines, as they conform post-structuralism with a critical view on common-sense and objective opinions. Deleuze and Guattari take their perspective on the matter of subjectivity by analyzing the connections of people and objects. When Deleuze and Guattari approaches subjectivity, they are not interested in uncovering subjectivity. Rather, they aim at creating new concepts that can articulate the subject in new ways. This approach enables a new perspective on contemporary phenomena. By addressing the examples, actions, utterances and connections that Big Data feature we intend to create a new perspective into Big Data and we mean to render the Facebook crisis as to see what this might entail on the production of subjectivity.

This project will dive into the debated topic of Big Data involving data, data privacy, and doing business. This, with a goal of analyzing the influence Big Data might have on society and the individual citizen.

Problem statement

Big data spurs many questions of how it might affect the daily lives of societies, consumers, managers, analysts, employees, corporations, etc. Following Big Data's role in light of Mayer-Schönberger and Cukier's *Big Data*, other related literature, and finally the recent scandal with Facebook and Cambridge Analytica, we then want to state the two following research questions:

1) *How does Big Data produce subjectivity?*
2) *How does Big Data in the case of Facebook and Cambridge Analytica produce subjectivity?*

Answering the research questions

To answer our research question, we will delve into the book by Mayer-Schönberger and Cukier, *Big Data*. Here we are presented to Big Data in an accessible manner. Furthermore, *Big Data* is also very illustrative and uses many examples from real life cases of how Big Data was used to solve problems in different contexts. This feeds well into our approach of analyzing with Deleuze and

Guattari's, as their method of analysis consists of reading texts closely and surface their underlying logics.

Our reading of *Big Data* will be through the lenses of Deleuze and Guattari. Deleuze and Guattari's concept of the machine will take a central role in our analytical work as it is closely connected to the production of subjectivity. The machine, for Deleuze and Guattari is an agile concept of flows, connections, and actualization. Their example of the breastfeeding baby exemplifies the machines and their flows where the baby's mouth is a machine, and the mother's breast is a machine. The mouth and the breast allow a flow of milk. They connect and actualizes the crave for milk that resides with the baby. And so the two machines of mouth and breast are connected, creating a flow of milk which actualizing a craving (Deleuze & Guattari, 2009, p. 1-48).

Deleuze and Guattari's concept of the machine is closely related to the question of subjectivity. Subjectivity is created in through the actions the machine produces. Thus, in the example of the mother and child, the drinking and the breast machine created the baby as a milk-drinking subject. And so, machine give us an insight into subjectivity as it is the machines that convey the flow that actualizes our actions. And it is our actions that determine our subjectivity. And as the actions are something in motion, subjectivity isn't just stabile. Rather it is in constant movement, in constant production.

Using the concept of machines, we will read Mayer-Schönberger and Cukier's *Big Data* to find what logics are at play in their account of Big Data which can give us a better look into how Big Data can produce subjectivity. Thus, this project will examine and find two different machines at play in their book on Big Data: An expansion-machine and a data-trust-machine.

Following our machine analysis, we will draw upon Lazzarato's reading of Deleuze and Guattari and their concepts of social subjection and machinic enslavement. These two concepts here will act as an extension to the machine concept. Social subjection and machinic enslavement encapsulates two different ways of how the machines work, different in the way they function but powerful in their relation:

> "Social subjection equips us with a subjectivity, assigning us an identity, a sex, a body, a profession, a nationality, and so on. (…) but the production of the individuated subject is coupled with a completely different process and a completely different hold on subjectivity that proceeds through desubjectivation. Machinic enslavement dismantles the individuated subject, (…)." (Lazzarato, 2014, p. 12)

Social subjection works by placing subjects into a binary (for example man/woman, teacher/student or worker/capitalist), while machinic enslavement encapsulates the desubjectivating processes, which addresses the machinic relations we are connected to by default when we engage in activities automatically – when everything flows and we don't need to think about what to do and how to do it.

These machinic concepts together form the action of submitting subjects and objects to certain roles of identities through social subjection while machinic enslavement enfolds the automatic movements that the flow of subjectivity follows. The flow in the two machines together with movement from one to the other will lead to the production of subjectivity.

Finally, we will investigate the expansion-machine as machinic enslavement as well as see how the data-trust-machine relates to the domain of social subjection. These concepts together will then be utilized to analyze the recent event of the data privacy scandal with Facebook and Cambridge Analytica as involved actors and look into Mark Zuckerberg's congressional hearing in that regard. Our aim is to analyze the online platform of Facebook with the expansion-machine as machinic enslavement and the congressional hearing with the data-trust-machine as social subjection. This, to arrive at a movement of production of subjectivity conveyed by Big Data through the case of data privacy and its alleged violation.

Main points

Through our reading of Mayer-Schönberger and Cukier's book on Big Data, we find two machines at work: an expansion-machine and a data-trust-machine. The former produces actions of growth and expansion from enlarging data volume, use cases for Big Data, and the Big Data scientists' competencies. The latter machine, the data-trust-machine, produces actions of trust towards Big Data as the engagement and usage of it demand a data-trust to expand and thus enable the expansion-machine. These machines have an inherent way of working together, and so by employing social subjection and machinic enslavement, lifted from Lazzarato's reading of Deleuze and Guattari, we are able to put them in relation to each other. The expansion-machine works in the domain of machinic enslavement and so it arranges expansion as something done 'automatically' in relation to Big Data. The automation of the expansion-machine and machinic enslavement will break down as every other machine. Here the data-trust-machine is mobilized as to re-establish data-trust, and guide the production of subjectivity back into machinic enslavement. It is through these movements that subjectivity is produced.

Chapter 2 – Method

This chapter aims to justify, evaluate, and discuss some of the decision we've taken along the course to answer our research question. We will elaborate on the methodological choices taken to write our thesis and show how our research question, choice of theory, and the empirical correlate.

Our place in post-structuralism

Our thesis is concerned about how big data may create subjectivity. And for the matter of subjectivity and the self has a rich philosophical history which stretches far back to Descartes, Hume, Locke, Kierkegaard etc. (Atkins, 2005), all the way up to more contemporary philosophers like Žižek, Foucault, and Deleuze and Guattari. Few philosophers have avoided the question of subjectivity and the Self altogether but some have made it a quest to inquire into what it means to be more than others.

In more contemporary philosophy, especially when speaking of French contemporary philosophers (Deleuze, Foucault, etc.), a post-structuralism approach to the question of subjectivity has been prevalent. The post-structural when speaking of the philosophy of science is among many things occupied with challenging points of authority. The deviance from structuralism, as the name implies, consists of a denial of hierarchy and finite reference points in language (Alvesson & Sköldberg, 2009, p. 179). The post-structuralist philosophers see the relationship between the signified and the signifier as an endless recourse around in language with no rational beginning and no rational end. Thus, it does not make sense to claim that language and the creation of meaning has an ultimate reference point or origin from which perfect chains of signification can elude. Also, the post-structuralists argue that any attempt of signification of meaning is highly context dependent and that meaning only can be created and understood in their contemporary contexts in contrast to universal conceptions of phenomena. Likewise, the positions of power determining signification and representation is not seen as being the same in every situation. Rather, in post-structuralism, points of power are questioned and are usually perceived as residing inside a given context. Thus, there are no central and universal point of power or signification, they are always different and produce different representation in different situations. (ibid.)

When speaking of the self and subjectivity, it then doesn't conform to post-structuralist thought to have a fixed and universal idea about subjectivity. The person's choices, thinking, morals and so on has to be understood from the context that the subject finds itself in. And it is along these conceptions that we have our approach to subjectivity as well. In this thesis, we don't see the subject as a fixed being of static subjectivity. Rather, the subject is a product of its surroundings and has to be understood through the situations it finds itself in.

As mentioned, Deleuze and Guattari aren't the only contemporary philosophers interested in subjectivity:

> "Among contemporary critical theories (those of Badiou, cognitive capitalism, Judith Butler, Slavoj Žižek, Ranciére, etc.), it is largely a question of subjectivity, the subject, subjectivation, and the distribution of the sensible. But what they neglect is hoe capitalism specifically functions – that is, through "machinic enslavements."" (Lazzarato, 2014, p. 13)

These contemporary critical theorists do not take the sphere of machinic enslavement into consideration when answering the question of how the subject forms. Rather they focus solely on the sensible and the domain of what Lazzarato denounces social subjection. Thus, we have chosen Deleuze and Guattari as our theoretical framework as they enable us to analyze the production of subjectivity in both a sensible and a non-sensible, 'automatic' domain.

We have chosen Deleuze and Guattari as our main theoretical stance in this project of investigating subjectivity in relation to big data. Deleuze and Guattari, like mentioned above, argues that subjectivity is something ‘produced’ by the context they find themselves within.

Deleuze and Guattari’s approach to the subject

Subjectivity for Deleuze and Guattari is not something stable. Neither is it something inherent and constant within human beings. Rather, subjectivity is ever changing and consists in our way of interacting and navigating the world. Action and ‘movement’ is central to their view on the subject.

Like mentioned earlier, Deleuze and Guattari argue that the subject is a matter of ‘production’. What they mean by production is closely tied to their theory of the machines, which we will outline further in the next chapter. Shortly, machines are pre-subjective entities which enables various possible ways of subjectivity. The machines form in certain contexts which then enables subjectivity within that same field. Machines produce certain behavior, and people within the same field as the machines thus connect to them and then certain modes of being takes form. It is important to understand that subjectivity is not inherent in the body of the subject, but subjectivity is rather seen as the subjects’ motions, behavior, habits and so on. Thus, subjectivity is more a process than a trait.

For our research question, this means that in order to speak of subjectivity we have to not look at people’s inherent personal traits or anything of that matter. Rather we have to look more at people’s doing as well as the contextual logics they find themselves within. Furthermore, it is important that we don’t look at people solely but that we fix our gaze at the whole ‘machinery’ of both subjects, objects and everything in between. In line with Deleuze and Guattari, it would be wrong to do research with a subject/object duality in mind as in the machinic analysis, subject, object, signs, indications, milieus, buildings, logics all form and constitute themselves as machines interconnected. Thus, in our study, this means that we cannot give any preference to either subject nor object but most consider all notions of affect in how subjectivity might be produced by big data.

Choice of philosophical concepts

As mentioned above, one of our central concepts to analyze the field of big data and answer our research question is the machine. As machines are pre-subjective and the components which produce subjectivity, it is a focal point for our research. Through readings of texts on big data we have mapped machines at play which we argue may give us a picture of how big data can produce subjectivity. To nuance our investigation, we have chosen to also use two other concepts from Lazzarato's reading of Deleuze and Guattari. Those two concepts are social subjection and machinic enslavement which are two modes of machines and between the two we find the production of subjectivity. The two domains have a circular nature in that when the production of subjectivity in machinic enslavement breaks down, the production shift towards the domain of social subjection which aims at re-establishing the order of machinic enslavement.

Our use of these three concepts (machine, social subjection, and machinic enslavement) will be the basis of our philosophical inquiry. The machines will help us answer how subjectivity may be produced and social subjection and machinic enslavement will be used to give momentum to the process of production – by showing how machinic production of subjectivity moves from one domain to the other.

Deleuze and Guattari's concepts tend to work in networks of other interconnected concepts (Parr, 2010, p. 54-55). It can thus prove difficult to speak of A and simultaneously refrain from mentioning B, C, and D too. Our focus on the concepts of machine, social subjection, and machinic enslavement is not merely a choice of concepts but also a limitation from the notions they have in their company of other co-concepts. One example is that for Deleuze and Guattari, their concept of desire is central also to their concept of machines. It is difficult to speak of the two separately. We have for instance chosen not use the concept of desire in our thesis as well as many other concepts that relate to the three.

Our formulated understandings of Deleuze and Guattari's concepts is also written with that in mind that their philosophical terminology can be quite cryptic and not easily accessible. At times, authors have different understandings of the Deleuze and Guattari's *Anti-Oedipus* and emphasize some concepts as more significant than others (Buchanan, 2008, p. 51). So to best avoid illegitimate and counterfactual use of their concepts, we have formulated our own readings of the concepts in the section of theory and it is those understandings which we use in the thesis.

Position of the body

It is our perception that the machinic way of investigation gains its strength from the account of affects in a certain milieu. This leads us to the question of the position of the body and the closeness of ourselves in relation to our empirical data. In this thesis we position ourselves 'above' the empirical and thus falls under what Donna Haraway would call the God-trick (Haraway, 1991, p.189). She distinguishes between researching using the God-trick and researching through a subjugated position with interviews and direct observations. The prior is, like mentioned, one of a broad overview of different landscapes/empirical fields. Haraway criticizes such position in research, as it is usually practices relativism. Haraway's definition of relativism in this context is the actions of comparing different phenomena across empirical fields, and such comparison would not be methodological sound as it lacks contextual generated insight which can only be gathered through the subjugated position:

> "The standpoints of the subjugated are not 'innocent' positions. On the contrary, they are preferred because in principle they are least likely to allow denial of the critical and interpretative core of all knowledge. They are savvy to modes of denial through repression, forgetting, and disappearing acts - ways of being nowhere while claiming to see comprehensively... 'Subjugated' standpoints are preferred because they seem to promise more adequate, sustained, objective, transforming accounts of the world. But *how* to see from below is a

> problem requiring at least as much skill with bodies and language, with the mediations of vision, as the 'highest' techno-scientific visualizations.” (Haraway, 1991, p. 191):

Here, Haraway explains the subjugated position as one where you can better identify central knowledge of your field of investigation as well as a closer interpretation thereof – something which the investigator of the God-trick wouldn’t be able to see. This close analysis is a consequence of our choice of method, as we let our empirical data solely consist of text in books, articles, news outlets as well as dialogue recorded on video found online while leaving ourselves not directly taking part in the fields we are investigating. We never engage in the empirical fields we wish to analyze.

However, had we chosen to engage with our objects of investigation through direct interviews, field research, direct observation, etc. (which is ways of Haraway’s subjugated position) it could be argued that we ourselves would become part of the object of investigation that we wish to pinpoint machines within. Thus, we would become things affecting other parts of the milieu and ultimately influence the machinic production of subjectivity. This of course can be both an advantage and a disadvantage. On one side of the coin, from an anthropological point of view, this may limit our investigations as the distance might blur our view and blind us to certain points of interest that could have been valuable to our analysis. In the ivory tower approach, we chose our own empirical data and we control where the discussion and analysis is going in favor of our research question – the empirical doesn’t move us. However, this also has the disadvantage of making us overlook very central aspects of big data present in the vast literature on it if we chose to steer away from it, thus making our project more naïve to certain central aspects of big data and thus less credible as it may not have taken them into account.

Alternatively, the research question could also have been answered by making a field study of an organization where big data is prevalent and embedded in the practices and culture. This would, however, force our research question to center itself around that organization and thus become very specific to it. Big data practices may very well differ to some extent all dependent on which organization we're speaking of. In this thesis we wish not to make a particular study of how big data might produce subjectivity in a certain organization but rather maintain a broader scope. The broad scope makes us less able to say something very specific but on the flipside it enables us to make more generic analytical points which thus can be taken to further analysis and used as stepping stones to more specific organizational or governmental studies.

To answer our research question, we have chosen empirical data which conveys big data and its implications in a broad and accessible scope. This now leads us to our reflection on our choice of empirical data.

Choice of empirical data

Our primary source of data comes from M&C's book, *Big Data: The Essential Guide to Work, Life and Learning in the Age of Insight* (2017). This book examines the concept of big data and its implications through simple language and plenty of real life examples of big data's application. M&C's book is rather popular and has been cited plenty by researchers within the field of big data. Also, M&C's book's aim is not to unfold a specific area within big data but tries to grasp its discipline as a whole. Finally, due to the book's general descriptions and outline of big data, it also enables us to ask how big data produces subjectivity in a more general notion rather than specific to a certain individual, position, title, profession, consumer, etc. We thus deem it relevant and fitting for our thesis for the above reasons.

M&C's book also serves as our starting point. Thus, we only use additional texts on big data as an extension of our analysis of M&C's book. Aside from M&C's book, we also let Lazzarato's reading of Deleuze and Guattari direct us towards empirical data. This is more the case when we actualize the concepts of social subjection and machinic enslavement along with the concept of the machine.

Chapter 3 – Empirical Data

The following section will aim to guide us into the world of our empirical case of the big data phenomenon. We will briefly assess some of the contents that the project takes into account, get a preview of the material and give an introduction to the literature. This will help us frame our empirical field of study and simultaneously give a quick preview to what the content entails. And so we can initiate by asking what is big data?

What is big data?

> "Smart homes, the Internet of Things, social media, mobile applications, and other technologies are generating an unprecedented amount of multi-structured data. This "big data" has the potential to transform businesses and industries and to unlock tremendous value." (Boston Consulting Group, n.d.c)

As we go through our daily tasks we often find ourselves dealing with highly developed machines, from our smart phones, computers, our social media usage and online news information absorption. These processes not only take us through our tasks, our work responsibilities, or keep us connected to each other. All these devices are collecting data simultaneously. And we produce a lot of data every day, every one of us, and all the time: IBM claim "2.5 quintillion bytes of data [is] created every day." (Jacobson, 2013) 2,5 quintillion bytes by itself simply makes no sense, the number is not even understandable on its own, but anyone can agree that this number is quite big. Behind this number of data we find everything from Snapchat videos, traffic videos, messages, emails, GPS navigation, Facebook-"likes", and shopping lists. We amass a lot of data and this data is big data. These masses of data have previously been impenetrable by classic analytic method and so "Big data" justifies not only the mass of data, but also holds the introduction of new methods of analysis that grab hold of the volume as to give whole new insights, whole new answers, and new approaches to our research.

To give our investigation an entrance to the great phenomenon of big data and to bring the whole event down to a level we can grasp we went and found one of the leading guidebooks for big data. This led us to Viktor Mayer-Schönberger and Kenneth Cukier's book, *Big Data: The essential guide to work, life and learning in the age of insight* (2017). And so given our initial point of departure of big data empirical literature we will continue along the lines and widen the perspective as we move into some management readings of big data, and an analyst guide. Last but not least we will look into the data privacy scandal that elapsed the spring of 2018 with Facebook and Cambridge Analytica as central actors.

M&C and Big Data

As this project's main advancement to the big data phenomenon we find Viktor Mayer-Schönberger and Kenneth Cukier's (from now on referred to as M&C) book on big data. Both authors are respected in the field of big data and data development. Viktor Mayer-Schönberger is a professor by Oxford university in UK and has written one book regarding big data prior to this big data culmination named, "Delete: The Virtue of Forgetting in the Digital Age." Kenneth Cukier is the data editor of the economist. He has numerous writings on technology, business and society under his belt. Together they matched up to write out the big data that arrived on the bookshelves in 2013. Now the book counts for more than 1 million sold copies.

Big Data: The essential guide to work, life and learning in the age of insight starts by setting up the world as it is today. Then the book takes the reader to a multitude of examples giving on real world examples of big data coming handy as to estimate, decide, distinguish or in some other way answer to a certain problem. Throughout this project we will dive into a number of different examples and drive out the analytic elements of these examples and all along follow the line of big data that M&C is revealing to us. But then how do M&C look into big data?

By reading *Big Data: The essential guide to work, life and learning in the age of insight* we can easily assume that the *guide* is normative in the way that they are both front runner for the big data phenomenon, they both speak highly of its capabilities and the book carries a message of a new era of insights that is to be found in the book:

'Anderson deserves credit for raising the right question – and doing so, characteristically, before others. Big data may not spell the "end of theory," but it does fundamentally transform the way we make sense of the world.' (Mayer-Schönberger & Cukier, 2017, p. 72)

M&C estimate the matter of how much big data will transform the world and surely concludes, along with Anderson, that big data will change the way we make *sense* of the world. Concluding this, M&C is well aware it's a bold statement but nevertheless what they believe and so the book will continue based on this conclusion. Given our project follows a guide with this assumption we are able to assess the phenomenon formed by forerunners rather than critiques. This will direct our analysis to some extend until we leap out of M&C and so we will engage this challenge of direction by inviting other empirical content into the analysis as well.

Other related literature

This takes us on to the next part of our empirical input. As mentioned we will be looking into Granville's take on how to become a big data analyst. Becoming a data analyst will partake as one of the feats to M&C's description of the analyst. This role is quite unique for the big data field as not only are we looking into new methods of analysis but these new methods will demand new kinds of researchers as well named the data analyst: Granville's book will join forces with M&C in the example as to encapsulate this new researcher's role and what his job description will be.

Another aspect of the big data phenomenon is big data business management and strategizing with data in mind. Here the project will draw on leading consulting firms' articles. BCG, McKinsey, together with KPMG will help us ascertain this field and we will look for their remark on big data phenomenon in relation to previous business management. We've taken into the analysis their articles on *Straight talk about big data* from McKinsey along with *Three keys to building a data-driven strategy* also from McKinsey & company consulting corporation. Bearing in mind these consulting have a goal of their own when they engage with the field, presumable to attract customers looking for consulting regarding big data. This might entail a pro-bias attitude to the field as that will work to their advantage of attracting customers.

Facebook, Cambridge Analytica, and data

Lastly we will dive into a recent case that caught major media attention during the spring 2018. Facebook, as a social network platform, holds a significant amount of data from all their users taking part in their community. This data and this community altogether are on regular basis being used on advertising strategies and marketing and it's through this platform and advertising Facebook drive their revenue. But the episode that unfolded during spring 2018 regards a data scandal, as a third party company, Cambridge Analytica, got hold of a chunk of the Facebook data. This was obtained by an app, that when users, of the Facebook platform, interacted with it, it would absorb all their personal data on Facebook along with all their friends' data. This data was then used to develop big data analytics. The aim here for, Cambridge Analytica, was to help drive the Trump Campaign to a US presidential election victory. This event sparked a political and legal intervention, Facebook CEO, Mark Zuckerberg, was put before a congressional hearing conducted the 11th of April 2018 (Associated Press in Washington, 2018). By diving into this congressional hearing, but without addressing the question of wrong or right, we aim to investigate this event's partaking in the development of big data and what effect, if any, it had on the production of subjectivity. This case will drive our third and last part of the analysis and by enabling our analytic notions from the previous analysis we will complete the analysis with a view into the real world through the Facebook and Cambridge Analytica data privacy scandal.

Big data is quite the new buzz word in many scientific disciplines. By big data we are talking about mass amounts of data that are being collected from a variety of modern days' tool, from GPS navigation systems, social media, and telephone devices. This mass amounts of data have demanded new analytic tools in order to cope with the scale. For this project's entrance into the concept of big data we are engage with one book's perception of the uses, the insights and the opportunities this concept offer us.

Summing up, we will engage with M&C who are front runners and advocates for the big data tool in general. We will from there widen our perspective of big data by reading literature on how to become a data analyst and leading consulting companies' take on big data through big data management articles.

Last empirical content for our analysis will be the recent event that took place in spring 2018. This event being the data privacy scandal that involved Facebook, Cambridge Analytica, and the US congress as the data privacy scandal went before a congressional hearing in the April 2018.

Big data's terminology

The purpose of this section is to define big data as it will be analyzed and used in this thesis. As the thesis primarily focuses on M&C's conception of big data, it is important that we outline how big data is understood in exactly their book *Big Data*. A lot of the terminology surrounding big data will be used later on in the analytical parts of the thesis, and therefore this section will serve to define that jargon.

Big data and related disciplines

Today, the concept of big data is often used alongside another similar field which is artificial intelligence (AI). Sometimes, big data and AI are even used interchangeably. M&C say that the general perception is that big data is a component of the AI science, as AI also is based upon very large amounts of data. AI in itself, however, is computer programs programmed to simulate human intelligence and learning (also referred to as "machine learning") (Mayer-Schönberger & Cukier, 2017, p. 11). However, M&C argue that big data is not about simulating human intelligence, but rather,

> "At its core, big data is about predictions" (ibid.) and "... not about trying to "teach" a computer how to "think" like humans. Instead, it's about applying math to huge quantities of data in order to infer probabilities..." (ibid., p. 12)

We can further conceptualize AI as being computer programs developed to solve problems themselves, independently, whereas big data is about finding insights data sets and making predictions but still leave the decision making up to people, as...

> "The algorithm [i.e. big data] itself may not be making the decisions, but the machines [i.e. big data] are doing what machines do best, to help human caregivers do what they do best." (ibid., p. 60)

However, according to M&C, the concept of big data is rather fluid and there is no definite and universal definition of it. Yet they emphasize big data as a practice which has emerged from more 'simple' data science disciplines, so to say. The shift from previous data science practices to big data came to be when the amounts of data and the processing power needed to manage it surpassed both the storage and processing capacities of regular computers. Thus, big data cannot be managed using ordinary machinery and requires more advanced computers and storage capacity than what is seen as normal. However, M&C makes it clear that "... the real revolution is not in the machines that calculate data but in data itself and how we use it." (ibid., p. 7), thus when speaking of big data, the focus should be on the data and its potential uses.

Finally, big data is also seen as a practice which covers many aspects of research across academic disciplines, as the data scientist (the usual title of a person hired to analyze and work with big data) usually has some knowledge in many academic disciplines. Later in the analytical sections of the thesis, we will speak more about the role of the data scientist drawing upon Granville and his book, "Developing Analytic Talent: Becoming a data scientist." In it he argues that the data scientist should have knowledge about statistics, mathematics, programming, and even an MBA would prepare him or her nicely as a data scientist (Granville, 2014, p. 45 and 73) M&C also exemplifies the data scientist as having skills with both statistics, software programming, storytelling, and infographics design (Mayer-Schönberger & Cukier, 2017, p. 125).

Leaving the general aspects of big data, we will now proceed to define the various relevant components of big data which are necessary to understand prior to reading the rest of the thesis. Frist we describe how they understand the algorithm.

The algorithm

The algorithm has a central part in big data. If we look to the Oxford Dictionary, the definition of an algorithm goes as follows: "A process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer" (Oxford Dictionary, n.d.a). M&C define the algorithm similarly in relation to big data as "… based on mathematical models that crunch mountains of data to predict gains…" (Mayer-Schönberger & Cukier, 2017, p. 8)

Drawing upon the two definitions above, the algorithm is a mathematical function by which big data's data streams are processed to form insights. The algorithm is instilled into the processing computer of the large data sets most likely by the data scientist. The algorithm acts like a set of rules from which it finds valuable information in the data. There can also be good and bad algorithms according to how well they find valuable insights in different data sets. They can be simple and they can be complex. Though, in the end, it is how well they manage to find the valuable information that matters.

However, the algorithm is not the sole reason of big data's efficiency of inefficiency. In one example, M&C shows us that sometimes a complex algorithm may perform worse than a simple algorithm which on the other hand has a lot more data to work with. In the example, two data scientists working on Microsoft Word's grammar check function discovered that their conventional and complex algorithm performed worsen than a much simpler algorithm in spell checking words. However, the simple algorithm had a billion words (which acted as the data in this case) to work with while the complex algorithms usually had way less words around a million or less.

As mentioned, the algorithm can't find any insights independently. It needs data to be accompanied by data before it can be productive. We'll now take a short look at what kind of conception of data we use in this thesis.

What is data?

Data plays a central role in the project. The Oxford English dictionary refer to data as: "Facts and statistics collected together for reference or analysis." (Oxford Dictionary, n.d.b) Apart from that, we understand data as points of information which can give no insight prior to analysis. This is the reference point on data we are working with in this project. Data will come to partake different actions throughout our analysis, but will always only refer to, facts or quantitative matter collected.

Chapter 4 – Theoretical concepts

Machines

In collaboration, Gilles Deleuze and Felix Guattari created different concepts that speak of the subject and its surroundings through interrelations and connectedness. One of the concepts they depend on the most is *machines*. We find our reading of Deleuze and Guattari's machine through a reading of *Anti-Oedipus*. Mostly we are seeking knowledge in second-hand literature that attempts to formulate the concept of the machine in different ways making it more accessible. This second-hand literature is found in Maurizio Lazaratto's book of *Signs and Machines* (2014), Ian Buchanan's *Deleuze and Guattari's Anti-Oedipus: A reader's guide* (2008) Claire Colebrook's *Gilles Deleuze* (2002), and Adrian Parr's *The Deleuze Dictionary* (2010).

The machine is all around

Machines are all around us, already connected, already engaged with other machines and already active. Machines are working on all levels, from individuals to society to state to the social to relations to the cup we pick up and drink from:

> "Everything is a machine. Celestial machines, the stars or rainbows in the sky, alpine machines – all of them connected to those of his body. The continual whirr of machines. "He [Lenz] thought that it must be a feeling or endless bliss to be in contact with the profound life of every form, to have soul for rocks, metals, water, and plants, to take into himself, as in a dream, every element of nature, like flowers that breathe with the waxing and waning of the moon." To be a chlorophyll- or a photosynthesis-machine, or at least slip his body into such a machine as one part among the others." (Deleuze & Guattari, 2009, p.2)

Here Lenz takes a stroll outside in the mountains and experiences much intensity. Intensity with everything around him, everything he connects with as he is "in contact with the profound life of every form" (ibid, p.2) as he is in contact with every machine around him. Lenz' experience reveals every potentiality around him: he even experience "flowers that breathe the waxing and waning of the moon." (ibid, p.2) If anyone else than Lenz walks through the mountains they would have felt the rocks and probably noticed the flowers but the flowers would not appear to have a form– a potentiality – to the strolling subject. The flowers would have swayed with the wind, maybe, but would not likely appear to "breathe with the waxing and waning of the moon". This, contact with the profound life of every form is the surfacing of the machinery in the entities. Lenz is not just strolling his way through the world, but rather he enables it. He *moves* it and *experiences* it and its machines. Lenz is just like us but with a more engaged psyche. We need to understand big data more like Lenz, to move it, to get in contact with its machines.

Before becoming a "Lenz" and notice the flowers' breath and waning of the moon, we stop stepping out over the edge and down in the abyss. Deleuze and Guattari's machine allows the reader to stare into the abyss and to understand it from a distance as we distinguish happenings and flow. This is the mission with machines. To understand potentialities that work around us everywhere enabling us, connecting us, and disconnecting us.

The machinic motion

Machines are Deleuze and Guattari's project to describe how the world moves, and how the subject engages with everything in the world. The machines function via flows and cycles of flow. These flows work by means of events such as actions, speech or mere thoughts. These events are actualizations of the flow and a flow is only possible between two interconnected machines. This is a central regard for the machinic nature. The machines will enable a flow, put the flow in motion through events and disconnecting the flow when the event fails to continue and break down. This flow, distinguished by the events, is the gateway to understand the machines that are connected. By addressing happenings, we are addressing the flow, which will emerge the machines that are connected. The machines are only distinguishable by means of the flow they enable. Claire Colebrook use the following example in explaining machines and flows,

> "A *machine*, however, is nothing more than its connections; it is not made by anything, is not for anything and has no closed identity. (...) Think of a bicycle, which obviously has no 'end' or intention. It only works when it is connected with another 'machine' such as the human body; and the production of these two machines can only be achieved through connection." (Colebrook, 2002, p. 56, italics in original)

It is only when the human body and the bicycle connect they achieve a flow of cycling. The bicycle alone doesn't make any 'intention', no cycling motion. Same goes for the human body, it cannot cycle without the bicycle. The connection makes the machine apparent as the event, here cycling, takes place and the machines enable flows as to how the events unfold: The bicycle as a machine and the human body as a machine, and the act of cycling as a flow. The flow of cycling is enabled as the connection of the machines is realized, once the flow stops, the two machines disconnect, and the machines breaks down.

This connotes the conception of machine and gives it its widespread nature. Machines are within the subject, outside it, without the subject, and part of something bigger as well as something in itself. Machines are in everything and everything is a machine. But, very important too, a machine is also nothing more than a connection. About this Claire Colebroke writes,

> "... a machine has no subjectivity or organizing center it *is* nothing more than the connections and productions it makes; it is what it does. It therefore has no home or ground; it is a constant process of deterritorialisation, or becoming other than itself." (Colebrook, 2002, p. 55-56)

Machines have an inherent nature of connecting and disconnecting, to engage with other machines. Without these flows the machines are nothing. This gives the machine a certain kind of analytic momentum, we either have to find a machine in its motion or there's no machine altogether. The project is thus forced to prove the happenings and the flow of the machine in order to create and work with the machine.

Machinic assemblages

As machines work all around us at all times, we have to work with a certain notion of the machines collectively as well. Two machines rarely perform alone, but are part of a bigger network of machines, a part of the assemblage, according to Deleuze and Guattari.

> "The assemblage to Deleuze and Guattari refer to a complex collective constellation of object, subjects, expressions, qualities, etc. These has by some mean a connection to one another and together 'an assemblage emerges when a function emerges; ideally it is innovative and productive. The result of a productive assemblage is a new means of expression, a new territorial/spatial organization, a new institution, a new behavior, or a new realization." (Parr, 2005, p. 18.)

With a car an experienced drive does not need to actively think to drive, if he or she is just engaging with the machinic assemblage of the car without having to make any effort and so:

> "Our actions and subjective components (memory, attention, perception, etc.) are "automized," a part of the machinic, hydraulic, electronic, etc., apparatuses, constituting, like mechanical (non-human) components, parts of the assemblage." (Lazzarato, 2014, p. 89)

We will use the machinic assemblage to distinguish some of the automation processes that subjects sometimes engage with, like in the car example. And this automatic connection to machine assemblages guides us through a significant part of our daily infrastructure. We engage in driving, shopping, cooking assemblages, etc. through this automatic approach subjectivity is produced as well.

The machine summarized

Machines partake as the essential infrastructure of our daily lives. Machines are essential to understanding thought and connections to everything around us, and are thus a determining factor behind the subject and especially in the production of subjectivity.

Machines and machinic assemblages will come to play a central initial role for this project. This, for a good reason, seeing that machines are all around us, in everything, and everything is potentially a machine. In the project we need to take a stroll, as Lenz in the mountains, through the literature of big data to distinguish and engage with machines it entails. The machine to Deleuze and Guattari and Lazzarato, and to this investigation, is examined and created by means of flows, of connectedness and disconnectedness.

The machines work by establishing flows in between them and they allow a flow to appear through events. Events can be seen in examples as acts, speech, thought, etc. And so the machine can emerges and is created in the events that form a flow and from there we can grip hold of the machines. The machine is nothing without its connection to another machine, and the connection forms a flow, which can be distinguished by events.

Further, machinic connections often play a part in a bigger network of machinic connections. This leads us to the final term within the realm of machines, assemblage. Assemblages are given by a complex constellation of machines and form a united body of production.

A final theoretical note has to be stated here. It has to be noted how the Deleuze and Guattari terminology of machines and machinic assemblages has a variety of concepts closely connected to the machine itself. Close to the machine concept we have Deleuze and Guattari's notion of desire. Desire enables the flow, searches flows and thus plays a big part to Deleuze and Guattari's notion of machines altogether. In this project we will overlook this complex notion of desire – as many insights it might have opened up for, as many potential theoretical pitfalls it invites along with a myriad of related concepts. Along with Deleuze and Guattari's concept of desire we've left out the three stages of synthesizes which places the machines in a map of production but again alter the scope of the research. The list of related concepts continues: rhizome, body without organs, and reterritorialization/deterritorialization. By having this stated, we acknowledge the multitude of the machine term in Deleuze and Guattari's terminology but at the same instance we have to prone this multiverse perspective, as the sheer complexity would far surpass that of our needs and alter the investigation altogether. As we investigate how the big data have an influence we obtain a more compliant result by decreasing the magnitude of theoretical approaches. Now, this aspect of theoretical complexity invites another implication to the project. Such complexity would lead to a problematic extension of the researcher's position, not only to the empirical case, but the theory as well. The theory surrounding Deleuze and Guattari's notion of machine is as Ian Buchanan puts it, "(...) both more complex and dare I say less irrational than Anderson allows." (Buchanan, 2008, p. 1) Perry Anderson, in the example here, is not alone; Deleuze and Guattari's terminology in *Anti-Oedipus* is very complex and at some moments irrational. This invites theoretical pitfalls and critiques that could cause a debate of creditability. By holding on to a reading of the machine and machine assemblages, isolated, we obtain a greater theoretical creditability and allow the investigation to pinpoint even small but still significant machinic aspects. This is why, the only addition we will have to our machine concept will be that of social subjection and machinic enslavement.

Lazzarato's conceptions of social subjection and machinic enslavement

Lazzarato's book, "Signs and Machines: Capitalism and the Production of Subjectivity" (2014), is a philosophical investigation of capitalism as a common denominator in today's society.

Though the word ‘subjectivity’ is usually used specifically for people’s being and their human traits, Lazzarato makes subjectivity refer to a broader notion which encompasses people as well as signs, objects, social relations and so on. This is mainly due to the philosophical framework which Lazzarato himself draws upon when constituting his analysis of capitalism. Lazzarato utilizes Deleuze and Guattari’s understanding of subjectivity as his theoretical compass. It is this perspective that enables the study of subjectivity not only one of human beings. As Deleuze and Guattari argue, such project must be undertaken by not singling out the human being as the main subjectivity – rather both signs, language, logics, objects, relations and so on has a crucial role in the matter. (Lazzarato, 2014, p. 26) Two of Deleuze and Guattari’s concepts which constitute the cornerstones of Lazzarato’s study, are the notion of the machine (which we have covered in a previous section) and the notion of signs. Through the next section we will elaborate on the social subjection and machinic enslavement. As a subcategory for both concepts, social subjection and machinic enslavement, we will write about *signifying* and *asignifying semiotics* as these will play a vital role for our investigation and our way of analyzing the social subjection and machinic enslavement

The performance of production of the subjectivity resides in the collaboration of the two domains of machinic enslavemenet and social subjection. Here social subjection equips us with identity, through a social model that subjectivities must conform to such as Man/woman, teacher/student, etc. But, social subjection, is coupled with machinic enslavement which is completely different as it dismantles the individuated subject we see in social subjection, but instead describe the spontaneous reaction and action in the world, the individual automation process with the world. Through these two domains we can understand the production of subjectivity. The real production happens between the two, as subjectivity moves from one to the other and with every movement it develops.

To put this in other words, we can use lazzarato’s example of driving a car. As we are heading down the highway driving the car, we are simply going along with the road, following the rules, the speed limit, and the lines on the road. Here we are engaged with a certain spontaneous relation to the world, as we are doing this by default, by ease and almost automatically. We've checked the fuel level and know the speed we are driving but we don't have to look at the gauges all the time. Here we are just being guided by our machinic assemblage which is the gauges we are keeping track off, the lines on the road that we follow and our speedometer that is within limit. We are doing this by ourselves, spontaneously. (Lazzarato, M. 2014, p. 89) This is the machinic enslavement mode that keeps us going, that guides on through the majority of our subjectivity production (driving the car).

Now, when suddenly we notice a car which is broken down in the distance we move out of our spontaneous world relation, we have to question the situation: "do I shift lane? decrease my speed and continue? or do I stop and help out? do I flash my hazard warning lights for the cars behind to notice the happening?" These questions take place within the order of social subjection. Now we must ask questions, now we break the automation that governs us to question the situation, we make a choice and follow through. We make a lane change, decrease the speed and pass the car situation, to increase the speed again as we've passed and find the inner lane once more to re-engage our machinic mode of automation once again and continue driving.

Another outcome could be that we brake, put hazard warning lights on, find the spare wheel and engage with the driver of the car that has broken down. These two different approaches lead to two very different outcomes: one takes you back to the same state of machinic enslavement, of driving down the highway, right away. The other leads to engagement with a problem, engagement with another person, and may lead to you giving away a spare wheel to another person.

Nevertheless, the approach you pick in the social subjection instance (when seeing the broken car ahead) might come to define your future machinic enslavement altogether – you are the kind of person who stops when something is out of order on the road, or the guy who doesn't stop. Social subjection and machinic enslavement in the car example both produce subjectivity.

Machinic enslavement

Machinic enslavement makes itself out through the flow between machines. By these flows of the machines we can see how non-human objects with human subjects engage to create subjectivity:

> "The [in]dividual is contiguous with machines. Together they constitute a "humans-machines" apparatus in which humans and machines are but recurrent and interchangeable parts of production, communications, consumptions, etc., process well exceeds them." (Lazzarato, 2014, p. 26)

For the machinic enslavement both humans and machines make up the production of subjectivity. Everything is components, cogs and gears in a assemblage to make the process go. In the car, the driver couples up with all the machines around him or her as to actually drive: the gauges guide, the lines lead the way, and the machine of the engine compiles to react on throttling and brakes on braking. We produce mere inputs and outputs in the whole system of a machinic assemblage to produce processes, to produce subjectivity. Now seeing that the machinic enslavement works through a variety of connected machines, the vital notion here is the underlying machinic notion that gives the flow. This enables the components to archive the collective process that exceeds them all. Here we must look beyond the machines themselves by looking into them and their function as to distinguish a flow that guides them. This is obtainable by looking into a technical grasp known as asignifying semiotics for the machinic enslavement.

The asignifying semiotics

To distinguish the machinic enslavement we have to understand and distinguish the asignifying semiotics. Now, to put the term in different words, asignifying means that which is not directly evident, marked, or signaled. Semiotics relate to the discipline of studying signs and symbols and their meaning altogether. For Deleuze, Guattari, and Lazzarato this is not the immediate function of asignifying semiotics, though it does carry the significance of the words with it, but we find asignifying semiotics further conceptualized with Lazzarato's reading of Guattari and Deleuze.

These entities are the gauges in the car, the speedometer, the fuel gauge, and the navigation system. They make up as an individual cog or gear for themselves in order to make up a movement and enable a flow that create subjectivity. This is, for Lazzarato, why they "(...) are not beholden to significations and the individuated subjects who convey them. They slip past rather than produce significations or representations." (Lazzarato, 2014, p. 80) They are not presenting something significant by themselves, but rather they *slip past* the significance they are made from. The asignifying semiotics are the gauges in the car, but in society they function numerous places, such as stock listings, currency, national budgets, fashion, matemathics, and music. (Lazzarato, M. 2014, p. 87) It governs us by default, we engage with the machines by default, they interact by default and through all of these subjectivity is produced.

But among some of them we can find a different flow mechanism. By establishing a network of machines we obtain an asignifying grasp. We need to look into the connections, semiotics, symbols or events, not to see what they are,

> "(...) but instead bring them into "communication in order to show how man *is a component part* of the machine, or combines with something else to constitute a machine. The other thing can be a tool, or even an animal, or other men."" (Lazzarato, 2014, p. 81)

And so by looking into mere communications, motions, and flows we can see what machine it constitutes and thus find the motion that guide the machinic enslavement, that guides the car, that drives production of subjectivity.

Machinic enslavement makes up the majority of the production of subjectivity but it does not account for it all because, as written, when the event changes, a broken car appears on the road ahead, something occurs that the default or automatic setting does not comply with. At this moment the machinic enslavement breaks down. At that instance we, as a driver, have to look up and thus break the machinic enslavement's flow. Here we enter into the domain of social subjection.

Social Subjection

Social subjection is the second realm to the two folded reality of production of subjectivity. Now where machinic enslavement breaks through events and signification, social subjection is the function for which events and significations come to play the vital role. In social subjection things has to fit into social models such as man/woman, capitalist/worker, teacher/student. These are found in clear utterances. It is in the very moment someone asks, "what's your profession?" that we place ourselves in a mode of subjection when we answer, "I'm a teacher". This is the event that makes the identity for a subjectivity take form, in this moment a stabile concept of you as a person has be conceptualized, the social obligation has been acknowledged: now you are a teacher. Social subjection always puts up a binary reality as a social obligation,

> "(...) it's an either/or: either someone, whoever it is, comes up with new methods for production of subjectivity, whether Bolshevik, Maoist, or whatever; or the crisis will just keep on getting worse." (Lazzarato, 2014, p. 15)

For the car example you are either the person to move along while something occurs or you are the one to stop and address the happening. Either way, the utterance, or the action in that moment decides the either/or outcome to the episode. This episode requires an action. It demands a taken position to the episode or "the crisis will just keep getting worse." (Lazzarato, M. 2014, p. 15) This process of social subjection becomes apparent in signifying semiotics.

Signifying semiotics

Along the previous lines of asignifying semiotics translated in machinic enslavement, here we find signifying semiotics translate to direct symbols and signs. Again, this does not grasp the whole concept on its own, but do tap into the function:

> "Only the speech act and not linguistic propositions has the property of accomplishing an enunciation, for it is through address and response that *it express values, points of view, emotions, affects, sympathies, and antipathies with regard to the situation, to the other, to utterances and, in particular, to utterances referring to the "true, the just, and the beautiful.""* (Lazzarato, 2014, p. 180, italics in original)

It is not the proposition that holds the power to enunciate, or articulate, but only through the speech act will we see the expressing of anything that matters. The utterance works as a sign here and establishes a social obligation. Here speech is the act in motion but speech could have been substituted with a physical act or the like. For with signifying semiotics we are looking for actual utterances, actions or remarks that appear as to express a certain value, a certain positioning in relation to a matter. This can be that of the driver, that either drives on or stops his car to help.

In order to engage with signifying semiotics in the analysis, first we must engage with the actual representation and look for what the act demands, performs or creates. So, by demonstrating the happening, the act, and exploit the situation we can achieve a sense of the social subjection in motion through the signifying semiotics. We have to distinguish the event as it represents itself. From here we can develop the social subjection.

Machinic enslavement and social subjection summarized

As part of the investigation of the production of subjectivity this project makes use of the two concepts of machinic enslavement and social subjection. By reading Lazzarato's book *Signs and Machines* we find the concepts described as they are lifted from Deleuze and Guatarri. The performance of the two concepts really shows in between the two and the subjectivity moves between them. Both concepts rely on Deleuze and Guattari's machine concept as they both function by connection, disconnection and flows.

Machinic enslavement plays a significant role in the production of subjectivity as it's the spontaneous realm of production. It is what subjectivity is guided by as it makes relations with a myriad of machines every day. It is like the gauges in a car, that show the speed and the lines on the road that we follow and do not question as we go. Machinic enslavement scopes in on all the cogs and gears that takes part in the process. The approach to distinguish and analyze machinic enslavement is to be found through what Lazzarato calls the asignifying semiotics.

Asignifying semiotics regard what is not present, rather that which slips past event. But by making out a number of events having the same machine we are able to determine what *slips past*, what is taken for granted and partake in the production of subjectivity.

Machinic enslavement maintains a flow up until it breaks down. It breaks down once the gauges stop working or disappear, when we ask questions to their function or when something unexpected happens and demands a reaction.

Social subjection works in binaries, and is distinguished by signifying semiotics. Signifying semiotics, opposite of asignifying semiotics, appear only through appearance, through the event, it becomes a signifying semiotic. It is only a semiotic (sign or symbol) once it is out there and represented. The semiotic being an act of speech or a physical act has to be demanding or acting according to a binary world view, the either/or perspective. And so the social subjection will function to demand or act according to the binary it was taken into account. By this the social subjection will re-establish a new machinic enslavement.

Production of subjectivity

While asking ourselves how might big data come to produce subjectivity we must first address another question: "(…) [W]hat does the concept of the production of subjectivity entail?” (Lazzarato, 2014, p. 12) This is the question that we have to ask prior to our own investigating of subjectivity. Luckily, for us, Maurizio Lazzarato already asked himself that question through his reading of Deleuze and Guattari. Lazzarato’s book *Signs and Machines* therefore gave the springboard for this thesis’ approach to that particular question.

How then does Lazzarato describe the production of subjectivity? For Deleuze and Guattari, read by Lazzarato, the production of subjectivity comes not as a single organism but is being produces collectively and in relation to the environment, or assemblage of machines, surrounding the subject and so,

> “[a]s a consequence, systemic crisis and the crisis in the production of subjectivity are strictly interlinked. It is impossible to separate economic, political, and social processes from the processes of subjectivities occurring with them.” (Lazzarato, 2014, p. 8)

This interlink between the production of subjectivity and the social, economic, and political systemic (crises) works in two ways: through social subjection and machinic enslavement. The first, social subjection works by assigning identities in binaries through representations. This, social subjection, is the production of the individuated subject, and is coupled with the other process, known as machinic enslavement. Machinic enslave is related to the automation processes and dismantles the binary identities that social subjection establishes.

So, in order for our project to investigate the production of subjectivity we will have to look into the social, economic and political relations that the subject relates to.

The two machinic concepts do not only form two ways of understanding the production of subjectivity but they play a fundamental role, for they enable the adjustment, modification, solicitation, assemblage, and stabilization of subjectivity (Lazzarato, 2014, p. 66). This is obtained through joining the “(…) two heterogeneous dimensions of subjectivity – the molar and molecular, individual and pre-individual, representational and pre-representational (or post-representational).” (Lazzarato, 2014, p. 31)

Through social subjection and machinic enslavement this project will seek to investigate the production of subjectivity. The production of subjectivity happens in the way machinic enslave and social subjection combine and work together. When machinic enslavement breaks down social subjection mobilizes and re-establish machinic enslavement.

Chapter 5 – Finding Machines

In this chapter we will delve into M&C's book *Big Data*. By engaging Deleuze and Guattari we will identify the underlying flows that reside in M&C's examples of Big Data and how these flows play out. By adhering to Deleuze and Guattari's concept of the machine, we aim to play out the examples and bring forth the flows at play in the book and show how machinic connections are established under Big Data.

In the following, we grasp the productions of two machines which operate in logics of expansion and data-trust. These two machines will be used for further analysis further on in the thesis. We will begin with finding the expansion-machine.

The expansion-machine

The example of Matthew Fontaine Maury

In this section we will outline a story found in M&C's book. However, this story will not be the only one we analyze. We will use a few more examples from M&C's book, but the example of Matthew Maury will play the largest part and we will analyze it in multiple sections in this chapter.

Matthew Maury was a U.S navy officer. Unfortunately, Maury got injured at an expedition and the following surgical procedures made him crippled and thus not able to fare the seas again. As a consequence of Maury's handicap, he was appointed head of Depot of Charts and Instruments. At his new desk he began his endeavor of collecting old maps and fare roads at sea. This interest was sparked by a long time wondering of why

> "... ships would zigzag across the water rather than take more direct routes. When he quizzed captains about it, they replied it was far better to steer a familiar course then to risk a less known one that might entail hidden dangers. They viewed the ocean as an unpredictable realm, where sailors faced the unexpected with every

> wind and wave. Yet from his voyages Maury knew that this wasn't entirely true. He saw patterns everywhere. On an extended stop in Valparaiso, Chile, he witnessed the winds operating like clockwork." (Mayer-Schönberger & Cukier, 2017, p. 73)

Maury thus began investigating the matter of sailing routes. He quickly found that the traditional maps were very inaccurate and hadn't been updated for a long time, and he was on the brink of coming up with something newer and better.

Maury gathered all old diaries written at sea. In these diaries details of journeys were noted down. However, most of those logbooks were merely written to pass the time during the lonesome days at sea. But all the information they contained proved very useful for Maury's project. Alongside the old books, which was regarded useless by most in the navy, Maury also "… inventoried the barometers, compasses, sextants, and chronometers in the depot's collection." (ibid., p. 74). Maury also spoke to experts in seafaring to learn from their insights about the different areas of faring the sea (ibid.). Maury also established other flow of information as he spoke with sailors about certain areas and sail routes.

What Maury managed to do was to gather all this information that was deemed useless regarding the sea and seafaring, which he transformed into something of great use to the navy. Maury's final product was in the form of new sea maps, which proved to be much better than what was used before.

But his maps were not done yet. Maury continued to improve the maps' accuracy by continuously adding more and more information. Maury "(…) created a standard form for logging ships' data and got all U.S. navy vessels to use and submit it upon landing." (ibid., p. 75). Soon, merchant ships from outside the navy also wanted Maury's maps. Maury would gladly give them to them in return for them also logging their trips in the same manner as the U.S navy vessels, making his maps even better.

Maury's maps became more and more popular and they proved themselves by showing more efficient routes (ibid.). Maury, now, had many flows of information. They came from among merchant ships and U.S. navy vessels. He even received useful information from…

> "… old sea dogs who rejected the newfangled charts and relied on the traditional ways or their intuition served a useful function: if their journeys took longer or met with disaster, they proved the utility of Maury's system." (ibid.).

Maury's charts ended up, not only helping sea farers, but they also played an important role in "… laying the first transatlantic telegraph cable." (ibid., p. 76). Maury's method also proved useful at the time when Neptune was discovered as a planet and not, as previously assumed, a star. Maury thus collected the old star maps which were made on the presumption of Neptune as a star to map its orbit (ibid.). Thus, Maury's data and method of converting new things into data and using them for various things grew and grew and he helped accomplish many things with Big Data.

We have now elaborated on the story of Maury and his sea charts and will now move onto another theme in M&C's book. This theme is about how Big Data's value is equivalent to the ways in which it can be used.

Use cases

In this section we will through our reading of M&C's book argue how Big Data's value is equal to its use cases. From the story of Maury, we will argue that when Big Data can be used for more and more things (making better routes for marine ships and merchant ships, laying the telegraph cable, mapping Neptune's bane), Big Data at the same time becomes more and more valuable.

We begin by looking at how Maury's charts only deemed valuable as the navy, the merchant ships, and the continents of American and Europe (laying the telegraph line between them) could use them to do tasks efficiently. This becomes apparent in how the old logbooks from the old sea fares were "… regarded as rubbish." (ibid., p. 74) as they consisted of "… odd limerick or sketch in the margins … [and] sometimes seemed more of an escape from the boredom of the passage than a record of ships' whereabouts." (ibid.). Strange plays with words ("odd limerick") and different drawings seemed to most navy personnel as things produced to merely entertain the sea farers themselves on the lonely sea and not as recorded experience or insight into the sea farers' travels. The old logbooks didn't seem to have any relevance in assisting the navy's personnel in completing their tasks of getting safer and faster from point A to point B at sea. Thus, they had no use and by that they were regarded as of no value.

The term *value* is in this example pieces of information that can assist in doing tasks better or more efficiently. The old logbooks didn't have any perceived value but they were regarded as merely an item to drift time with. But when Maury used the logbooks, among other things, to create charts that proved themselves *useful*, we then saw Maury produce value from something without value. He made something useless useful. Here we can then see that Maury establishes a new flow of information and ship logs usefulness.

If we shortly turn to a quote from M&C, we see their argument of how make Big Data yield more use cases is to make it extensible into more uses:

> "One way to enable the reuse of data is to design extensibility into it from the outset so that it is suitable for multiple uses. Though this is not always possible – since one may only realize possible uses long after the data has been collected – there are ways to encourage multiple uses of the same dataset." (ibid., p. 109)

In the above quote, M&C explore the concept use cases and where it may be found in Big Data. Thus, they explore how Big Data can become valuable which is through a process of designing extensibility of the datasets use cases into their design. We can reuse data to find multiple use cases. Let is now return to Maury again. Data in the context of extensibility and reuse can also be seen in how Maury collected information from the old logbooks, the barometers, the compasses, the sextants, the chronometers, and the information which he gathered from the logs that he made ship captains fill out after their journeys. Here Maury's *information* equals to Big Data's flows of *data*, as it is exactly this data that enables the usability of Big Data itself.

Not only does Maury produce use cases for the navy but also for astronomers, as he managed to map Neptune's orbit from data which was initially regarded as useless. In this accomplishment, we find that Maury again used useless information (the star maps were made on the presumption that Neptune was a star and were thus discarded) and made it usable. He made use of navy mapping techniques in the astronomical discipline. By this he not only mapped Neptune's orbit, but expanded the research methodology of astronomy altogether and too expanded the uses for the discarded star maps.

Shortly, turning to a quite from M&C, they describe the uses and potential uses of Big Data as follows:

> "Data's true value is like an iceberg floating in the ocean. Only a tiny part of it is visible at first sight, while much of it is hidden beneath the surface. Innovative companies that understand this can extract that hidden value and reap potentially huge benefits. In short, data's value needs to be considered in terms of all the possible ways it can be employed in the future, not simply how it is used in the present." (ibid., p. 103)

This notion of potential multiple uses for the same data, M&C coin as *option value* (ibid., p. 102). The option value is potentially endless and only has virtual limitations as the company's ability to surface its potential uses. In the possibilities of data use cases we find potential value for the company (ibid., p. 104). The data's value is not inherent in itself, it is rather tied to the processual abilities and the know-how of the company. Data is not valuable by itself but only to the extent that companies or people find uses for it. Much like the navy in the example of Maury. The navy had all the data they needed to make the maps. However, they regarded the logs useless as they lacked the skills to utilize the data's potential of making good sea charts. Maury on the other hand had this ability and made the data useful. He was able to connect with its *option value* and realize the potential.

Rounding up this section, we can now conclude that when we speak of value in terms of Big Data, we are speaking of its uses.

Maury took something of no value – information which on the surface served no use – and made it valuable by making sea charts that made the sea farers' journeys shorter and safer which served great usability. Through Maury's example we learned how he opened up new flow of information in the navy industry. He engaged old logs, created information formulas and thus established a flow of information to develop the mapping charts of the seas. He even engaged astronomy and expanded the method of research, as his navy methods would determine the orbit of Neptune. Now, Maury's approach to logs, sailors, and diaries does open one more notion to this analysis of Big Data, as he shed light onto *datafication*, the process of translating the world by data. This leads us to the next section.

The process of datafication

In this section we will explore the notion of datafication. Datafication is defined as: "To datafy a phenomenon is to put it in a quantified format so it can be tabulated and analyzed." (ibid., p. 78). Thus, datafication is the process of translating phenomena into measurable data. We will here explore how datafication is a production of subjectivity by the expansion-machine. This will be shown in how Big Data enables people engaging with it to perform this act of datafication and thus expand the amount of incoming data streams they have.

An example of datafication is found in M&C's story of Shigeomi Koshimizu, a professor at Japan's Advanced Institute of Industrial Technology in Tokyo.

Koshimizu managed to datafy the act of simply sitting in a car seat. He placed sensors in car seats that would measure a multitude of pressure points on the seat.

> “Koshimizu and his team of engineers convert backsides into data by measuring the pressure at 360 different points from sensors in a car seat and indexing each point on a scale from zero to 256. (...) [T]he system was able to distinguish among a handful of people with 98 percent accuracy.” (ibid., p. 77)

Koshimizu and his team aspired to take this data on sitting postures in car seats and use it for detecting irregular posture. But the data proved to be useful for preventing car theft as well by identifying a driver’s reflexive shift of position prior to an accident. The car would thus be able to react before the incident occurred. We here see datafiction in action. Koshimizu took the phenomenon of sitting in a car seat and datafied it into data. Kosimizu now had a new data stream of car seat postures which he didn’t before.

Moving away from the example of Koshimizu and turning back to the story of Maury, we find that Maury also practiced datafication. as

> “(...) [H]e understood that the Navy’s musty logbooks actually constituted “data” that could be extracted and tabulated. In so doing, he was one of the pioneers of datafication, of unearthing data from material that no one thought held any value.” (ibid., p. 76)

Maury found that the old logbooks could be used as data to form valuable use cases in form of his sea charts. He thus took a something of seemingly no use for Big Data and converted it into data which can become valuable and be used for different things. Maury expanded the amount of data streams available to him.

We’ve have seen in the two examples above how datafication as an act produces new flows of information and opens up new use cases. At the same time, what slip past constantly, is the momentum of datafying as many phenomena as possible in order to extend and enhance the scope of the Big Data tool. Maury did it with the sea charts, expanded the charts with different types of information and even expanded the method into a whole different discipline. Koshimizu found ways of not just registering car theft, but may have opened a gateway to more secure car accidents in the future as well. We also see this when M&C mention that the datafication of literature around the world has enabled Big Data to "understand human behavior and cultural trends through the quantitative analysis of texts." (ibid., p. 84). The same notion of expansion arises in when we look into our pocket phones. These devices produce immense data every day, as we message each other, use the location feat, and scroll through Facebook news. The fact that today's cellphones are so good at recording data has led the director of MIT's Human Dynamics laboratory, Sandy Pentland together with Nathan Eagle, to create the concept of 'reality mining' which describes how cellular devices can help supply the data that can predict accurate human behavior and social behavior (ibid., p. 90).

These datafied areas are just the tip of the iceberg, according to M&C. Big data and the ones who use it will always want to expand the scope of data as it opens ever more opportunities for new uses and efficiency. As M&C write,

> "Through datafication, however, in many instances we can now capture and calculate at a much more comprehensive scale the physical and intangible aspects of existence and act on them." (ibid., p. 97) and "Once the world has been datafied, the potential uses of the information are basically limited only by one's ingenuity." (ibid., p. 96)

Drawing upon the concept of the machine from Deleuze and Guattari, we can see that datafication is an expression of an expansion flow inherent in Big Data. If we analyze the above examples with the concept of the machine, we find that there is an expansion-machine at play in Big Data. The expansion-machine enables new connections through datafication.

The fact that data from a car seat can now enable the identification of theft and prediction of car accidents is a new connection formed by this expansion-machine.

Koshimizu can be seen as working the flow of the expansion-machine as his link with the expansion-machine has managed to produce this new connection through the expansion of incoming data streams (making the car seats record data) and expanding uses of data (identifying theft and predicting accidents).

When Maury datafied the old logbooks to improve his maps, we also saw a new connection form due to the expansion-machine. Prior to Big Data and the expansion-machine, no one thought of making such a connection. However, when Big Data and Maury entered the picture, they made this new meaningful connection between the logbooks and the sea charts. Thus, Maury expanded his sources of data to be used for his maps, acting according to the flow of the expansion-machine.

Sandy Pentland and Nathan Eagle's concept of reality mining is also an example of new machinic connections established by the expansion-machine. Through Big Data and the expansion-machine, a connection between cellular devices (gathering data from social interactions, GPS locations, etc.) and predicting human behavior takes form. Thus, we see how datafication can be interpreted as a flow of a machinic expanding and a confirmation of new connections.
Ending this section, we have analyzed the two examples of Maury and Koshimizu to see how the expansion-machine produced subjectivity in enabling their acts of datafication. We have thus in this section established a machinic production of subjectivity which is the expansion of the number of available data streams.

We will now look towards how the expansion-machine also enables the expansion of volume in data in already established data streams.

Bigger is better

In this section we will move away from the theme of datafication and look into another aspect of Big Data. This aspect addresses how Big Data and the expansion-machine produces the activities of expanding the data from already established data streams. More data can yield more uses and more precise results. We will begin by looking back to the example of Maury.

After Maury had made his first charts, his project was far from over. Maury kept improving his maps by adding more and more data to his research. He extended his scope of data to not only include the components and information he found at the depot but he continually began to log ships' routes and even made ship captains throw out bottles in the ocean with relevant information about their positions and the weather (ibid., p. 75). Those water bottles would then be collected by other ship captains and given back to Maury's department to contribute to his maps' accuracy. As he added more data to his project, the charts became better and more precise.

The above is one example of how the logic within Big Data is such that it always seeks to grow larger in size of data. With more data, as with the Maury example, Big Data continuously yield better and more precise results. The expansion-machine makes the people connected to it collect more data from the flows of data it has already established.
the quantity of it is large enough, it can to a large degree disregard the quality.

To analyze the story of Maury with the concept of the machine, we see this expansion-machine present in how not only unusual components proved useful, but also in how Maury kept expanding his already established data streams (the fact that ship logs became more and more common among seafarers for example). This increased volume of data played a part in Maury's creation of new, more efficient and precise, maps. Thus, these new connections enabled new modes of being for the people in the navy that used such maps in how their routes now became different, safer, and faster. In short, the expansion-machine makes a new connection between quantity of data and better results. At the same time, the expansion-machine also forms a new connection in terms of connecting new and unconventional sources of data to better charts.

Going from Maury's story towards a new example in M&C's book, we find the story of Google and IBM that wanted to use technology and algorithms to translate languages (ibid., p. 36-39). Google, however, wanted to take the next step and have Big Data assist in translating languages. To do that, Google first wanted to approach the task in a different way that IBM had done prior to them.

IBM had at one time tried to translate Russian into English by translating "... sixty Russian phrases into English ... using 250 word pairs in computer's vocabulary and six rules of grammar." (ibid., p. 37). IBM's translation-machine was successful in translating sixty sentences, however, IBM had to admit further down the road, that their future aspirations in the field of machine translation could not be realized. It turned out that "The problem was harder than they had realized it would be. Teaching computers to translate is about teaching them not just the rules, but the exceptions too." (ibid.) IBM then learned from their previous mistakes and launched a new translation-machine which

> "(...) used ten years' worth of Canadian parliamentary transcripts published in French and English – about three million sentence pairs. Because they were official documents, the translations had been done to an extremely high quality." (ibid., p. 37-38)

Another difference between these two translation-machines was that the first one was augmented with a dictionary, while this newer one looked for the best matches through the use of statistics and mathematics. Statistics and mathematics were useful in this case as they had many transcripts which were perfectly translated into both English and French which laid the groundwork for a computer to statistically determine which translation would be best to solve the task put ahead of it. However, after some time they ran into problems again as the translation-machine was to solve more difficult questions, the engineers had to constantly rewrite the statistical and mathematical rules the machine translated by. The task of improving those rules proved too resource demanding when compared to the output it gave them (ibid., p. 37).

Google on the other hand wanted to perfect automated translation and achieve what IBM didn't. Google chose a different approach to the task. While IBM had very accurate and correct translations serving as the data foundation in their translation program, Google "… availed itself with of a larger but also much messier dataset: the entire global internet and more." (ibid., p. 38) From the messy depth of the internet where all sorts of translations could be found – both precise and imprecise ones – Google found fit data. Their translation-machine also did not adhere strictly to any type of data – all sorts could be used. And the result was very successful and became the tool known as Google translate today.

In the end, this method enabled Google to do two things: 1) translate languages much more precise than ever done automatically before and 2) they, due to the large sums of messy data, could now also translate one language into another for which very little translation had been done. They enabled translation bridging by using English as a mediator between the two languages far apart. Prior of google translate, Hindi and Catalan was almost not translated and the process of translating the two was but today everyone with internet can translate them. This was achieved by allowing data to be messy and imprecise as it enabled them to collect very large amounts. Google accompanied by messy but large datasets won the translation program race of IBM and Google.

If we delve into the above example of IBM and Google with the concept of Deleuze and Guattari's machine in mind, we see multiple logics at play. Firstly, we see the logic of expansion in how Google with more, bigger and messier data utilized its Big Data translation-machine to many things. These new results were not only to translate between common sets of two languages but virtually between any other language as long as there was a connecting link present in form of a third language acting as an intermediator (as English was in the example). So, the expansion-machine formed new connections between Google, large amounts of messy and unstructured, data and the ability to translate languages. And it is exactly this new connection demanded by the expansion-machine that made Google's translation tool better than IBM's. Another connection established here is between languages that was not directly translatable before. However, with Big Data and the expansion of the amounts of data, Google managed to bridge between languages. Google now make new connections by the expansion flow between languages which was hard to translate before.

In this notion of expansion, we see other new machinic connections take form. A new connection formed between translating languages and the means of doing so. This is evident in how higher quantity opposed to quality led to Google's automated translation-machine. The Google example compared to the case of IBM shows how Big Data and machinic expansion (at play in Google's case) contrasted the usual common sense-driven approach of IBM.

Another logic at play in the example is that messy data is better and only because it enables for gathering large amounts of data. We see that Google was able to produce more uses with their data because they broke free of the limitation that IBM worked with, namely, that the data had to be of high quality (IBM only used governmental transcripts). However, this logic has another component to it which is the natural effect of allowing the use of messy data. When you allow the use of messy data you at the same time enable yourself to gather data at much larger quantities – and large quantities of data can yield many new uses and much more value. This pairing of messy data contrary to precise data and usefulness can thus be seen as another machinic connection established by the expansion-machine.

We have in this section analyzed the examples of Maury, Google, and IBM and seen how the expansion-machine produced new flows of subjectivity. The produced subjectivity was seen in how expanding the data from already established data streams, even if the quality of it is low, can lead to better results. Thus, the expansion-machine produced actions of expanding the volume of data in already established data flows. We will now move on to the next section which is about the data scientist.

The Generalist

A key component in making data valuable is the data scientist. In this section we will explore how the expansion-machine produces subjectivity for the data scientist. This becomes evident in how the data scientist, working with Big Data, always will try to expand his or her fields of competences.

We find in M&C's book that it is the data scientist that designs the Big Data-tools, finds the valuable insights, and formulate his or her findings into potential uses. However, the data scientist is an interesting person. In the example of Maury, we are briefly introduced to the data scientists which helped Maury making his charts at that time:

> "Maury and his dozen "computers" – the job title of those who calculated data – began the laborious process of extracting and tabulating the information trapped inside the deteriorating logs." (ibid., p. 74)

Here we are not only introduced to Maury's 'data scientists' but also to the fact the Maury himself acted as one as he worked besides them and made similar work: "Maury aggregated the data and divided up the entire Atlantic into blocks of five degrees of longitude and latitude." (ibid.) Maury's engagement with the data was what enabled him to divide the Atlantic into section which was one pillar in making his charts. Such work is exactly that of the data scientist.

We are also presented to how Maury tried to expand his competencies in order to make his maps:

> "Whenever Midshipman Maury arrived at a new port, he would seek out old sea captains to gain their knowledge, based on the experience passed down for generations. He learned about tides, winds, and sea currents that acted in regularity" (ibid.).

To make his maps, it wasn't enough that Maury was good at tabulating and calculating simple data which he gathered, he also needed to gain knowledge about the field for which his data and its value should become relevant.

We can we see a flow of expanding one's field of expertise is necessary if one want's to work directly with Big Data. In the example above, it would not suffice to be neither a skilled statistician (who knew how to handle data) nor a skilled geologist (who knew about the "(...) tides, winds, and sea currents" (ibid.)). Maury had to be both, so he sought out the professionals that could give him the knowledge needed. One's skills and knowledge had to extend further, and through different research disciplines.

M&C are not the only ones who speak of the data scientist as a multi-skilled worker. If we turn to Granville, he describes the data scientist as follows:

> "Data scientists are not statisticians, nor data analysts, nor computer scientists, nor software engineers, nor business analysts. They have some knowledge in each of these areas but also some outside of these areas." (Granville, 2014, p. 73)

So, the reason why people working with Big Data are not coined 'statisticians' is that statistics isn't a field which operates with so large amounts of data, which inherently necessitates more delicate IT-skills than the average statistician possesses (ibid., p. 74). The data scientist has to have more skills than that. In general, the data scientist isn't categorized under any of the above mentioned titles, as he or she both possesses knowledge within all those fields,

> "... data science = Some (computer science) + Some (statistical science) + Some (business management) + Some (software engineering) + Domain expert + New (statistical science) ... (ibid., p. 75).

Finally, the data scientist cannot be limited to the above competencies. He or she must also have a solid understanding of the enterprise he or she is working within. Granville suggests that relevant enterprise knowledge as well as general business knowledge too is vital for the data scientist. This is because the data scientist not only needs to analyze data streams but also make systems for handling data as well as turning the data into valuable insights, tailored for the company's specific challenges and goals (ibid., p. 73-78).

Thus, Big Data demands that not only itself expands but that the scientist working with it also expand in terms of skills. The data scientist must thus expand their scope vertically instead of specialize horizontally in specific fields (statistics, mathematics, etc.) (ibid. 76).

M&C agree with Granville, that the data scientist is a must for the data-driven company and at the same time, he or she should have a wide combination of skills.

> "... the "data scientist," ... combines the skills of the statistician, software programmer, infographics designer, and storyteller. Instead of squinting into a microscope to unlock a mystery of the universe, the data scientist peers into databases to make discovery." (Mayer-Schönberger & Cukier, 2017, p. 125).

The above skills are what enables work with Big Data and this work entails finding discoveries within datasets. Elaborating further on this quote, the data scientist is no longer a person with a few specialized skills which would enable him or her to work with specialized areas of research. To work with Big Data and extract value from it, one must expand one's skills to encompass many fields of knowledge. Following this logic, the data scientist's skillset would be expanding across constantly counting fields of expertise.

Speaking of the data scientist in terms of the machine-concept, the expansion-machine forms new connections between the data scientist and different academic disciplines. First it may be statistics, but then Big Data may become even more complex through its expanding. Then it must form a new connection between the data scientist and another discipline (domain expertise, mathematics, etc.). This continuous expanding and formation of new connections could continue as long as Big Data keeps expanding itself – as long as we maintain a connection to the expansion-machine.

In the example of Maury, we saw how the expansion-machine demanded the expansion of Maury's fields of knowledge. The expansion-machine in that case established new connections between Maury and the professionals which he maybe never would have met if it wasn't for his connection to the expansion-machine. Other than that, the expansion flow made a connection between the expansion of competencies and the creation of the sea charts.

This section has explored how the expansion-machine creates the expanding behavior of the data scientist. This expanding behavior is one of always trying to acquire new competences, and this behavior is how the expansion-machine produces subjectivity for the data scientist.

Expansion-machine surfaced

Through the first half of chapter five: *Finding machines* we've went through a number of relevant examples as to surface our first of two machines at play in big data. The machine connects to other machines and in between a flow is enabled, creating and enabling the production of subjectivity.

Our first encounter of the expansion-machine in connection emanates as Maury engages the navy logs, diaries and notes with a different approach. Maury collects useless information to create and expand its use cases. The development makes apparent how Maury connects his collections of old information from seas with the expansion-machine. This connection created a flow of information and developed what would be his popular sea maps. Thus, Maury's activities in producing the maps was subjectivity produced from a connection to the expansion-machine.

Second notion to the expansion-machine surfaces in the event of Koshimizu. Koshimizhu would take an event, in his example car seats, and perceive it through data. He would understand the car seats by measuring pressure points. Through his data he would view the car seats performance and thus his perception of the car seat became *datafied*. The next he did was to connect this datafication to the big data expansion-machine and soon he had opportunities of detecting car theft and predict accidents. Koshimizu would create a new flow of information through datafication and connection to the expansion-machine. His action of datafication, of establishing new data streams, as well as expanding the uses for his data is an example of subjectivity produced by Koshimizu's connection to the expansion-machine. This event recoils in various other events as we saw with reality mining.

These two notions make up our third expansion-machinic flow. Now once the connection establishes to either existing data flows, or the connection establishes to new data flows – even both, we undoubtedly see an expansion of the dataset altogether. Here, the example of Google shows how the expansion of data produces more accuracy, and yields better results. The connection to the expansion-machine becomes a way to achieve better analysis, to achieve more precise Big Data tools.

Fourth notion addresses production of subjectivity in relation to the expansion-machine's connection to the data scientist. The data scientist has to connect constantly to the expansion-machine as to become a good data scientist. The knowledge pool of this professional simply keeps expanding and expands into new disciplines. These expanding activities of enlarging his or her own competences is how the expansion-machine produces subjectivity for the data scientist.

Our realization of expansion manifest itself as the expansion-machine connects and enables a flow of expansion of volume, of dataficaiton of world phenomena, use cases, and even as the role description of the data scientist expands. Every connection produce subjectivity as flow articulates.

Now that we have investigated M&C's book to find the expansion-machine at play, we will now turn towards a different focus. We argue that there is another underlying machine at play in M&C's *Big Data*. This machine has to do with trusting Big Data: the data-trust-machine.

The data-trust-machine

Following the analysis of the expansion-machine, another motion is also at play in M&C's *Big Data*. The following chapter will aim at analyzing *Big Data* again using the concept of the machine with a different focus. In order to do so we must reach into Big Data in order to *create* the machine in function. And so through palpating the examples and events of Big Data and by *doing* Big Data we are able to arrive at a conceptual stability that take the shape of the machine in production. This time, we will speak of an establishment of a data-trust-machine and how it demands that subjects must trust the data and its profound way of working. This data-trust is essential, in order to utilize Big Data's potential.

It sees something we don't

As covered, Big Data is a widespread phenomenon that we can engage with in many ways and so M&C invite their readers to enter their understanding of Big Data. The first notion of Big Data we will describe is its large scale. Scale, put simply, is the question regarding the amounts of data that Big Data has available. Scale is especially important in relation to Big Data as, "… it isn't just that "more trumps some", but that, in fact, sometimes "more trumps better."" (Mayer-Schönberger & Cukier, 2017, p. 33).

Scale is shown to be a vital advantage to Big Data as scale altogether makes for the most accuracy and increases the possibilities and uses of the Big Data as we spoke of in the section on expansion. An experiment proved a simple algorithm accompanied with a large amount of data beat a complex algorithm and even high quality data if it holds a significant lower amount of data points. We see this in M&C's example of the developers behind Microsoft Word's spell checking feature:

> "In fact, a simple algorithm that was the worst performer with half a million words performed better than others when it crunched a billion words. Its accuracy rate went from 75 percent to above 95 percent." (ibid., p. 36)

Scale becomes central and we are going to trade away quality in order to acquire *larger scale*, more data points. But as the scale serves to increase the Big data capabilities it limits the human comprehension of the information.

Big data set becomes more beneficial for Big Data's computing power, but simultaneously more incomprehensible for humans, as we cannot process the content due to scale. This is displayed in M&C's example of prematurely babies or "preemies". Prematurely babies were occasionally suffering deathly infection but doctors couldn't figure out why. Dr. McGregor, a researcher at the University of Ontario Institute of Technology, was determined to help dealing with the infection problem. By gathering large amount of data from "preemies", Dr. McGregor and her team was able to make quite a finding through Big Data analysis. M&C write the following:

> "Strikingly, Dr. McGregor's big-data analysis was able to identify correlations that in some ways fly in the face of physicians' conventional wisdom. She found, for instance, that very constant vital signs often are detected prior to a serious infection. This is odd, since we would suspect that deteriorating vitals would precede a full-blown infection. One can imagine generations of doctors ending their work-day by glancing at a clipboard beside the crib, seeing the infant's vital signs stabilize, and figuring it was safe to go home – only to get a frantic call from the nursing station at midnight informing them that something had gone tragically wrong and their instincts had been misplaced. (ibid., p. 60-61).

In this quote we see a very good example of how Big Data's analysis yielded quite accurate results and Dr. McGregor is designing to tool to measure and react according to her finding.

By engaging the preemies example with the concept of the machine we can reach new analytic understanding. Big data creates machinic connections between things that seem contrary to common sense. With the premature babies example we see how Big Data analyzed new knowledge but it went against common practice. This connection showed the infection occurred when vital signs would stabilize. But the connection yields on a notion of trust. The connection was only possible if they trust in what Big Data and its analytics can do without necessarily understand the result. As the scale of data needed to make this discovery of the preemies' proneness to infection had to be very large, people can no longer analyze the data by themselves. It was out of the hands of the ordinary doctor to analyze the data, and theory couldn't prove an answer. This means that a connection to a data-trust-machine must take place before the insight about the preemies' infections could be possible. Thus, the data-trust-machine opens up for investigations and the use hereof. Only if trust in data is engaged, they can utilize Big Data and the flows it produces. And so,

> "… it [Big Data] can detect subtle changes in the preemies' condition that may signal the onset of infection 24 hours before overt symptoms appear. "You can't see it with the naked eye, but a computer can."" (ibid., p. 60)

It is not what we can see, but what *it* can see that unlocks the possibilities of the data. It is not a matter of what is apparent for humans but what is *not* apparent for humans and what that may produce. Here we see a split. We see a split of what is understandable for machines and Big Data, and what is understandable for human comprehension. Here we saw a moving away from what can be analyzed by people. And a move towards Big Data and what it can unlock of information and potential of data analytics. But only conveyed in so far as one is willing to sacrifice understanding of the data.

Trust forming the data basis

By following the notion from before we can continue investigating this split between humans' and Big Data's perception by looking into how Big Data tools analyze. One leading example of Big Data is to be found among social network platform. These internet-based pages work by engaging people socially and having people interact and share content online. Here we cannot avoid talking of major actors such as Facebook, Twitter, and Instagram. These Big Data sets are all around as,

> "With Facebook, Twitter, LinkedIn, and other social network platforms, our personal connections, opinions, preferences, and patterns of everyday living have joined the pool of personal information already available about us." (ibid., p. 100)

All this information is coming from us and made by us. All the time we are online and engaged with these platforms. From these databases, Big Data will have access to our friends, our hobbies, and everything in interest – as long as we have trusted these platforms, used, and engaged with them.

This not only enables opportunities for marketing, along with knowing your political tendencies or what your next kitchen brand will be. This potential of Big Data analysis simultaneously attracts and demands trust. Trust to the owners of the data; trust that the content is not misused for the wrong purposes or shared without consent, and that the data security is on par. Here we invest trust into the domain of data once again, data trust.

This data that makes up this companies' data set are generated form user activities. Now once we connect with Facebook we are creating data. So, without our consent, and therefore without our data the Big Data tool becomes irrelevant. If we distrust the platform, we may close down our profiles, disconnect to the data environment, and break down the expansion of our personal data.

In M&C's book, we see a data-trust-machine is necessary when we speak of Big Data based on social platform or user data in general. We will now look deeper into these two analytic notion of: trusting data you don't understand. And through user trust, the Big Data creates data. Now, we will turn to an event of it in an example from M&C's own book regarding manholes in New York City.

Big data in NYC

In New York City manholes were blowing up as the sewer catch fire underground and an explosion would follow. The explosions would shatter the structure above or throw the metal cover several stories up in the sky. (ibid., p. 68) No one knew the cause of these fires but something had to be done. Through Big Data analysis, data scientists were able to locate the next manhole catching on fire before it would occur. Con Edison was a public company that provided NYC with electricity which also did the maintenance of the same manholes that would randomly explode. At one point,

> "Con Edison turned to statisticians uptown at Columbia University in hopes that they could use its historical data about the grid, such as previous problems and what infrastructure is connected to what, to predict which manholes were likely to have trouble, so the company would know where to concentrate its resources." (ibid.)

The researcher was not able to investigate why the manholes would blow up, as it was far too risky. Furthermore, the testing of any such hypothesis would take too much time. The team turned to Big Data analysis to find a pattern for the manholes catching fire. This turned out successful as "The top 10 percent of manholes on their list contained a whopping 44 percent of the manholes that ended up having severe incidents." (ibid.)

The scientists did not know of the cause, but through Big Data, they could detect a tendency in the pattern of the sewer fires and thus predicted the next happening before it would occur. M&C sum up the example by stating that we have "… to put our trust in correlations without fully knowing the causal basis for the predictions." (ibid., p. 70).

In this example, we see researcher make a decision how to approach the problem. Sending out resources would be both expensive and dangerous, so they decided to use Big Data analysis which led to connect a pattern of the incidents, and helped bring manhole incidents down.

As show here Big Data are not just finding new insights when answering many of the questions that were previously costly but, rather, Big Data pose a new way of investigating: “Big data may not spell the “end of theory,” but it does fundamentally transform the way we make sense of the world.” (ibid., p. 72) We may have Big data come in and fundamentally alter our ways of researching, our way of perceiving the world, and way of thinking altogether. We may have to engage with new types of researching and we may have to encounter our questions and our problems in new ways.

If we engage this example with the machine concept we assess machine flow given by predicting the pattern of manholes incidents. This flow of information was only obtained as research sought for new methods as the old way were too costly and risky, and found Big Data analytics. By privileging Big Data analytics, Con Edison obtained new information as for where the incident would occur, but had to give up understanding the causational basis of the predictions. The example serves to underpin the data-trust-machine’s connection to Con Edison. Here we see establishing of a new flow and disconnection of another. Big data would feed new information as to where the incident occur, but would take away the causal knowledge for the incidents. By engage the data-trust-machine Con Edison achieved new insights, but had to give up a causal reasoning.

Summing up the data-trust-machine

A notion of data-trust is inevitably at play in our reading of Big Data. Data unlocks and enables flow of information and opportunities coming from Big Data as was the case with the premature babies – thus producing subjectivity. But in this case, it was apparent that the research was not penetrable by human comprehension due to vast dataset, but for the computer it’s gold. And so in order to yield the Big Data results we have to trust the computing analysis without understanding the data. In the premature babies example the output of the analysis was even against common sense: as vital signs stabilize, the baby was going into infection.

Leaving the hospital and babies example, the social network platform, websites and forums, internet users engage with every day, rely on a user data-trust as well. When uploading pictures, liking newsfeed and reacting to events, the user leaves behind a trail of data. Much of which holds personal value. Data-trust becomes apparent as a machine in how it enables companies' incoming streams of consumer data which both companies and governments can use for various purposes within their business or society's interests. The data-trust-machine is fundamental for Big Data as it is only through it that incoming streams of data can become relevant. Thus, the data-trust-machine produces subjectivity of internet users' in their activity on the internet and social platforms as well as companies' usage of data streams for their businesses.

The NYC manholes proved that the data-trust-machine made connections between datasets on the manholes and finding the manholes that were about to explode, thus becoming more effective. If Con Edison had never connected to the data-trust-machine, they may not have been able to prevent the problem as urgently. Thus the action of trusting Big Data in this instance and use Big Data to solve the problem was an example of how the data-trust-machine produces subjectivity. This trust relation generated new methods of research to Con Edison's problem, but trusting Big Data also disconnected him from explaining the basis of their problem. Con Edison's data-trust connection produces new subjectivity as the circumstances changed.

Big data poses a new approach to knowledge: either we trust data based research or we leave all the potentialities of Big Data. Big data and the data-trust-machine produces subjectivity in how we have to make a leap of faith by connecting to the data-trust-machine or else we leave Big Data completely. We make a choice of enabling Big Data through the data-trust-machine or we disconnect from the data-trust-machine and dislodge Big Data in the same instance.

When connecting to Big Data, we operate something which we do not understand and utilize something we do not know how functions. Data-trust produces subjectivity and enable Big Data's analysis findings. It is through the data-trust-machine that we can connect with Big Data, it is also the data-trust-machine that leads us into the expansion-machine and the modes of being that it produces. We will look more into that in later chapters.

Now, with the help of Lazzarato, we will try to show how both the data-trust-machine and the expansion-machine can be linked to Lazzarato's social subjection and machinic enslavement.

Chapter 6 – Machines, Machinic Enslavement, and Social Subjection

Now we will take our previous notion of the expansion-machine and further nuance it using Lazzarato's concept of machinic enslavement. We will here re-read parts of *Big Data* through Lazzarato's theoretical perspective and find both conceptual commonalities between the expansion-machine and Lazzarato's machinic enslavement. This will give us a certain grasp on the expansion-machine which in turn will stand in opposition to the data-trust-machine and Lazzarato's social subjection to show how Big Data might create subjectivity.

In the first part, we will look into how the process of datafication (of translating phenomena into data) is a mode of machinic enslavement and also a trait of the expansion-machine. Datafication makes people engage with Big Data by datafying event into data for Big Data to use. Furthermore, we also see how data and people in a sense merge together, as everything in the machinic enslavement of Big Data can only be seen through and made sense of through data. In this merging, the asignifying semiotic component of the assemblage in machinic enslavement is data. Everything is datafied and translated to fit Big Data use through an automation process.

In the second part, the example of the preemies from M&C's book will show how a data scientist practiced social subjection by making a distinction between what people can comprehend and what computers can comprehend. The data scientist emphasizes that the useful results of Big Data cannot be understood by people but only Big Data-tool. Thus, she subjects people to a category of trusting that which they don't understand in order to use it, through a connection to the data-trust-machine. Apart from that, we will take a look into three management consulting firms' way of speaking of big. The three companies propose companies to become data-driven as opposition to not data-driven companies. Here inciting the necessity of connecting to the data-trust-machine in order to use Big Data and become data-driven.

The expansion-machine as machinic enslavement

To quickly sum up the concept of machinic enslavement, it is a concept which operates beyond the domain of social subjection. Thus we are moving beyond the individuated subject. The machinic enslavement resides in the asignifying semiotics, which regard all the things around us that smoothly moves us along in machinic enslavement without demanding explicit attention. Lazzarato refers to asignifying semiotics as the significations that "… slip[s] past rather than produce significations or representations." (Lazzarato, 2014, p. 80) It's all entities we connect with as we adhere to automation, or as we do processes by default. Here the subject is no longer an individuated subject, but rather also just a component in the assemblage, in the collective constellation of the event. Guattari describes the subject in machinic enslavement as "a gear, a cog, a computer component part in the "business" and "financial system" assemblage" (Lazzarato, 2014, p. 25). The subject is part of something bigger, it's a component along with everything else, human or non-human, to form an assemblage. Within an assemblage we find a myriad of components that make machines, which partake in performing the enslavement.

With this, we aim at addressing the things that slip past the representations in the text or put otherwise, we aim to read between the lines in M&C's book. We will try to make out which entities slip past, as we are connected with the expansion-machine. We will extract what is implied and how these implications make up the asignifying semiotic at function as we are in machinic enslavement. And so the initial approach here is to go back to the Big Data examples and try to make out the semiotics of the expansion-machine in function.

Expansion and asignifying semiotics

Earlier on in the previous chapter, we identified the expansion-machine through a reading of M&C's *Big Data*. The expansion-machine relies and connects to datafication as to constantly expand its territory, significance and utility. Its capabilities are ever-increasing. We saw with the Maury example how old ship logs became reused and recycled through mapping techniques that lead to a better solution for ship maps. These logs of navigation, gathered from various ships were combined, and collectively they form as the assemblage of the navigation map. In this process we see how data is consumed and integrated into the assemblage of maps that makes up the complete map that Maury had designed. This mapping structure would be increasing information. The maps was expanding as more possible sailing routes got datafied and collected into the map assemblage. Along this development of the navigation maps there's an expansion of the assemblage of routes, navigation notes and map details. This development of the map assemblage is slipping something past the event as the process is going on. Along the development, the expansion-machine of maps and navigation information is constantly forcing event to fit data, to fit the map, to fit the assemblage. When Maury gets one note, diary or drawing, that map assemblage is developed and expanded. Always producing the new map and a better map:

> "Regardless of the kind of assemblage, expression and content are continually to process of deterritorialization which asignifying semiotics and machines allow to be harnessed, controlled, as well as produced." (Lazzarato, 2014, p. 87)

Here Lazzarato outlines the machinic enslavements and the asignifying nature of the concept. Maury is producing the map together with harnessing an expanding a field of information. Always developing the assemblage as the expanding production is going on. The Maury map is developing along with new information as to make up the expansion as machinic enslavement motion.

Koshimizu

This motion of gathering and constantly expanding is not just at play in the Maury example. When we close in on expansion in datafication - the phenomenon of converting everything into data, we find other examples such as Koshimizu's car seat data: The Koshimizu example where sensors on car seats and analyses of the data led to new insights on many different levels, from detecting car theft to predicting car accidents from bodily posture before the accident would happen. Big data will, through expanding datafication, be able to help out in many different ways. Koshimizu's car seat example connects with expansion and datafyies the world and so the asignifying semiotic datafies the world, molding everything, controlling and harnessing everything.

Datafication not only poses a variety of opportunities, but even offers enhanced ways of conducting research. Datafication drives the examples of Maury and Koshimizu every time it enables more precise results and a new way of perceiving the world. This datafication then is a clear expansion-machinic movement as part of the Big Data phenomenon. As part of this movement of datafication we find a notion that is invisible throughout the examples. This notion slipping by is data, as we quantify the world in the process of perceiving it with Big Data. Datafication is not only a matter of reading and understanding the world. It is conveying and decoding the world through quantitative values. Datafication and reading the world as quantitative data creates the world in the quantitative way alongside the reading process. As we want to assess events with Big Data it produces the very understanding of event. And with this subjectivity is engaged with the machinic enslavement, and every entity it come across is decoded and harnessed into the assemblage.

Datafication entails a certain type of mode of enunciation, a certain way of perceiving, as it forces everything into quantitative data:

> "The asignifying semiotics and machines (economic, scientific, etc.) these functions put to "work" are connected with subjectivity and

> consciousness. But it is not solely nor mainly a matter of reflexive consciousness or human subjectivity. Above all, they mobilize partial and modular subjectivities, non-reflexive consciousnesses, and modes of enunciation that do not originate in the individuated subject." (Lazzarato, 2014, p. 89)

The datafication is not only an enabling of the expansion-machine but it also proves to mobilize certain consciousnesses that move the subjectivity altogether. As we are conceptualizing entities by data we are also changing everything into data altogether.

Through the previous paragraph we've established Lazzarato's concept of asignifying semiotics that constitute the machinic enslavement. By applying asignifying semiotics into the expansion-machine's flow in the examples above we've distinguished the machine in motion. These data examples feed into the expansion-machine and poses the very creation of new relations, new insights, ideas and change of state of old practices even.

Partnering up with data unlocks many new territories and in the conveyance of these new territories Big Data seems to imply a certain reading, perceiving and decoding notion as well. Along with all the examples stated above the expansion-machine also entails a certain data decoding. Everything that feeds into the machine of expansion and in Big Data in total has to be data in the first place. First a phenomenon has to be interpreted through data before Big Data can function with it. In order for Koshimizu to establish new findings with our car seats, first he had to sensor everything in data and quantities, he had to absorb the event by datafication. This puts the perceived and the perceiver into a deadlock as everything read as 'data' inevitable has to be understand in 'data'. Data becomes what is slipping past every event. Data becomes the asignifying semiotic that enables the process of the expansion-machinic flow. Subjects are enslaved to perceive the world through data, and decode everything into data as we continuously expand our use of Big Data. Like Lazzarato says,

> "In machinic enslavement, the individual is no longer instituted as an "individuated subject," "economic subject" ..., or "citizen." He is instead considered a gear, a cog, a component part in the "business" and "financial system" assemblages..." (ibid., p. 25)

We have now seen how the expansion-machine in Big Data works as machinic enslavement. We will now take a step backwards and re-read M&C's *Big Data* along with several articles by leading consulting management firms to investigate how the data-trust-machine resides in the domain of social subjection.

The data-trust-machine as social subjection

In this section we will take one more analytic assessment of the data-trust-machine. That is to describe the data-trust-machine in relation to social subjection. Social subjection has a conceptual relation to machinic enslavement. The two are different domains enabling two ways of producing subjectivity. More than that, social subjection substitutes machinic enslavement when it breaks down. This movement too is a part of the production of subjectivity.
In the previous chapter *finding the machines*, we touched upon some of the notions that will be further outlined here. The aim of this part is to outline the domain and nature of the data-trust-machine as social subjection. Again we will see much of the input was initially established in the previous section on the data-trust-machine will be used again. First a remark on social subjection.

Social subjection operates with a dualistic logic which aims at dividing the world into dualities such as man/woman. One way social subjection articulates itself is by use of language. Lazzarato writes,

> "The "linguistic machine" and its theories are pressed into the service of the law, morality, Capital, and religion. They systemize, structure, consolidate, and enable power formations." (Lazzarato, 2014, p. 79)

Social subjection is a matter of exercising language to subject individuals into positions in the society, and so by establish roles and identities.

We will in this section use two examples to show how language and data-trust establishes divides and dualistic relations. We will begin with a look back into the example of the "preemies". Then we will look towards three international and interdisciplinary consultancy firms and their way of articulating the importance of trusting and using data in your organization as opposed not to.

You don't have to see it to believe it

We will start by returning to Dr. McGregor in the example of the preemies. In this example, Dr. McGregor found that the stabilization of the preemies' vital signs was not a sign of betterment but was rather an indicative sign of a breakout of a serious infection. When Dr. McGregor was elaborating on Big Data's efficiency in predicting the preemies' coming infection she stated the following:

> ""You can't see it with the naked eye, but a computer can."" (Mayer-Schönberger & Cukier, 2017, p. 60)

The above sentence will act as our starting point for elaborating how the data-trust-machine also functions as social subjection. The quote from Dr. McGregor is taken in a context of Big Data's ability to answer questions. In the example of the preemies, Big Data found that preemies' stabilization of certain vital signs would be a sign of infection occurring hereafter. Big data do not offer us any concrete explanation as to why this is so. M&C does follow up with a suggestion as to why this might be present:

> "McGregor's data suggests that the preemies' stability, rather than a sign of improvement, is more like the calm before the storm – as if the baby's body is telling its tiny organs to batten down the hatches for a rough ride ahead. We can't know for sure: what the data indicates is a correlation, not causality." (ibid., 2017, p. 61)

We cannot know the cause, we cannot ask Big Data for it, but the important thing is that we know the infection occurs prior to its actual occurrence.

What the example of the preemies with Dr. McGregor's statement emphasizes is that Big Data is highly efficient at producing results. At the same time, you do not need to understand the data to know that it works and is valuable. You don't need to know the explanation as to why the preemies' vital signs stabilize prior to an infection. You just have to trust it. You have to put your trust in Big Data.

Dr. McGregor's "You can't see it with the naked eye, but a computer can" (ibid., p. 60) is an example of the data-trust-machine as social subjection. The fact that her project with predicting the preemies' infections has been successful adds greatly to the authority of Big Data. Big data can yield significant results, and Dr. McGregor must take a leap of trust as she cannot see what the computer can. The signifying semiotics is at play here, as she divided Big Data from human comprehension with her statement. Now acknowledging this divide Dr. McGregor exemplifies the identified subject as she reaps the benefits of Big Data use and conform to the established divide.

Lazzarato speaks of the national language as serving as signifying semiotic in social subjection. Everyone in society must conform to the official language of the state and its bound rules and structures to be able to navigate in society. This sets the fundamentals needed for controlling and subjugating individuals through exactly that language:

> "By assigning an individual subjectivity, an identity, sex, profession, nationality, and so forth, social subjection produces and distributes places and roles within and for the social division of labor. Through language it creates a signifying and representational web from which no one escapes." (ibid., p. 24)

In relation to Big Data, we can draw upon different words and distinctions within the use of language to determine in which position the subject is situated in relation to Big Data. This use of language can, according to Lazzarato, be utilized to fit people (companies and governments) into certain roles as subjects according to what they signify of meaning. In this case, through the utility of the data-trust-machine.

The data-trust-machine and the survival of the fittest

Now we will move away from M&C's book and the examples therein and look towards how large management consultancy firms speak of Big Data. We will look into a few articles published by BCG, McKinsey, and KPMG. They have a certain position of expertise in business management in common and they all have both lots of experience with and clear views on the subject of Big Data. We will now analyze their use of language and try to see how they relate to the data-trust-machine to and produce social subjection.

In the McKinsey published article, "Straight talk about Big Data" we are introduced to the following about the shift to a data-driven company:

> "Changes of this magnitude require leadership from the top, and CEOs who embrace this opportunity will increase their companies' odds of long-term success. Those who ignore or underestimate the eventual impact of this radical shift—and fail to prepare their organizations for the transition—do so at their peril." (Henke et al., 2016)

McKinsey posits multiple things. Firstly, they emphasize that becoming data-driven is a matter of life or death. That is emphasized in the use of the word "peril" which translates to danger and bankruptcy. They underpin how the companies won't keep up with Big Data companies and will become a victim of the disruptive effects of Big Data, if they do not join the movement themselves. It is not simply a matter of survival, but it is a question of either business success or perishing altogether. If we bring the data-trust-machine into play here, it may very well be the utterance needed for companies to pursue Big Data. The company must change to Big Data and with it achieves success or peril. Thus, the subjection of companies is defining them as data-driven with a bright future or the opposite: remain a traditional company moving toward bankruptcy.

To further justify the necessity of the data-trust-machine in McKinsey's social subjection, let's have another look at the same article:

> "Almost by definition, Big Data analytics means going deep into the information weeds and crunching the numbers with a technical sophistication that can appear so esoteric that senior leadership may be tempted simply to "leave it to the experts" and disengage from the conversation. But the conversation is worth having." (ibid.)

McKinsey refers to Big Data analysis as an esoteric [1] discipline for senior management implying that they very likely do not understand how Big Data analytics works. But they urge them to have the conversation. They continue to tell senior leadership to embrace this esoteric, difficult, but highly rewarding practice. They are encouraged to look away from the fact that Big Data seems esoteric but states, if they connect to the data-trust-machine they will see great results for their company.

In another article by McKinsey titled, "Three keys to building a data-driven strategy", we are presented with the following:

> "The lead concern senior executives express to us is that their managers don't understand or trust Big Data–based models and, consequently, don't use them." (Barton & Court, 2013)

McKinsey here tells us that the data-trust-machine is a focal point for data-driven companies and also exemplifies how the establishing a connection to the data-trust-machine is the enabler for using Big Data. Whenever the connection to the data-trust-machine is lacking, the new ways of operating the data-driven company

[1] Esoteric refers to something that is understandable to only few number of people due to specialized knowledge

are not taken seriously and thus shows as a hindrance in becoming a more successful and valuable company. This lack of connection still accommodates the established divide, we are still seeing the signifying semiotic producing the duality here, as either data-driven or traditional. This produces still same social subjection outcome as we've seen before, but it is a different connection to the trust machine. Here manager has explicitly disconnected trust from Big Data.

McKinsey is not the only consultancy firm addressing Big Data directly and recommend companies to become data-driven companies. KPMG write in their article, "Data-driven business transformation: Driving performance, strategy and decision making" that…

> "If you find your organization lags in data understanding and capability, it's time to take remedial steps." (KPMG, 2015)

In this quote, KPMG compare a company that lags data understanding and capability to a state of disease that needs a cure, or a problem that needs a solution – thus the remedial steps are of need. In this example, KPMG tell companies to make a the necessary towards the data-driven business which is the remedy for their problem of not being data-driven. The language here produces the same social obligation. The data-driven business stands in opposition to an "organization which lags in data understanding and capability" (ibid.). KPMG performs the same establishment as we've seen before: if you are lagging data you must change and become data-driven.

For our final example of consultancy firms practicing social subjection, we turn to BCG. In their article, "The Digital Imperative", they write the following:

> "Agile leaders try to find ways to better use internal and external data. BCG research shows that leaders in the use of Big Data generate 12 percent higher revenues than companies that don't experiment with Big Data. They are three times more likely than weak innovators to mine Big Data for new-project ideas…" (Dreischmeier et al., 2015)

Here we see a distinction between agile leaders and weak innovators. We can see the data-trust-machine playing a role here by means of 'experimenting'. Experimenting necessitates a discovering engagement with the object. To engage with something, we don't fully understand, here Big Data, is to ask companies to connect to the data-trust-machine and start experimenting. To reap the benefits (12 percent higher revenues) that Big Data yields you have to experiment and become the agile leader. The subjection here takes shape in BCG defining the successful subject as the "agile leader" (ibid.). To become the agile leader and not the weak innovator requires exactly a connection to the data-trust-machine which again enables the uses of Big Data.

The above examples show how consulting companies establishes identity role. They do so through language be emphasizing the downfalls of not using Big Data in your company. Other than that, the consulting firms states how understanding Big Data can be difficult and thus companies do not use Big Data. However, businesses should overcome this mistrust and connect to the data-trust-machine, make the leap of faith to use what they don't necessarily understand, and become data-driven and successful.

We have now seen how both the expansion-machine functions as machinic enslavement and how the data-trust-machine produces social subjection.

Production of subjectivity in machinic enslavement and social subjection

We will now both sum up our analytical points in the previous chapter. In addition, elaborate on how subjectivity is produced in Big Data by reading M&C's book and consultancy firms' articles through the lenses of machinic enslavement and social subjection.

We found that the expansion-machine is in the domain of machinic enslavement. On the basis of the examples of Maury and Koshimizu, what surfaced was the expansion-machine produces flow in an automatic manner: As with the expansion-machine and Big Data, everthing has to be converted to data to analyze it. Data

becomes an asignifying semiotic that guide the machinic enslavement to expand data and constantly datafy more.
By this data we are not just analyzing the world, but the very act of datafication is slipping past an understanding of the world through data. Now, as we datafy more and more of the world, more and more of the world altogether is understood as data, is managed as data and is by default becoming data. Data as the asignifying semiotic guides the machinic enslavement to produce a datafied world and the subjectivity altogether. Once an event has been datafied, it joins the machinic assemblage and the production of the subjectivity.

In addition, we argued that the data-trust-machine is in the domain of social subjection. We saw how subjectivity was created in how social subjection surfaced when Dr. McGregor made a distinction between people who could not interpret big datasets and machines which could. McGregor's invitation to connect to the data-trust-machine is also made in the context of preemies suffering infection or not. Here Big Data can "see" when the infection will occur and so doctors can treat them, without Big Data we cannot know when the infection occurs. This is also a signifying semiotic which produces subjectivity by establishing a duality, a choice between trusting Big Data or don't trust Big Data.

This chapter also argued that subjectivity is created likewise in social subjection in the articles written by the management consultancy firms on Big Data. Here in the context of companies. Like McGregor, the consultancy firms made a distinction between data-driven companies and non-data-driven companies by painting a picture through signifying semiotics of how becoming data-driven is needed if the company wants to succeed and even survive in the future. The consultancy firms established a binary for companies to adhere to. We here see subjectivity produced in how these signifying semiotics produce an establishment where companies either connect to the data-trust-machine and become data-driven or they go extinct.

Next we will look into the recent case of the data-analysis company, Cambridge Analytica which allegedly illegally acquired data from Facebook to use it to influence the recent US presidential election in 2016. This example will both serve to both activate the notions we have drawn out from our previous analyses but also contribute on how Big Data produces subjectivity.

With our analytical notions of expansion and data-trust machine along with the machinic enslavement and social subjection we have the analytic foundation from which we will continue our investigation on how Big Data produce subjectivity.

> "And so the mission will be to 'lodge yourself on a stratum, experiment with the opportunities it offers, find an advantageous place on it (...) produce flow conjunctions here and there, try out continuum of intensities segment by segment, have a small plat of new land at all times." (May, 2005, p. 25)

Chapter 7 – The Case of Facebook and Cambridge Analytica

In this chapter, we are going into a case that sprouted alongside the development of this project: The “Cambridge Analytica” scandal - or, as it is named as well, the “Facebook privacy crisis” or the “Unfair game of Donald Trump’s presidential campaign”. This event goes back to Ted Cruz’s campaign in 2016.

We will in this case see how Facebook users go through a cycle of production of subjectivity. The first part of the cycle consists of the subjects residing in machinic enslavement. Connected to the expansion-machine produces expansion two ways. Facebook users expand the Facebook data pool when adding more content like photos to their profile, connecting with friends, etc. Second expansion is seen as companies use the Facebook data to expand use cases and Big Data methods.

The second part of the cycle happens when it is brought to attention that Cambridge Analytica used Facebook data to manipulate American voters. Facebook didn’t manage to keep their data safe from such misuse. Here, we find a breach in the trust to the Big Data-platform Facebook. This moved the case into the light, and before a congressional hearing. Here we will analyze Zuckerberg’s speech with connection to the data-trust-machine. We will see how social subjection again is producing establishments for identities.

Machinic enslavement in the Facebook example

Facebook, Cambridge Analytica, and the US presidential election 2016

Departing from an article released by The Guardian titled, “Ted Cruz using firm that harvested data on millions of unwitting Facebook users” (Davies, 2015). Ted Cruz was collaborating with a small IT company named Cambridge Analytica and they were engaging with Facebook users’ data in order to enhance the Ted Cruz campaign:

> “(...) Cruz has turned to Cambridge Analytica for its unparalleled offering of psychological data based on a treasure trove of Facebook “likes”, allowing it to match individuals’ traits with existing voter datasets, such as who owned a gun.” (ibid.)

Cambridge Analytica would extract Facebook information to look for people’s preferences and spot potential voters from the Facebook data. The Big Data analysis aim was to improve Ted Cruz’s campaign and help distinguish what focus his speeches should have in order to persuade potential voter. This story was not revealing anything out of the ordinary, Ted Cruz answered. But there was one notion to the whole concept that was standing out: the process of the whole dataset and the functions of it. Ted Cruz went on to lose the election to become the republican candidate for the American presidential election.

He lost the race to Donald Trump’s campaign. The whole affair of Cambridge Analytica and data handling in the Ted Cruz camp wasn’t anything outspoken and, without going into the discussion of the media attention, the reveal of Donald Trump’s cooperation with Cambridge Analytica soon became the biggest focus.

Donald Trump’s campaign was now making use of Cambridge Analytica’s help as part of the republican presidential campaign in 2016. Cambridge Analytica gathered more than 50 million American Facebook users’ data through a survey tool. The survey tool did not just gather the data of the individuals who were doing the surveys, but also gathered all information on their Facebook profiles and all their Facebook friends’ profiles data as well. A survey counting 270.000 users soon ballooned to have not just the 270.000 users’ data but all their friends’ data as well (Collins & Dance, 2018). This was the data foundation Cambridge Analytica would use to develop a Big Data tool. They designed it read user habits, segment, and utilize to run personalized ads on Facebook (Cadwalladr & Graham-Harrison, 2018). Cambridge Analytica, through Big Data, came to wield a political tool which unheard of before and allegedly helped Donald Trump win the 45th United States presidential vote.

Now the theme that is relevant for the project in this case is on how Big Data was used to produce results and what reaction followed the event. We will begin by analyzing the case with the expansion-machine as machinic enslavement. In doing so we will palpate into the event at play here, starting off with the data gathering methodologies, the data handling, and tool development to the output and the management of this Big Data-tool.

A bigger Facebook is a better Facebook

As part of the Cambridge Analytica instance outlined above, the first engagement with Big Data that we can unravel is found in the particular data gathering, the consummation of the data set and in how this foundation of data that led to the Big Data-tool came into place.

We have to find the data gathering of this tool in Facebook as the mine itself. Facebook holds approximately 2,1 billion Facebook users (Statista, n.d.) that all have uploaded information into their account. The initial information that is inputted is email address that enables the account. From there you can find a myriad of opportunities of uploads, from pictures of yourself, pictures of your friends, family members, etc. The Facebook account also holds, if you let it, your legal civil status – whether you are married, single, "it's complicated", etc. It can hold information on your current job situation, your friends that you have accepted or added, your hobbies and a large number of other information. Ultimately Facebook can hold almost every form of information describing you, that you will allow and that you have uploaded. You produce data, as you upload information. The users see information, picture and newsfeed. But all along produce more asignifying semiotic, as data is constantly slipping past and expanding.

The outcome then is, many users are making use of this feature and uploading information as they see fit. And so the data pool of Facebook amounts to an almost unimaginable size. Though this pool of data is held only by Facebook and not up for the take of a third party. Now even though Facebook will not let a third party have the information as they please, the third party may harvest the information of a Facebook user if the Facebook user press “I agree” on an app “terms of service”. So, terms of service might very well include, that the company may download your Facebook information.

Seeing that data is the very foundation of the Big Data tool the data amount is, without doubt, the most central entity to acquire. The expansion-machine connects to prevail another flow: Aleksandr Kogan went on to develop a Facebook app called this-is-your-digital-life (Cadwalladr & Graham-Harrison, 2018). This app would pay the user a small amount of money if you completed a simple survey containing few questions. Along with the survey you had to accept the user terms, which would allow your Facebook data to be gathered including your friends’ content. At this instance the expansion-machine is connected.

The true feature of the expansion-machine came to play as the survey’s results alone did not play the major part of the data gathering. Rather the survey was a gateway into a realm of Facebook friends, a data expansion that would multiply the initial data output of the survey multiplied by the number of friends a profile had connected with on Facebook. Through a short period of time Cambridge Analytica amounted more than 50 million users’ data and thus had obtained a big dataset.

Molding the tool, enabling the environment

Now, with a large dataset, Big Data analysis comes handy as to work with the data. Cambridge Analytica are experts on this topic and soon they developed tools, segmentations and categories out of the data. Many Facebook users are aware of the vast information they are entrusting Facebook with. Some even realize this data is being used to advertise just the right advertisement for their profile, but

> “What you probably missed is that researchers had figured out how to tie your interest in Ms. Kardashian West to certain personality traits, such as how extroverted you are (very), how conscientious (more than most) and how open-minded (only somewhat). And when your fondness for Ms. Kardashian West is combined with other interests you’ve indicated on

> Facebook, researchers believe their algorithms can predict the nuances of your political views with better accuracy than your loved ones." (Collins & Dance, 2018)

The way these researchers believe to know you is through the OCEAN model. The OCEAN model consists of five different personality types. The model itself wasn't anything new, but instead a well-known psychology model, but the use of it across millions of profiles was new. Aleksandr Kogan's app, this-is-your-digital-life, through Facebook friends, unlocked the dataset and Cambridge Analytica made Facebook content eligible for the model. Now they had a datafied you to fit a model, from which they could personalize newsfeed for the individual Facebook user.

By knowing your interests through the pages you've 'liked' like favorite music, TV-series, sports hobbies, etc. Cambridge Analytica was able to compare your Facebook profile with the personal profiles of their survey subjects, who had undergone the survey. By the survey results Cambridge Analytica would distinguish all other Facebook profiles' specific OCEAN model profile based on their data.

Stanford did a similar model inspired from the Cambridge Analytica model. To give an example following their model would suggest: You like Tom Waits (you've "liked" her fan page), follow Miami HEAT (you've hashtagged their franchise when updating your Facebook-status), and have a lot of Christian friends (your friend's on your friend list). Through these pieces of information Cambridge Analytica could determine you as an open person, very agreeable, and not neurotic (Collins & Dance, 2018).

Now, the whole idea here is not to discuss whether the method was working, was precise enough or whether it was ethical. Rather the process itself is of interest here. Cambridge Analytica, as established, used highly relevant Facebook information in order to determine your OCEAN model profile. Cambridge Analytica utilized what was produced by the expansion-machine, amplified their initial test survey input with the data they obtained alongside the survey by connecting to people's friends. Through comparing methods, simple psychological models and a combination of data, Cambridge Analytica developed a new Big Data tool. Much like the traits of the Maury example, we see how bigger is better. Here with a political agenda the expansion-machine comes to combine various data sets, across surveys and Facebook user patterns to map out specific traits of subjects.

This is the very emphasis that Deleuze and Guattari has to their machine: "Machines do not fill lacks; they connect, and through connection create." (Buchanan, 2005, p. 124). The machine's function is to create new relations, new flows as to create the world. This is well proved in the example of the Cambridge Analytica case, where new relations, new expansion forms gave way to new insights, produced new relations, and new momenta for which the world is actualized anew.

Trump campaign, politics and automatization

Continuing the analysis, we will move to the last part of the Cambridge Analytica and Facebook case. In engaging with Big Data and connecting new data and disciplines Cambridge Analytica did so with political agenda, allegedly, to enhance the Trump campaign's chances of winning the United States presidential election. Their development of the segmentation led to a new way of approaching potential voters. The segmentation would distinguish you in the OCEAN model and so they could feed into the environment that you, as a Facebook user, interact with. Through their model they knew what would be of interest to you and by that not just target you as a voter but as a person. Through their psychological profile of you, and your "like" patterns on Facebook, Cambridge Analytica was able to feed into your Facebook content and target political nudges directly to you.

The Facebook accounts, Cambridge Analytica, and the Facebook users found themselves in an assemblage of machinic enslavement. This engages with Lazzarato's notion of enslavement as we see with his example of a car:

> “Driving mobilizes different process of conscientization [production of subjectivity], one succeeding the next, superimposing one onto the other, connecting or disconnecting according to events. Often as we drive we enter “a state of wakeful dreaming[.]”” (Lazzarato, 2014, p. 89)

When the subject engages with the data platform of Facebook, the data, information, and the subject become one single body. And the data as a asignifying semiotic interacts and produces the subject along with the subject interacting and producing the data. These flows goes both ways and co-create subjectivity in the realm of Big Data. Here Cambrigde Analytica is partaking in the production. Not only did they extract massive amounts of data about a number of Facebook accounts, they took part in the interaction altogether and thus became part of the assemblage environment of the subject. Cambridge Analytica took part in the production of subjectivity in machinic enslavement. The Facebook user, was unaware of the other player, Cambridge Analytica, whilst Cambridge Analytica was well aware of the Facebook users’ presence. The lesson here is not only, that Cambridge Analytica was able to reach out to American voter on a personal level. Cambridge Analytica took place as a component in machinic enslavement and thus became part of the automation process of subject, and became part of the production of subjectivity.

The episode involving Facebook, Cambridge Analytica and Big Data has revealed traits of the expansion-machine as machinic enslavement. By means of methods, Aleksandr Kogan and Cambridge Analytica, had engaged with new approaches or rather, they expanded the use of a Facebook feature in order to construct a big dataset that was eligible for Big Data techniques. By approaching Facebook users through an app, they would extract immense amounts of data by extracting not only survey results but simple “like” patterns, friend lists, photos and conversations. Expansion became a method in the data gathering process became the gateway into Facebook’s data pool, that would increase the amount of data into Big Data.

Once data was gathered the project went on to engage a psychology discipline by medium of the OCEAN model. Hereby enable datafication of the Big Data set and segment Facebook users. Following this enabling of method, the big datasets would prove useful, as they did not just target the Americans as voters, but on a personal level.

This feature played into the third momentum of the Cambridge Analytica and Facebook data adventure. As they now had immense data, Facebook users decoded into segments along the OCEAN model, Cambridge Analytica engaged with the users by interacting with the user environment, the machinic assemblage of the Facebook profiles. Not only were they reading Facebook data, but through segmentation techniques they were able to relate to the individual Facebook user, engaging the environment, and partake as machine in the enslavement automation process of the user. They didn't investigate the data trails the profiles left behind, but now they became part of the content of which the subject was engaging with and became part of the production of subjectivity.

The Scandal

Trump went on to win the 2016 election; Facebook's stocks continued to rise, and Cambridge Analytica went on to find new clients. The whole episode didn't spark much attention, until journalists started investigating the matter and moved the event into the spotlight around March 2018. Much of the content used to previous analysis is based on the articles produced around that time. With this attention, Christopher Wylie, who worked for Cambridge Analytica, come out with his story on what went on. The Guardian conducted an interview with Christopher Wylie explaining the Big Data-episode in detail, revealing techniques to the data gathering, analysis and its output (Cadwalladr, 2018). The culmination of the reveal led to interrogations, law cases, and a congressional hearing. Mark Zuckerberg went before a congressional committee hearing (Griffin, 2018) and Cambridge Analytica together with Facebook are facing lawsuits (Bowcott & Hern, 2018). This will be the entrance to a further analysis of the case, where we look into how Mark Zuckerberg connects with data-trust and how social subjection surfaces.

Social subjection in the Facebook example: Moving into the spotlight

Now that the case went public, got a lot of attention and the public heard a lot of news on the case, the central people to the event stepped into the light such as Zuckerberg who was called to a hearing before the US congress. Also, experts on IT and data security were mobilized as to help users understand data and how they could keep data safe from misuse (Stallman, 2018). The event altogether went from being in the shadows, something hidden from the public, to be visible and outspoken in the public. This moved the data questions into the light, from something we just produced by our actions to something the broad attention of society needed to investigate, understand and secure. Facebook and data security altogether got outspoken and put into contexts of law, ITs security, and even governmental interference. This is why this analysis will move along with the event, from regarding asignifying semiotics to signifying semiotics, and from machinic enslavement to social subjection.

Signifying semiotics and data-trust

Back to our case of Cambridge Analytica and Facebook, we find that data as asignifying semiotic was a central point. We know how Big Data was central to Cambridge Analytica's acclaimed success to influence American voters. When this all broke out in the news, the questions that we will follow are: what happens when we start being skeptical to data and when our trust to data breaks down? Trust to the system, like Facebook, on which we read news and keep in contact with friends and relatives. The platform we use on a daily basis. What happens when we ask question to the platform itself, when we ask Facebook's CEO to present before a congressional hearing, when ”60 minutes” newsmagazine start questioning the alleged involved people in the scandal, and we start filling law suits?

> “(…) [W]hen a breakdown, an accident, or a malfunction occurs, the subject function, consciousness, and representations must be mobilized in order to ”recover” from the incident, explain it, and mitigate its effects with a view to return the automatic functions and enslavement procedures to their normal state.” (Lazzarato, 2014, p. 38)

This subject function, consciousness, and representation adhere to the domain of social subjection. Our domain social subjection connects and relies on the data-trust-machine. We will now engage the congressional hearing with the data-trust machine to analyze social subjection as to mitigate and establish explanation for the scandal.

Chairman Grassley, at the US Congressional hearing of Facebook CEO Mark Zuckerberg, invites data-trust into play in his speech before the congressional hearing with Mark Zuckerberg by stating:

> “As “we get more free and low cost services, the tradeoff for American consumers is to provide more personal data. (…) All the contours of the Cambridge Analytica situation are still coming into light. There was clearly a breach of consumer trust and a likely improper transfer of data.” (The Washington Post, 2018, 4:43 – 4:44)

Social subjection as a machine functions with the signifying semiotics and events such as speech and symbols in society. Social subjection semiotics produce a binarity of modes of identity and roles. The translation makes binaries and subjects people and things into dual identities. This produces the individuals according to their respective signification. And so

> “[s]ubjection refers to the transcendence of models into which subjectivities must fit and to which they must conform (man/woman, capitalist/worker, teacher/student, consumer, user, etc.).” (Lazzarato, 2014, p. 31)

Further, this dualistic comprehension follows a certain premise. by re-reading one of Lazzarato's notions of signifying semiotics in capitalism we can move closer to this comprehension of dualistic in social subjection and its symbols:

> "Language, icons, gestures, the language of things (urbanism, commodities, prices, etc.). all semiotics must be compatible with and adapt to the semiotics of capital, especially those having to do with the labor force."
> (Lazzarato, 2014, p. 71)

We will engage signifying semiotic of language as to penetrate the event and investigate the data-trust machinic flow as social subjection.

Zuckerberg's hearing

The fragment from Mr. Grassley's presentation of the case was part of the first speech and how the hearing started. He presented the case, the problem, and the significance through approximately ten minutes. Following the chairmen's speech and the ranking member's speech at the congressional hearing, Mark Zuckerberg would have his initial speech, his time for free words. He initiated his speech with a description of Facebook as a tool for people to interact and connect with each other,

> "But it is clear now that we didn't do enough to prevent these tools from being used for harm as well. And that goes for fake news, for foreign interference in elections, and hate speech, as well as developers and data privacy. We didn't take a broad enough view of our responsibility and that was a big mistake. And it was my mistake. And I'm sorry. I started Facebook, I run it, and I'm responsible for what happens here. (...) But I'm committed to get this right. This includes the responsibility of protecting people's information which we failed to do with Cambridge Analytica." (The Washington Post, 2018, 4:53 – 4:54)

This speech will be our analytic case, and we will in this part focus on two different notions of the speech, “I’m sorry” and “I’m committed to get this right”.

“I’m sorry”

The first section of the speech involves his statement of guilt, as he allegedly didn’t do enough to prevent these tools from being used for harm. But not only so, he continues by placing the responsibility: “and it was my mistake. And “I’m sorry.” (ibid.) The enunciation accomplishes more than speech. By saying, it was my mistake and “I’m sorry” we do not need to question who (or what) is responsible, because Mark Zuckerberg has placed the guilt upon himself through these statements. The speaker assigns himself a role and he assigns his listeners a complementary role. This will be the force of “performative”, “[t]he force of the performative resides in the distribution of “rights” among speakers.” (Lazzarato, 2014, p. 171) Mark Zuckerberg is not only stating that he is responsible and sorry, but he performatively uses the function of language as to produce social obligations.

Social obligations in that he establishes a relation among subject and object. The social obligation establishes a frame of perception of the matter. But in order for the social obligation to exist, it must be acknowledged in the social. The rights, role and identities will have to be acknowledge. Establishing of the social obligation is by language, and with an answer the social obligation will be acknowledge.

Mark Zuckerberg challenges and acts on precisely this, to create a social obligation, to place roles, where *he* is responsible and *he* is sorry and the congress and all Americans watching become identified, but only if it is acknowledged. In the same instance, when Mark Zuckerberg states he is sorry, his speech act requires a certain connection to data-trust as a machine. And this connection to trust requires response from the other, an acknowledgement. Without it, his social obligation becomes irrelevant for, “[f]or the enunciation, “nothing is more terrible than lack of response.”” (ibid, p. 179) For the enunciation the response is a constitutive element of the utterance as it agrees on the terms in the presented speech but it isn't necessarily an agreement on the argument itself.

What is important to notice for Mark Zuckerberg, his utterances of "I'm sorry" are not supposed to ask the audience to trust him, but to get them to agree to rules and categorizations inherent in his speech. For Mark Zuckerberg it is important to get the social setting to acknowledge his social obligation. Zuckerberg's act establishes a formulation of the matter. We can either recognize it and go along with it or deny it altogether. If we answer according to his statement we are acknowledging the social obligation, but we can also build another social obligation.

If the audience accepts or even responds to his speech, then they agree with his establishment of the social obligation – that the fault was his. By this the outcome is Mark Zuckerberg is responsible for what has happened. At the same time, Big Data cannot be blamed. This establishment will subsequently re connect the trust machine, as we can now trust data, but we need to decide on whether we trust Mark Zuckerberg. Mark Zuckerberg creates the distinction between him and Facebook, and ultimately between him and Big Data. We can trust data again, but we may not be able to trust the person.

"I'm committed to get this right"

Now the second tempi of this enunciation concerns Mark Zuckerberg's utterance where he says, "I'm committed to get this right" (The Washington Post, 2018). Now this remark, following the established obligations, put the situation into new light. The speech and in its act of establishing a social obligation challenges to modify the situation. Now, it's not just a question of trusting the establishment of the event, where Mr. Zuckerberg is the one to blame, but he offers the solution. Mr. Zuckerberg is that solution as well, as he is "... committed to get this right" (ibid.). So, what he offers is his commitment to make amends and for him to do better. The listener's tradeoff is data-trust: trust to continue to use Facebook as a social network platform, and to trust that Big Data do not misuse personal data again. This enunciation altogether obligates the listener to take a position, to express by online actions what their stand is. Such a stance can be interpreted as to whether or not Facebook users/the American voters would return and use Facebook as a platform built again. By connecting to Facebook, the user reconnects the data-trust-machine. The enunciation from the Facebook CEO is not only linked to the Facebook data privacy case but may also speak for those data scandals that may come in the future. "I'm committed to change this", is a performative speech-act to shape the event and by that shape the subjectivity of the data users, so as to all listeners must subsequently position themselves with respect to his enunciation and connect to the data-trust-machine yet again. Mark Zuckerberg offers the whole solution for the data scandal if we acknowledge and establish the social obligation.

Sum up

Following the episode that took place, this analysis also "stepped into the light", as the media and news magazine brought it to attention. Along with the congressional hearing of Mark Zuckerberg, this analysis moved from the domain of machinic enslavement to the domain of social subjection. By doing this, the data-trust-machine came into play, as the congressional member, Chairman Grassley, stated that the incident led to a breach in trust between consumers and the Facebook platform. The whole episode was questioning the premise of "data" and "private data security".

By questioning this data and its relation to the data-trust-machine we also looked at a second statement before the congressional hearing which came from Mark Zuckerberg, the man on the stand. Now, his speech involved a clear expression of not only Facebook as a platform, but his speech produced social subjection as well. As he blames himself for the incident he establishes roles and identities in the matter. His enunciation was not only a speech for itself but an action altogether, as he was establishing a social obligation, a relationship to the matter.

Continuing with Mr. Zuckerberg's testimony he did not only offer an establishment of social obligation, but he offered the whole unity of blame and solution. His solution to the event is only obtainable through him. Mr. Zuckerberg, as he is committed to make things right. This will remove data from the question of blame and the question of solution. This altogether will re-create a state of automation by re-enabling the flow of data-trust to data: We may continue using social networks such as Facebook. This may bring us back to our way of living life, bring us back to machinic enslavement with Big Data: A life of expanding our online Facebook profile's content, our online friends, our communication streams, etc.

Subjectivity produced in the case of Facebook and Cambridge Analytica

In this chapter we have analyzed how subjectivity was produces in both spheres of machinic enslavement and social subjection in the case of Facebook and Cambridge Analytica.

In the first part of this analysis, we saw an assemblage in machinic assemblage consisting of Facebook, Cambridge Analytica, the OCEAN model, Facebook users as well as Aleksandr Kogan and his surveys. We have argued that all parts of this assemblage was connected to the expansion-machine of Big Data which produces subjectivity in various ways. We also analyzed how the asignifying semiotic governing the movements of expanding activities in machinic enslavement was data. As we seek to utilize Big Data analytics we are translating the world into data.

We saw that the expansion-machine produced data in form of Facebook becoming a large social platform with lots of personal data stored. The Facebook users expand their personal profile with more personal information as well as connecting with their friends online, datafying themselves and producing subjectivity. These data pools were gathered by Aleksandr Kogan and through his survey he expanded into tapping the survey results, along with the Facebook data on the profile and their friends' data as well. Kogan datafied the Facebook users' profiles and their peer connections. This led to Kogan's large dataset.

Now, we saw expansion flow as Cambridge Analytica found new ways they could use the datasets made by Kogan. This was evident in how they could influence

American voters' political preferences through small advertisements online. The expansion-machine also produced how Cambridge Analytica was able to datafy the American voters by putting them into categories of the OCEAN-model. Furthermore, we also saw it in how Cambridge Analytica was able to influence American Voters in completely new ways through Big Data and target people as individuals. The tool altogether allowed content to take part of the machinic enslavement. This practice is a new production of subjectivity as new flows of influencing persons' preferences was surfaced.

However, this whole assemblage broke down as the activities of Kogan and Cambridge Analytica came into the spotlight of the media and thus user trust to the Facebook platform was broken. This broke down Facebook's connection, along the users', Kogan's, and Cambridge Analytica's connection to the expansion-machine. The breach of user trust signaled a disconnection to the data-trust-machine as well. With the scandal in the spotlight, the domain moved from machinic enslavement to the domain of social subjection.

Zuckerberg's speech at the congressional hearing corresponded to signifying semiotics in the domain of social subjection which produced subjectivity in a number of ways. First we saw how Zuckerberg's utterances, "I'm sorry" and "I'm committed to get this right" established a binary distinction between him and Facebook. We analyzed this distinction as a way of using signifying semiotics that put up the social establishment by blaming Zuckerberg himself. By saying "I'm sorry" and "I'm committed to get this right" Zuckerberg established a social obligation where we are looking away from data, and focusing on him as a person.

Producing subjectivity through social subjection is this act of establishment where a certain social obligation frames the roles, putting Zuckerberg as a central focus and leaving data to itself. If the audience chose to engage on Zuckerberg's terms, production of subjectivity would have to follow the established social obligation. Another notion here is that the audience could also choose not to respond Zuckerberg's speech at all, thus not recognizing his social obligation. This would create subjectivity in how the event then would have framed a social obligation.

Chapter 8 – Conclusion

We have in our thesis asked the questions of how Big Data produces subjectivity and how subjectivity is produced in the case of Facebook and Cambridge Analytica data scandals.

Through Deleuze and Guattari's concept of the machine as well as Lazzarato's social subjection and machinic enslavement, we analyzed various literature and articles on Big Data to examine how subjectivity is produced by Big Data through two machines. The two machines that surfaced was the expansion-machine and the data-trust-machine, where the former is related to the sphere of machinic enslavement while the latter is related to social subjection.

When people engage with Big Data, they inevitably find themselves connected to the expansion-machine. The expansion-machine constitutes flows which produces subjectivity in various ways. In our thesis, we have shown how the expansion-machine enables people to expand the volume of already established data streams, expand their scope by making new data streams through datafication, expand the uses of Big Data, and expanding their competencies into more and more fields of expertise in order to manage Big Data better.

The other machine at play in Big Data is the data-trust-machine. This machine is apparent in how Big Data demands a data-trust from the ones who engage with the results of it and as users supply data-trust when engaging with the social network platform. The data-trust-machine surfaces in how big datasets becomes too large for people to comprehend. In extension to that, Big Data's insights and research are not comprehendible for us either.
So if people want to make use of Big Data, they must trust that which they don't comprehend. Every information a user supplies on Facebook is an act of production of subjectivity, as it produces data. The data-trust-machine simultaneously produces subjectivity when the resulting output demands a choice between trusting Big Data or not.

When we are connected to the expansion-machine we too find ourselves in the realm of machinic enslavement. Here we engage in expanding activity produced by the expansion-machine in an automatic way. This machinic enslavement is the act of datafication, where we follow the asignifying semiotic of data. Data slips past events in machinic enslavement, as we see, interpret, and understand everything through data in this realm. It is already always part of our perception on the matter.

The data-trust-machine functions in the sphere of social subjection. Here we have found that social subjection produces a necessity in language of trusting data or not trusting data. This makes the basis for choices we have to make in relation to Big Data. These choices are the subjectivity which is produced in social subjection. Social subjection usually makes a distinction between using Big Data or not using Big Data with an emphasis on that in order to engage with it, you have to put your trust in Big Data as well. This binary distinction is usually emphasized as choosing to trust Big Data to be both vital for survival and highly beneficial where the opposite leads to undesired consequences.

In the case of the data privacy scandal with Facebook and Cambridge Analytica, Big Data produced subjectivity in both machinic enslavement and social subjection.

Our analysis of machinic enslavement distinguished a production of subjectivity as Aleksandr Kogan established a new approach to creating big datasets on Facebook users' profiles. We examined how the expansion-machine in machinic enslavement produced subjectivity as new methods where allegedly enabling Cambridge Analytica to use Big Data to influence US voters by means of psychology, prior to the US election in 2016.

The production of subjectivity in machinic enslavement broke down, as the incident of Cambridge Analytica's work was brought into light by the media. The event mobilized the social subjection as Mark Zuckerberg, CEO of Facebook, was placed before a congressional hearing. Here Zuckerberg stated, "I'm sorry" and

“I’m committed to get this right”. These two utterances made a distinction between himself and Facebook – only one is to blame. And he proposed the solution to the problem, again appointing himself. Through this binary distinction, he emphasized the blame on himself. These utterances produced subjectivity by framing a social obligation on the listener to make a choice between blaming Zuckerberg or Big Data. If they chose to consider blaming him, then they are already engaged with the social obligation he proposed, as they would agree with the establishment, the framing of the perspective on the problem.

Chapter 9 – New Perspectives

The following sections will delve into some of the analytic findings we've made and discuss outcomes, shortcomings and implications that might follow. The subjectivity produced by the data-trust-machine, the expansion-machine as well as social subjection and machinic enslavement may have consequences for Big Data, the people working with Big Data (directly or indirectly), government, society, education, and organizations.

These themes are meant to form a discussion of our thesis' practical and theoretical implications. In practical themes, we will discuss the question of whether or not we may fall back into machinic enslavement in the aftermath of the scandal of Facebook and Cambridge Analytica. We will also discuss how a growing popularity of data scientists might influence organizational structures and the role of theorists which data scientists may substitute in some cases. Apart from that, we will discuss future social subjection in terms of distinguishing data-driven businesses from non-data-driven businesses and how they might have consequences for business models, the competition, and relationship between those two types of businesses.

Theoretically, we will discuss what implications our research might have in general Big Data and Big Data management literature. Furthermore, also how our analytic notion of the data-trust-machine could enhance M&C's theory on Big Data.

The aftermath of the Facebook scandal

One notion that this project does leave empty regards the outcome of the Facebook case. We saw how machinic enslavement was apparent as users have been engaging with Facebook, uploading photos, interacting on forums, and expanded their data content. As the scandal occurred people became more aware of data security and data altogether. Zuckerberg's act of speech may have enunciated a social obligation, but how will the response be? Will we see the acknowledgement of the social obligation? Will we interact with Facebook once again? Will our data-trust be re-established? Many would probably argue that we are back in machinic enslavement, but even the smallest event has an impact on machinic enslavement. So, how has this event had an impact on machinic enslavement? How has this Facebook scandal had an effect on the machinic enslavement we engage with today? Have we become more aware of data or are we more aware of Mark Zuckerberg and his mistakes, not blaming data or Facebook, but blaming the people standing behind the data platform? This investigation would be beneficial for this project, but we might not be able to do the analysis just yet.

Implications for the data scientist

We have previously found that Big Data's expansion-machine demands that the data scientists always try to further expand their field of technical and academic competency. Granville found that a good data scientist had knowledge within statistics, mathematics, computer science, software engineering, etc.

This expanding trait may be argued to follow the expansion of Big Data and the different tasks it will undertake in the future. We cannot speculate in what kind of new objectives Big Data might have in the future in an organizational context but if they should expand in number, such new tasks may again demand more new knowledge from the data scientist.

A further expansion of the data scientist's competencies may be at a cost. If Big Data's ability to solve different problems across fields in the organization, the field experts which used to deal with such types of topics may become replaced by more data scientists. Rather than having the in-depth knowledge of any such given field the data scientist would have only some knowledge about it which would prove enough to operate the Big Data tools to find better solutions than the field experts could.

If we follow this argument, we might see a future where the expansion-machine's impact on subjectivity in the context of the combinations of people employed in the average data-driven business change. It may change to an environment with many more data scientists and fewer field experts. Each department within the organization may only consist of one field expert while the rest of the team would be data scientists. Such change begs the question: what role will the field expert then have?

If business decisions based on Big Data would expand to more and more departments of the organization (demanded by the expansion-machine), it be could argued that the role of the field expert would become one of aiding the data scientists rather than leading them. Aiding them in the sense of making sure that the Big Data tools operated on the proper theoretical foundations – making sure that Big Data looked in the right directions.

If such proposition would take place, it would imply new ways of approaching work, organizational structures, and education. If data scientists will become abundant, this could mean that universities would have more programs where data management would be a part of it. Also, we could see the theoretical and specific academic courses become fewer and rarer. This could also change the public and corporal perception and value of the theoretician. Can we expand the Big Data tool and the data analyst into every scientific discipline? Following our expansion logic this will soon be tested.

Data-driven companies who will expand their hiring of data scientists and expand the data scientists' roles in the company may restructure the departments to accommodate such expansion. An accommodation like that could be in form of making their departments overlap according to the data scientists' general expertizes in many fields expands. Thus anticipating they will move easily from department to department.

Moving back to focusing on the data science profession. Following the expansion-machine, the data scientist's expansion of skills could also pose a challenge in education. If universities would begin to offer programs in data science in the way we have investigated it in our thesis, you could argue that such programs over time would become more and more broad encompassing more and more disciplines for the students to learn. Thus, a challenge may surface, as 'data science' will not be about data science as such but rather a myriad of other things.

Another thing to consider in this regard is how Maury in M&C's book became a data scientist. He had some knowledge about seafaring and the weather and gathered knowledge from other specialist in related fields and he fairly quickly became a data scientist. In this there may reside a theoretical challenge for governments, corporations, and educational institutions in defining what the data scientist is. What will be the description and the requirements for the future data scientist?

Sharper distinctions between data-driven and non-data-driven

We looked in our thesis at how leading management consultancy firms used social subjection to subject companies to partake the identity and practice of a data-driven company. By that, they subjected them to trust Big Data, to connect to the data-trust-machine.

One argument made by the consulting companies resonated with all: becoming a data-driven corporation reaps more and more benefits along with increased performance. This way of doing business might set off a new segregation of markets: those who are using Big Data and those who don't. The divide between the two types of doing business may become greater and greater. This might create a ripple effect influencing politics, education, etc.

If the use of Big Data will enable data-driven companies to become much more efficient and generate much more value than their counterparts, we may see that consultancy firms like BCG, McKinsey, and KPMG would push social subjection in a direr fashion. The divide in value between using Big Data and not using Big Data may become greater and greater as the Big Data tools expand and can yield more uses and more efficiency. If such development takes place, the dramatic distinctions between being data-driven and not drawn by these consultancy firms might become more intense than they are now.

Consultancy companies are speaking of Big Data and asking companies to comply to it or fall behind their competitors. This might lead to a segregation of the type of company altogether. Social subjection might form a binary understanding of the companies: dividing the companies of Big Data from all other companies. And might change the perception of companies altogether.

Such change in social subjection might make the question of connecting to the data-trust-machine somehow easier but it also intensifies a feeling of emergency. Big data is here and companies have to react according to it. The use of data might even expand to a point where data becomes dire for new start-ups. In such a reality, entrepreneurs with new businesses may have to purchase datasets from large data-driven companies that gather data on everything they can (presumably Google, Facebook, Amazon, etc.). This, making the process of starting a new company much different than it is today.

This divide between data-driven companies and non-data-driven may cause non-data-driven companies to focus on theory rather than data, thus focusing on hiring theoretical specialists. It may even be so that non-data-driven companies will become smaller in size corresponding to the number of theoretical specialists available and the demand for such. Such companies may be more prone to offer consultancy services for data-driven companies in helping them adjust the Big Data tools to search for the right tendencies – offering the theoretical foundations that Big Data needs. Likewise, today most large companies use both in-house and external consultants from consultancy firms and it may become the same way where data scientists stand as the normal employee and theoretical specialists will act more like consultants.

Effect on data management literature

We have now discussed some practical implications of our study on work, business, and society. Now we will move towards looking at how our findings can influence data management theory.

The notion of the expansion-machine is, as we have argued, one which enables growth in many different directions. It is that logic we follow when we find ourselves working with in machinic enslavement. Our conception of expansion may be seen as having worrying implication from a data management perspective. The assemblage wherein Big Data, the organization/company, and employees find themselves within might be perceived as one with too much growth for what is healthy for the company.

Scholars might emphasize the importance of having breakdowns now and then to make sure that the company's structure, strategy, mission, and vision stays intact. Scholars might argue that the expansion-machine in Big Data leads to unstable businesses that are led by the unlimited possibilities of Big Data rather than their own strategy and ideals. Thus, not to let Big Data control the company but making sure that the company controls their Big Data tools.

The cycle of moving from machinic enslavement to social subjection and back again can also set the theoretical foundation for how companies can manage breaking their employees' machinic enslavement and thus come out of the mode of ever more expansion of Big Data uses. However, it may not be done through a breaking of data-trust but rather some sort of healthy guidelines for data management that employees and management have to apply to and frequently check to see if they live up to them.

Data-trust to Big Data

In M&C's book *Big Data* which has been our primary object of analysis, we find a great emphasis on the transition from our usual way of reasoning to one focused on Big Data's reasoning. To shortly summarize the distinction between the two ways of reasoning, causational reasoning is the way people are used to make sense of the world. Causational reasoning is characterized with finding explanations for why something reacts to something else. Big Data's reasoning is only concerned with finding probability and tendencies between different phenomena. Such reasoning can thus link phenomena together but can never give an explanation as to why the tendency is there. This lack of explanation is something that people cannot understand as such – however, M&C argue that to understand the tendencies is now irrelevant due to the high value Big Data and its findings can yield.

However, from our analysis of how Big Data produces subjectivity, we argue that to accept Big Data tools and reasoning is not as simple as M&C may do. In our thesis we have showed the users of Big Data perform a leap of faith to trust data.

Thus, to see the benefits from Big Data through its utilization and the expansion of its uses it requires a trust in the way data works. It may be quite significant to make this trust connection. Even more so for an entire organization to shift collective mindset to connect with Big Data and leave old practices behind. How can a company manage to do such? How can an individual make the leap of faith and connect to the data-trust-machine? Do we need guiding principles?

There is also the question of data-trust breaking down from time to time. We exemplified in the case of Facebook and Cambridge Analytica how a breach in data-trust leads to an exit from machinic enslavement and made the users of Facebook question the assemblage of Big Data they were engaging with.

Though, the case of Facebook and Cambridge Analytica was one example of a breakdown in the data-trust-machine, we might argue that similar breakdowns may happen in organizational contexts. An event such as mistakes in the programming of the algorithms which would alter the Big Data tools and could spur the question of how subjects within the organization may respond to such. If the Big Data tools find tendencies from which decisions are made that yield little results, how long will it take before a breakdown in the data-trust-machine will occur? When will people recognize that the entity they trusted is not functioning correctly? And when it happens how will the company go about reinstituting the employees' connection to the data-trust-machine again? And more importantly, how will organizations deal with such breakdowns which may happen from time to time? What kinds of management practices could prove themselves effective in dealing with such breakdowns? Would such practices take the form of social subjection by creating a narrative about the employees and the company and their role in relation to Big Data? We could also pose the question of how organizations could prevent breakdowns in the data-trust-machine.

So, the theory on Big Data and the shift to Big Data could take our findings into account and further investigate what problems and challenges may lie in such a transition. Such an investigation could take the form of further theoretical inquiries into what the two machines in Big Data as well as the cycle might entail for the transition and how it may produce subjectivity in certain contexts.

Finally, in M&C's book, we are often presented to a few large data-driven companies (Facebook, Amazon, IBM, etc.) which indirectly are portrayed as the leaders in the field of Big Data. We then suggest that in potential future case studies of large data-drive companies, inspired from M&C's conception of Big Data and the two ways of investigating, such studies could use our concepts of the expansion and data-trust-machine as well as social subjection and machinic enslavement to nuance their research. This could enable an understanding of the transition towards a data-driven company and would prove valuable for companies.

Such investigations may contribute to a further enhancement of M&C's conception of Big Data and may be able to conceptualize further theory on how to make the transition from our usual way of reasoning to Big Data's reasoning successfully.

Bibliography

Books

Alvesson, M., & Sköldberg, K. (2009). *Reflexive Methodology: New Vistas for Qualitative Research: New Vistas in Qualitative Research* (2nd ed.). London: SAGE Publications.

Atkins, K. (2005). *Self and Subjectivity*. Oxford: Blackwell Publishing.

Buchanan, I. (2008). *Deleuze and Guattari's 'Anti-Oedipus': A Reader's Guide*. London: Continuum International Publishing Group.

Colebrook, C. (2002). *Gilles Deleuze*. London: Routledge.

Deleuze, G., & Guattari, F. (2009). *Anti-Oedipus: Capitalism and Schizophrenia*. London: Penguin Group.

Granville, V. (2014). *Developing Analytic Talent: Becoming a Data Scientist*. Indianapolis: John Wiley and Sons.

Haraway, D. (1991). *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.

Lazzarato, M. (2014). *Signs and Machines*. South Pasadena, CA: Semiotext(e).

May, T. (2005). *Gilles Deleuze: An Introduction*. Cambridge: Cambridge University Press.

Mayer-Schönberger, V., & Cukier, K. (2017). *Big Data: The Essential Guide to Work, Life and Learning in the Age of Insight* (Rev. ed.). London: John Murray.

Parr, A. (2010). *The Deleuze Dictionary: Revised Edition* (Rev. ed.). Edinburgh: Edinburgh University Press.

Internet sources

Allen, N. L. (2015, February 15). Study: Data Science Curricula Surges. Retrieved May 13, 2018, from https://poetsandquants.com/2017/02/15/study-data-science-curriculum-continues-surge/

Associated Press in Washington. (2018, April 4). Mark Zuckerberg will testify before House panel on Facebook's use of data. Retrieved May 14, 2018, from https://www.theguardian.com/us-news/2018/apr/04/mark-zuckerberg-facebook-testify-data-cambridge-analytica-house

Barton, D., & Court, D. (2013, March). Three keys to building a data-driven strategy. Retrieved May 13, 2018, from https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/three-keys-to-building-a-data-driven-strategy

Boston Consulting Group. (n.d.)a. What Defines Big Data Success? Retrieved May 13, 2018, from https://www.bcg.com/capabilities/big-data-advanced-analytics/insights.aspx

Boston Consulting Group. (n.d.)b. Big Data & Advanced Analytics. Retrieved May 13, 2018, from https://www.bcg.com/capabilities/big-data-advanced-analytics/publications.aspx

Boston Consulting Group. (n.d.)c. Big Data & Advanced Analytics: Strategic Assets for Every Industry. Retrieved May 14, 2018, from https://www.bcg.com/capabilities/big-data-advanced-analytics/default.aspx

Bowcott, O., & Hern, A. (2018, April 10). Facebook and Cambridge Analytica face class action lawsuit. Retrieved May 13, 2018, from https://www.theguardian.com/news/2018/apr/10/cambridge-analytica-and-facebook-face-class-action-lawsuit

Cadwalladr, C. (2018, April 4). ‘I made Steve Bannon’s psychological warfare tool’: meet the data war whistleblower. Retrieved May 14, 2018, from https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump

Cadwalladr, C., & Graham-Harrison, E. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. Retrieved May 13, 2018, from https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election?CMP=share_btn_tw

Collins, K., & Dance, G. J. X. (2018, March 20). How Researchers Learned to Use Facebook ‘Likes’ to Sway Your Thinking. Retrieved May 13, 2018, from https://www.nytimes.com/2018/03/20/technology/facebook-cambridge-behavior-model.html

Davies, H. (2015, December 11). Ted Cruz using firm that harvested data on millions of unwitting Facebook users. Retrieved May 13, 2018, from https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data

Dreischmeier, R., Close, K., & Trichet, P. (2015, March 2). The Digital Imperative. Retrieved May 13, 2018, from https://www.bcg.com/publications/2015/digital-imperative.aspx

Griffin, A. (2018, April 11). Zuckerberg Congress Hearing: The Most Excruciating Moments From The Facebook Boss' Questioning. Retrieved May 13, 2018, from https://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-mark-zuckerberg-hearing-congress-moments-whatsapp-twitter-a8299511.html

Halevi, G. (2012, September). The Evolution of Big Data as a Research and Scientific Topic: Overview of the Literature. Retrieved May 13, 2018, from https://www.researchtrends.com/issue-30-september-2012/the-evolution-of-big-data-as-a-research-and-scientific-topic-overview-of-the-literature/

Henke, N., Libarikian, A., & Wiseman, B. (2016, October). Straight talk about Big Data. Retrieved May 13, 2018, from https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/straight-talk-about-big-data

Jacobson, R. (2013, April 24). 2.5 quintillion bytes of data created every day. How does CPG & Retail manage it? Retrieved May 14, 2018, from https://www.ibm.com/blogs/insights-on-business/consumer-products/2-5-quintillion-bytes-of-data-created-every-day-how-does-cpg-retail-manage-it/

KPMG. (n.d.). Big Data. Retrieved May 13, 2018, from https://home.kpmg.com/be/en/home/services/advisory/technology-advisory/data---insights/big-data-.html

KPMG. (2015). Data-driven business transformation: Driving performance, strategy and decision making. Retrieved May 13, 2018, from https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2017/01/data-driven-business-transformation-final.pdf

McKinsey & Company. (n.d.). [Overview of Big Data analytics related articles]. Retrieved May 13, 2018, from https://www.mckinsey.com/quarterly/overview/big-data-analytics

Oxford Dictionary. (n.d.)a. Algorithm. Retrieved May 13, 2018, from https://en.oxforddictionaries.com/definition/algorithm

Oxford Dictionary. (n.d.)b. Data. Retrieved May 13, 2018, from https://en.oxforddictionaries.com/definition/data

Stahl, L. (2018, April 22). Aleksandr Kogan: The link between Cambridge Analytica and Facebook. Retrieved May 13, 2018, from https://www.cbsnews.com/news/aleksandr-kogan-the-link-between-cambridge-analytica-and-facebook/

Stallman, R. (2018, April 3). A radical proposal to keep your personal data safe. Retrieved May 13, 2018, from https://www.theguardian.com/commentisfree/2018/apr/03/facebook-abusing-data-law-privacy-big-tech-surveillance

Statista. (n.d.). Number of monthly active Facebook users worldwide as of 1st quarter 2018 (in millions). Retrieved May 13, 2018, from https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

Suffolk University. (n.d.). Big Data and Business Analytics Major. Retrieved May 13, 2018, from http://www.suffolk.edu/business/undergraduate/66667.php

TED. (n.d.). Data: A collection of TED Talks (and more) on the topic of data. Retrieved May 13, 2018, from https://www.ted.com/topics/data

The Washington Post. (2018, April 10). Mark Zuckerberg testifies on Capitol Hill (full Senate hearing) [YouTube video]. Retrieved May 13, 2018, from https://www.youtube.com/watch?v=6ValJMOpt7s&t=10s