Postulating Consumers: How Marketers Conceptualise Consumers in the Era of Big Data Analytics

Roger Ferdinand Francois von Laufenberg

University of St Andrews

This thesis is submitted in partial fulfilment for the degree of

Doctor of Philosophy (PhD)

at the University of St Andrews

June 2020

Candidate's declaration

I, Roger Ferdinand Francois Von Laufenberg, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 73,300 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree.

I was admitted as a research student at the University of St Andrews in January 2017.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

Date 15/06/2020 Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 15/06/2020 Signature of supervisor

Permission for publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Roger Ferdinand Francois Von Laufenberg, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

Date 15/06/2020 Signature of candidate

Date 15/06/2020 Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Roger Ferdinand Francois Von Laufenberg, understand that by declaring that I have original research data or digital outputs, I should make every effort in meeting the University's and research funders' requirements on the deposit and sharing of research data or research digital outputs.

Date 14/06/2020 Signature of candidate

Permission for publication of underpinning research data or digital outputs

We understand that for any original research data or digital outputs which are deposited, we are giving permission for them to be made available for use in accordance with the requirements of the University and research funders, for the time being in force.

We also understand that the title and the description will be published, and that the underpinning research data or digital outputs will be electronically accessible for use in accordance with the license specified at the point of deposit, unless exempt by award of an embargo as requested below.

The following is an agreed request by candidate and supervisor regarding the publication of underpinning research data or digital outputs:

No embargo on underpinning research data or digital outputs.

Date 14/06/2020 Signature of candidate

Date 15/06/2020 Signature of supervisor

General acknowledgements

First, I want to thank Kirstie Ball, my PhD supervisor, who has been a tremendous support throughout the entire length of this project. By providing me with valuable feedback, assisting me with my questions while at the same time enabling me to grow as a researcher throughout this period, she has helped me immensely. Along the way, I have learned a lot more than just the skills of how to write a PhD thesis, which I appreciate a lot. Furthermore, I would like to thank the staff at the School of Management for being such a great support and for providing valuable feedback on my research as well as on my teaching. I want to particularly thank Boyka Bratanova, Jeffrey Hughes, Alina Baluch, Shona Russell, Laurence Lasselle, Lorna Stevenson, and Philip Roscoe. I am also thankful for the fabulous work of the School of Management administrative staff, especially by Sam, Dorothy, and Lynn. As for my PhD colleagues, I want to thank all of you, as it has been a joy working in the same office and sharing thoughts about research and the world. Here in particular, I want to thank Toma, Alex, Kirsty, Jasmin, Andreea, and Matthew, who made me feel welcome from the beginning and who are all much more than just colleagues, but rather amazing friends.

Outside of St Andrews, I want to thank first and foremost all my participants who I cannot name, as without their assistance, this thesis would not exist. I am sincerely grateful for their time and insight. I would also like to thank everyone who I have encountered during this time and who have been a great inspiration, in particular the amazing research communities I am now part of, in the fields of Surveillance Studies, and in the Science and Technology Studies, as well as my former and new colleagues at VICESSE.

To all my friends and family: you have been an amazing support! My mom and my sisters for always believing in me, despite not always understanding what I am doing. Chris, Julia, Raphi, Kathi, Andi, and Lisi for the past and future great times together, be that in Scotland, Vienna or somewhere else in the world – thank you. Vera for being such a great friend as well as an amazing academic inspiration – you never fail to amaze.

And most importantly, Verena. I am not sure if you realise the amount of support, care and love you are providing. Without you, this PhD would not have been feasible. Thank you!

Funding

This work was supported by the Social Sciences and Humanities Research Council of Canada through the Big Data Surveillance Project and the University of St Andrews.

Research Data/Digital Outputs access statement

Research data underpinning this thesis are available at https://doi.org/10.17630/af965a99-1c23-4a3d-a49d-f2b54529fcca

Abstract

The proliferation of big data analytics in marketing appears to be having significant effects on the field, such as changing how marketers perceive their consumers and how they act on them. As I discuss in my research, marketers are not satisfied to work solely with approximate, imagined conceptualisations of consumers as a basis for advertisements and offers. Instead, they are looking for exact virtual data doubles of existing and potential consumers, which is something they hope to achieve through big data analytics. In my thesis, I explore the question of how and why marketers conceptualise consumers differently when using big data analytics compared with traditional market and consumer research methods. This is embedded in the theory of the co-production of knowledge and empirically relies on interviews with marketers and data analysts, case studies, and participant observations at industry conferences.

In my research, I show to what extent the idea of the data double consumer conceptualisation is considered an ideal case for marketers, and that it is believed to be made possible through big data analytics, which is expected to create an exact knowledge about consumers. However, my findings show that in practice, big data analytics should be considered a sociotechnical assemblage that produces knowledge which contains inaccuracies, errors and uncertainties. Knowledge about consumers is not just discovered – neither through traditional market and consumer research methods nor through big data analytics. Instead, it is the outcome of a co-production that involves different steps, individuals, teams, normativities, and technologies. Hence, knowledge about consumers is never an exact representation of reality, irrespective of its methods of production. Consequently, consumer conceptualisations expected to be exact data doubles cannot be attained. Instead, postulations are established that are believed to be accurate, without having actual proof. Yet, my findings show that knowledge resulting from big data analytics has a higher credibility and epistemic authority amongst the participants, explaining the persistence of the data double consumer conceptualisation in digital marketing.

Table of Contents

List of Tables

List of Abbreviations

AI	Artificial Intelligence
ANT	Actor-Network Theory
B2B	Business to Business
BDS	Big Data Surveillance
CAQDAS	Computer Assisted Qualitative Data Analysis Software
CEO	Chief Executive Officer
COO	Chief Operating Officer
CRM	Customer Relationship Management
CTO	Chief Technology Officer
DMP	Data Management Platform
FATML	Fairness, Accountability and Transparency of Machine Learning
FCE	Field-Configuring Events
GDPR	General Data Protection Regulation
KPI	Key Performance Indicators
LGBT	Lesbian, Gay, Bisexual, and Transgender
ML	Machine Learning
P&L	Profit and Loss
R&D	Research and Development
ROI	Return on Investment
SSK	Sociology of Scientific Knowledge
STS	Science and Technology Studies
UTREC	University Teaching and Research Ethics Committee

1. Introduction

This thesis explores the effect of marketers' increasing reliance on digital consumer data and big data analytics, and their use of this datafied knowledge to conceptualise consumers. It establishes that big data analytics as a means of creating knowledge about consumers is perceived as more credible and accurate by marketers, in comparison with more traditional market and consumer research methods of knowledge production, particularly in comparison with qualitative methods, such as focus groups. Marketers are under the impression that consumer conceptualisations through traditional means are mere imaginations, approximate concepts of groups of consumers. In contrast, consumer conceptualisations resulting from big data analytics are perceived as exact copies of the consumers in the form of virtual "data doubles" (Haggerty and Ericson, 2000). The main question I explore throughout my research is how in practice marketers conceptualise consumers differently through big data analytics, and why they do so.

The proliferation of digital marketing goes hand in hand with the rapid technological development of the recent decades. A combination of an increase in computing power (Zhang et al., 2018), data storage constantly expanding while at the same time becoming less expensive (Harvey, 2019), and the development of artificial intelligence (AI) and machine learning (ML) algorithms (Günther et al., 2017) has resulted in complex statistical analysis becoming increasingly feasible and more precise (Côrte-Real et al., 2017). The use of (digital) consumer data for marketing purposes is certainly not a new development. Consumer data collection and analysis have a long history in the finance industry, in which credit rating systems have been in place since the 19th century. These systems have been boosted considerably through the implementation of digital repositories and computer systems since the 1950s (Lauer, 2010). Loyalty schemes are another example of a means of collecting vital consumer information, such as individualised shopping and consumption behaviour, which led to an early form of one-on-one marketing and individualised consumer profiles (Pridmore, 2010; Pridmore and Lyon, 2011).

The promises behind these technologies are extensive and offer precise profiling of consumers in a way that has not been possible before (Hofacker et al., 2016; Kosinski et al., 2013; Matz and Netzer, 2017; Netzer et al., 2019). The aim is to create exact consumer

profiles, or virtual data doubles (Haggerty and Ericson, 2000), for marketing purposes, with the potential of predicting exactly how consumers will behave in the future. These consumer profiles, in the form of data doubles, hold the promise that offers and advertising can be targeted in such a way that the probability of the consumer buying or converting is close to certain. At least, that is what is to be believed. Certainly, when looking at the big e-commerce corporations, for example Amazon or Alibaba, and their recommendation systems, or Google's online advertisement platform *Google Ads*, there is little doubt about the possibilities tendered by these operations.

The promises and expectations that emerge in the mainstream big data analytics and marketing literature are rarely addressed critically from within that strand of literature. Instead, the fields of Critical Data Studies and Surveillance Studies have raised considerable doubt regarding the promise of creating exact knowledge through big data analytics. While José van Dijck (2014) has termed these expectations and promises the ideology of dataism, a belief in the potential of objectively quantifying and tracking all aspects of human behaviour, boyd and Crawford (2012, p. 663) have criticised the "mythology" of big data analytics, which is said to create a "higher form of (…) knowledge". Both criticise the same point, namely that even though big data analytics promises objective, true and infallible knowledge, in practice it will never be able to provide that. Big data analytics always operates within a certain area of probabilities and accuracies (Symons and Alvarado, 2016) and thus will always experience errors and produce inaccuracies (Busch, 2014). Producing an exact truth is not possible, nor is producing a virtual data double which factually represents the real consumer. Both will always be an approximation of "reality". Thus, if consumer conceptualisations in the form of data doubles are used in practice based on the premise of being factual individual consumer profiles, marketers rely on a fallacy. The data double consumer conceptualisation turns into a postulated consumer conceptualisation, believed or assumed to be true without any evidence.

However, as little is known about how widespread big data analytics practices are amongst different organisations, and how these operations exactly play out in practice, there are several questions which have emerged and which drive the interest of this research. So far, empirical research on the practice of big data analytics for consumer research and marketing has been sparse. Some organisational research has been done on

the implementation of big data analytics and potential difficulties which emerge in the field of marketing (Quinn et al., 2016). Pridmore and Hämäläinen (2017) have done some research on digital marketing and consumer segmentation. And, while the results of their research showed that there were still many limitations to the use of big data analytics in practice and that old segmentation practices would persist, the rapid development of technology might have changed this in the meantime. Certainly, the complicated and messy technology of big data analytics was considered to be the main obstacle in the studies of both Quinn et al. (2016) as well as Pridmore and Hämäläinen (2017). Apart from certain exceptions (Ariztía, 2018), there has been little research on how big data analytics in marketing is actually used in practice, marking a gap in studies on the practical issues of big data analytics and consumer surveillance.

Additional research in this area, however, could be vital and make a fruitful contribution. For example, there are many aspects which have a significant influence on the practices of big data analytics and consumer conceptualisations, which are at the core of marketing and consumer profiling. These aspects include the complex social, organisational, material, and political relations in which these practices are embedded. But this also involves the realities that are enacted through these practices and which have an impact on and are performative on the recipients – in this case, the consumers (Callon, 2006; Cochoy, 1998). Within the scope of my thesis and my research interests, this specifically involves research into one of the practices that is at the core of modern consumer surveillance – to be more exact, the practice of using big data technologies to analyse consumer data and to create datafied knowledge about these consumers, which are then used as a means of creating consumer conceptualisations – profiles, predictions, and the personalisation of the consumer for marketing measures.

In particular, the creation of data double consumer conceptualisations based on the ideology of dataism is a prevalent expectation in the mainstream big data analytics and marketing literature. This is an ideology which might become even more evident amongst practitioners, as the use of big data analytics replaces practices that have been considered inaccurate, slow and unreliable – such as the more traditional forms of consumer research. This increasingly raises the potential of marketers relying on the false premise of a data double, which is a postulated consumer conceptualisation.

1.1. Research question, theoretical approach, and methodology

My research thus seeks to investigate how the practices of marketing and consumer surveillance – and, more specifically, of creating consumer-oriented knowledge and establishing consumer conceptualisations – change through using big data analytics and how the practice of conceptualising consumers is affected by the ideology of dataism. Or, to phrase it more concisely in a research question, my thesis asks:

> How and why do marketers conceptualise consumers differently when using big data analytics in comparison with traditional market and consumer research methods?

An entry point to the research of these surveillant practices in marketing can be found in Critical Marketing Studies, in which researchers have turned their gaze "*inside* marketing" (Zwick and Cayla, 2011, p. 7) to better understand its underlying practices. Also using theoretical concepts from Science and Technology Studies (STS), such as Actor-Network Theory (ANT), the cultural materialist strand of Critical Marketing Studies is interested in the processes of marketing in the same way as STS is interested in the processes of doing science. In the process of marketing, Critical Marketing Studies is particularly focused on the performativity of markets and marketing. The concept of performativity means that theories about the market, knowledge about the market and consumers, as well as the actors that are involved in the market and its practices – subjects and objects alike – all actively shape how the markets and marketing function (Callon, 2006; Roscoe, 2016). From my research perspective, particular interest is focused on the performative processes of producing knowledge about consumers, as well as the use of this knowledge for the conceptualisation of consumers – traditionally in the form of segmentation which nowadays occurs in the form of individualised data profiles.

Due to my focus on the production of knowledge in marketing, particularly through the use of big data analytics, and, thus, being interested in the practices and methods of doing marketing as well as doing science, my thesis is primarily embedded in the theories of STS. Particularly, as "STS suggests that methods are never simply techniques", it becomes obvious that the methods of big data analytics "are materially complex and performative webs of practice that imply particular arrays of subjects, objects, expressions or representations, imaginaries, metaphysical assumptions, normativities,

and institutions" (Law, 2017, p. 47). In order to analyse the complex "webs of practice" of big data analytics as a method for producing datafied consumer-oriented knowledge and as an important part in the conceptualisation of consumers, I specifically rely on Jasanoff's (2006b) concept of the co-production of knowledge. The co-production of knowledge, on the one hand, directs its focus to the social, cognitive, material, and normative procedures behind the practices of producing knowledge. On the other hand, it further extends its focus to the translation and stabilisation of that knowledge and thus its enactment in practice.

The research question and the theoretical approach will serve as a solid framework to investigate how and why big data analytics influences the practice of conceptualising consumers, through which marketers enact the datafied knowledge they possess about consumers within marketing measures. Empirically, this will be researched through a triangulation of data collection methods and data sources. The core of my empirical data consists of organisational case studies with practitioners as key-informants, i.e. data analysts as well as marketers. These are supplemented with single key-informant interviews, again conducted with practitioners as interviewees. Additionally, participant observations at industry conferences and exhibitions – as Field-Configuring Events (FCEs) – provide an extra layer of insight into the practices of big data analytics in marketing.

1.2. Thesis outline

The following Chapter 2 will expand on the issues I have laid out here and will further address how big data analytics in marketing promises the creation of data double consumer conceptualisations, which should rather be perceived as postulations. In Chapter 3, the theoretical concepts of Critical Marketing Studies as well as those of the co-production of knowledge will be explained in detail. In that chapter, I will also point out how these concepts contribute to answering my research question. From there, Chapter 4 will discuss the philosophical perspectives with regard to the methodology used, and will detail my data sources, data collection methods as well as my data analysis.

The findings of my research will then be presented in Chapters 5 and 6. Addressing the question of *how* consumers are conceptualised differently, I will explain the different practices of conceptualising consumers, as described by the participants, in Chapter 5.

While imagined consumer conceptualisations are produced through the use of traditional market and consumer research methods, the implementation of big data analytics leads to the emergence of data double consumer conceptualisations, which are expected to factually represent the consumer. However, as big data analytics will never have the ability to produce such a factual representation of the real consumer, these conceptualisations are rather postulated consumer conceptualisations. In the second part covering my findings – Chapter 6 – I will explain *why* these different consumer conceptualisations emerge, and *why* participants are under the impression that they can rely on actual data doubles. By detailing the practice of co-producing knowledge, I can demonstrate, on the one hand, how datafied knowledge is deeply embedded in social, cognitive, material, and normative processes. On the other hand, I am able to show how uncertainties and errors of datafied knowledge are omitted from the stabilisation process of knowledge, in which only the advantages of big data analytics remain: advantages that are constantly being considered in relation to the disadvantages of traditional market and consumer research methods.

In the final chapter, Chapter 7, I will refer to my research question and discuss how my findings can answer the research question. There, I will also consider my contribution to the research fields of Science and Technology Studies, Critical Marketing Studies, and, importantly, Surveillance Studies. I will then reflect upon the limitations of my study and present suggestions for future research opportunities before concluding with some final thoughts.

2. Big data analytics in marketing – the hype and its critics

2.1. Introduction

The topic of big data analytics is not new. Recent years have seen an increase in the popularity of big data applications in all sectors – from governmental sectors, such as healthcare, policing, intelligence and justice, to private sectors, such as research and development, human resources or marketing. This research focuses on the application of big data analytics in marketing. In this domain, true datafication has taken place, which has led to the "transformation of social action into online quantified data, thus allowing for real-time tracking and predictive analysis" of consumers (van Dijck, 2014, p. 198). Since consumer data is increasingly within the reach of private companies and organisations, it can be used for in-depth analysis of consumer behaviour and allows companies to profile and segment their consumers in greater detail (Ball, 2017). Equally, the increase in computing power and the development of new analytical capabilities holds the potential of exposing fine-grained insight into consumers, identifying historical, and anticipating future, consumer behaviours. In combination with insights into personal traits and psychological states and emotions, uncovered through psycho-graphic and socio-economic data, this can create substantial and accurate consumer profiles (Hofacker et al., 2016; Kosinski et al., 2013; Matz and Netzer, 2017; Wedel and Kannan, 2016). The goal of big data analytics in marketing is to profile, predict, personalise and target consumers – or, in short, to establish a virtual data double consumer conceptualisation of real consumers (Haggerty and Ericson, 2000).

While this is the theory and the expectation of big data analytics in marketing settings, recently, an array of critical literature has emerged to counter some of the claims that are made in the more mainstream big data analytics literature. The main critical argument is that through the datafication of marketing, big data analytics is largely perceived as a new, all-powerful technology that has the ability to create exact data double consumer conceptualisations (boyd and Crawford, 2012; Crawford et al., 2014; Vaidhyanathan, 2012). This expectation and perception of big data analytics is what van Dijck (2014) considers to be the ideology of dataism, in which people believe that human behaviour and social life can be objectively measured, quantified and classified through data analysis. Furthermore, through the ideology of dataism, van Dijck (2014) also encapsulates the trust of people in the individuals, technologies and institutions that are

responsible for the collection, analysis, interpretation and communication of data and datafied knowledge. This idealised view of big data analytics fails to consider the politics behind doing big data analytics (Jasanoff, 2017), the wide array of subjective interpretations that are involved in the analysis (boyd and Crawford, 2012), the statistical limitations of the analysis (Kaplan et al., 2014; Symons and Alvarado, 2016), the problems with the big data technology (Pasquale, 2015), and, overall, the entire complexity of the sociotechnical assemblage that is big data (Kitchin, 2014).

A similar and equally applicable critique to the use of big data analytics for the conceptualisation of consumers emerges from the sociology of commensuration (Krenn, 2017a; Mennicken and Espeland, 2019). Adding to the above, the critique of the sociology of commensuration specifically addresses the view that in almost every part of society, the perceptions and expectations of apparent 'natural' boundaries of the population and the society persist. These 'natural' boundaries serve as the basis for measurements and classifications. As a result, the classifications of individuals, behaviours and groups are considered to be objective and to accurately represent reality (Bowker and Star, 2000, 2008). However, as elsewhere, the politics, the sociality and the technical, all of which influence how people and consumers are measured, categorised and classified, are largely ignored. In the process of measuring, categorising and classifying, a wide range of decisions are taken that are not only based on natural boundaries but that are also socially constructed (Espeland and Stevens, 1998). This can also be observed in marketing settings, where the apparent natural boundaries of the consumers are used to segment and classify them, often ignoring the sociality and normative politics behind the practice of establishing classes and profiles (Krenn, 2017a). This is a practice that is likely to intensify with the use of big data analytics (Fourcade and Healy, 2017a, 2017b [2013]; Krenn, 2017a).

Based on these problems, in this chapter, I will start by explaining what big data analytics is and how it is used in consumer research and marketing. This will be largely based on the literature, in which expectations are created of the potential of big data analytics to accurately create the consumer conceptualisations of individualised consumers upon which marketers can exactly profile, predict, personalise and target the consumer. This will be followed by a critical examination of these expectations, with a focus on the practice of doing big data analytics (Beer, 2017). Focusing on the practice reveals the

sociotechnical assemblage of these analytics (Andrejevic et al., 2015), as it is influenced by different actors, technologies, and ideologies which affect the datafied knowledge about consumers. Closely connected to this, in the last section, I will critically address the practice of measuring and classifying consumers through the analytics, which is also heavily dependent on actors, power structures, ideologies and technological limitations (Krenn, 2017a).

2.2. Defining big data

The definition of what big data is has seen significant developments over the last two decades. The first mentions of the term *big data* dates back to the late 1990s and was mainly used as a term to describe the processing and visualisation of large data sets (Cox and Ellsworth, 1997). Compared to the definition that I use here, it has become apparent that the term has evolved since its first use. Based on the plethora of different definitions that are used in the literature, particularly in the information technology literature, I define big data as: *the combination of techniques and technologies for the (automated) analysis of exhaustive, complex and unstructured datasets in real-time, with the aim of revealing statistically unidentified connections and patterns in the datasets, creating new knowledge that assists in decision-making* (Hashem et al., 2015; Kitchin, 2014; Kitchin and McArdle, 2016; Laney, 2001). Big data refers to the datasets and technologies, as well as the techniques, of carrying out the analysis. As a result, the term is often used interchangeably with the term *big data analytics*. In my research, I focus specifically on the practice of analysing big data, which is why I will be using the term *big data analytics* throughout my thesis.

The definition of big data analytics has changed over the last few years, going hand-in-hand with the technological development and the expansion of its analytical capabilities. An early-stage definition of big data analytics, as we know it today, was provided by Laney (2001), who defined it as being based on three dimensions, all starting with the letter *V*, prompting it to be dubbed the three V's of big data. This definition is interesting and important as it encapsulates the essence of what constitutes big data. According to Laney (2001), the defining dimensions of big data are the *volume*, *velocity* and *variety* of the data that is being used.

The most obvious dimension of big data – as is hinted by its name – is the high *volume* of data, implying that substantial amounts of data are processed and analysed. What is considered a high volume of data often depends on factors such as what type of data is being processed: e.g. numerical or video data. But this notion of high volume has changed over time. As the number of internet users is continuously increasing, having reached an estimated 4.3 billion users at the end of 2019, this means that the amount of digital data that is being produced and processed is also increasing (ITU, 2019). Furthermore, the rapid expansion of storage capacity and processing power, with simultaneously diminishing costs, increases the magnitude of data and changes the notion of what is considered high volume (Bello-Orgaz et al., 2016; Gandomi and Haider, 2015; Kambatla et al., 2014). For example, in 2010, the amount of digital data created per year was considered to be around one zettabyte. This year, in 2020, it is expected to be around 44 zettabytes of data per year (Desjardins, 2019; Tien, 2013).

Data size or volume is, however, not the sole dimension used by Laney (2001) for defining big data. High *velocity* of data is another defining attribute, meaning that the data is often created in (or near) real-time and analysed in a timely manner. The high velocity of big data often requires multiple individuals and processes being involved in the practices of data processing and analysis (Kitchin and McArdle, 2016). The velocity of data, furthermore, means that data is collected at different instances and by different organisations and then combined to be analysed, subsequently, by other organisations (Matzner, 2016). The content of the data is also constantly changing: complementary data collections are aggregated with existing data collections to provide more comprehensive and accurate analyses (Laney, 2001). This also requires increasingly improved data-processing, data-manipulating and data-analytical methods, often in the forms of algorithms, to cope with this continuous increasing flux of information (Bello-Orgaz et al., 2016).

The wide *variety* of big data as a third defining characteristic refers to the different types of data that are collected, such as textual or numerical data, but also includes videos, images, audio, and metadata which can be provided in a structured or unstructured format (Bello-Orgaz et al., 2016; Laney, 2001; Tien, 2013). The broad variety of data enables a much more fine-grained analysis. However, this often comes at a cost since the data is

generally provided in an unstructured format and requires a major effort in structuring and preparing it for analysis.

The changing definitions of big data in conjunction with the development of its capabilities can be observed with the 3V definition which, over time, has seen further *Vs* being added to it. The additional dimensions which have been proposed are the *veracity*, *variability*, *visualisation*, *validity*, *volatility*, and *value* of data (Degli Esposti, 2014; Ducange et al., 2018; Laney, 2001). The *veracity* of data designates the necessity of awareness about the quality of data, which is instrumental in the accuracy of the analysis. This is sometimes also referred to as the data *variability*, and is closely related to the *validity* of data, indicating the statistical variance which needs to be considered and the use of uncorrupted data. This also includes the notion that the data and the analysis vary in time and through context and are thus situation-dependant (Ducange et al., 2018). The *visualisation* of data refers to the necessity of making data, the analysis and the results comprehensible to the end-users of the analysis, leading to the representation of the data in a visualised manner, such as graphs and tables (Kennedy et al., 2016). The *volatility* addresses the retention periods and erasure of data, which are important as storage capacities can be limited through the size of the data. Finally, the *value* characteristic of big data indicates the extraction of valuable information, as well as their potential economic value. This is often referred to as big data analytics because the extraction of the information and its subsequent use is never given, but is the result of analytical processes (Bello-Orgaz et al., 2016; Ducange et al., 2018; Wang et al., 2016).

Shifting away from the *V* characteristics of big data, Kitchin and McArdle (2016) have carried out extensive research into the characteristics of big data by analysing 26 datasets from different domains that have been labelled as such. They concluded that in many of these datasets, while the *volume* and *variety* of data are often present, they are not considered to be constituting factors of the datasets. Instead, an overarching characteristic of big data is its *exhaustivity*, meaning that, compared to standard datasets, big data is not based on samples but tries to capture the entire research population. The exhaustivity of its data, combined with the *velocity* of collection and the analysis in near real-time, is what defines big data and sets it apart from classic data analytics. Other characteristics of big data analytics that have appeared occasionally in the researched datasets are that they are "fine-grained (in resolution) and uniquely indexical (in identification)" (ibid., p. 1);

relational through multiple datasets being connected; extendible and scalable; as well as having veracity, value and variability of data, as seen above.

Apart from focusing on the characteristics of the big data sets, there is merit in looking further into the capabilities of big data and the underlying processes, to understand and define what it is and does. For example, Brock and Khan (2017) discuss in their paper the possibility of analysing data in near real time – which still relates to the high velocity of data – through the use of predictive algorithms for the use of decision-making. The authors, furthermore, consider big data able to radically challenge typical information technologies that process and analyse data as they do not have the capabilities to deal with the big data datasets. A final point here is that, according to the authors, the term big data also includes all the technology or the infrastructure required to handle and analyse the data. As others also have implied, big data analytics is strongly dependent on the technology in use.

Hashem et al. (2015), for example, discuss how cloud or distributed computing is a technology which enables many of the operations that are necessary for big data analytics, such as the processing and storage of the high volumes of data. Cloud computing is key to big data analytics as it does not require individual entities or organisations to install and maintain the technological hardware infrastructures that are necessary. Instead, these infrastructures are combined, and outsourced to external providers, turning big data from a technology into a service. Additionally, the systems needed to analyse big data can also be used to specify big data analytics, as well as to differentiate the process from that of classical data analytics. These can, for example, include data management systems, which have become progressively more complicated through the nature of big data, or the statistical calculation and the modelling, which are at the basis of big data analytics. One further consideration is that the inclusion of machine learning algorithms has become an increasingly important factor in big data analytics. As the analysis becomes more and more automated, ML algorithms are taking over the role of analysing the data and producing datafied knowledge (Thompson, 2019).

Finally, the definition of big data can be further expanded by looking at the processes that are enabled through big data analytics. For example, Wang et al. (2018) have emphasised the analytical capabilities of big data as an additional component in its definition. These

capabilities range from descriptive to predictive analysis of large, complex and unstructured datasets. Importantly, these types of analyses are not feasible with classical data analytics, which are severely constrained in their capabilities in comparison to big data analytics. All these examples have contributed to the proposed definition of big data analytics as: *the combination of techniques and technologies for the (automated) analysis of exhaustive, complex and unstructured datasets in real-time, with the aim of revealing statistically unidentified connections which assist in decision-making*.

2.3. The big data promise in marketing

Big data analytics is used in many different settings and disciplines. Examples can be found in the governmental and public sector, such as in urban planning, through the Smart City (Coletta and Kitchin, 2017) or in national statistics (Kitchin, 2015); for policing and intelligence (Sanders et al., 2015); or in the health services (Raghupathi and Raghupathi, 2014). But big data analytics is also widely used in the private sector, where the main driving point of users is to obtain economic value through data and subsequently to gain an advantage over competitors (Cabrera-Sánchez and Villarejo-Ramos, 2019). In private companies, big data analytics closely relates to business intelligence and analytics and includes the commercial use of big data analysis, but also analysis for other means, such as research, science and technological development, wellbeing, safety and security, etc. (Chen et al., 2012). Côrte-Real et al. (2017), for example, have highlighted how organisational operations can be assisted and improved through big data analytics, which specifically relates to knowledge management and communication in general. The benefit from big data analytics in this area is that it creates organisational agility and efficiency in their operations, which ultimately will lead to an advantage over competitors.

Looking more specifically at the use of big data analysis in consumer research and marketing in organisations, as the main focus of my research, big data analytics has the goal to improve knowledge management and knowledge communication, as identified by Côrte-Real et al. (2017). However, here, specifically, the use of big data analytics lies in gaining a deeper and more exhaustive knowledge about the consumers of the companies. The main promises of big data analytics in consumer research and marketing lie in the knowledge about consumers and, particularly, how this improved knowledge can help to profile, predict, personalise and target them (Hofacker et al., 2016; Moro et al., 2016; Wamba et al., 2017). This includes the potential of uncovering their psychological traits,

states and emotions (Ascarza et al., 2018; Kosinski et al., 2013; Matz and Netzer, 2017; Netzer et al., 2019) and which, consequently, allows for the highly-personalised profiling and targeting of individual consumers, out of which the end goal is value creation (Dekimpe, 2019; Elia et al., 2019; Raguseo, 2018). In the following sections, I will address these promises of profiling, predicting, personalising, and targeting consumers.

2.3.1. Know your consumer – big data profiling and the personalisation of marketing

The core characteristics of big data – the enormous volume of data marketers have at their analytical disposal combined with the wide variety of data being collected, processed and analysed, almost in real-time – means that a precise analysis of the data can create an important economic value for those extracting consumer information (Symons and Alvarado, 2016). Individual companies have at their disposal exhaustive data about their consumers (Elia et al., 2019). As consumers are generating ever more data about their consumption behaviour, as well as about their socio-demographics, the potential for organisations to leverage this data has certainly evolved in recent years (Dekimpe, 2019). In particular, as consumption activities have shifted into online environments, they have become more visible and traceable (Hofacker et al., 2016). Examples of consumption activities include media and entertainment, which are ever more consumed online through news websites, streaming platforms and social media channels. Also, offline consumption is, in many instances, linked to the digital through wearables and activity trackers, or through a constant sharing of consumption situations on social media. Furthermore, the increasing availability of the internet of things, such as Amazon's Alexa or Google's Home, reduces the boundary between offline and online consumption (Shareef et al., 2016). All of these circumstances and tools lead to an extensive stream of data ready to be analysed and to provide a much broader and detailed insight into consumers than was possible before the implementation of big data analytics. The expected knowledge about consumers and other market actors seems beyond what would be possible in a traditional consumer research environment (Hofacker et al., 2016; Shareef et al., 2016).

A crucial aspect of big data analytics in marketing is to better understand consumer needs and behaviour (Yao et al., 2014). In particular, e-commerce organisations and companies can better identify the consumption behaviour of individuals on a very detailed level. As an example, consumer behaviour includes the decision-making by the consumer in terms

of why she or he is looking for a specific product or service – out of necessity, desire, through dissatisfaction with another product or service – and how other factors, such as reviews, can influence this behaviour (Hofacker et al., 2016; Tata et al., 2020). Consumer behaviour data can include much more, including items that have been searched for, which search terms were successful in leading a consumer to the site, how long the consumer browsed the products, on which items the consumer clicked or only hovered over with the cursor, which products were added to a shopping cart and deleted again, which products were compared, and more. (Mazzei and Noble, 2017; Shareef et al., 2016; Tzafilkou et al., 2014; Wedel and Kannan, 2016). This online consumer behaviour leaves traces in the form of consumer data and cookies, which are further analysed and used by companies for profiling and targeting (Chester, 2012; Mellet and Beauvisage, 2020).

Big data analytics in market and consumer research can be used not only to analyse consumer behaviour. Often, combining various kinds of information about consumers has immense potential for big data analysis in marketing. For example, merging information about consumer behaviour with the underlying psychological aspects of the consumers, such as their psychological traits and states, can lead to improved consumer segmentation, for which targeting measures can be specifically tailored (Matz and Netzer, 2017; Pridmore and Hämäläinen, 2017). The use of psychological traits for marketing is not a new development. Consumer behaviour research often links the emotional and cognitive acceptance of product or brand marketing by consumers if these are tailored to their own psychological traits (Aaker, 1999; Hirsh et al., 2012; Wheeler et al., 2005). However, the observation, identification or even prediction of these traits, such as personality, political orientation, openness, risk aversion, self-esteem, perfectionism, etc., has been challenging (Matz and Netzer, 2017). Through the analysis of the digital traces left by consumers, this seems to have become considerably easier. Kosinski et al. (2013) have, for example, shown that many personality traits, such as openness, extraversion, satisfaction with life, as well as other attributes, such as sexual orientation, political preference etc., can, with a certain level of accuracy, be automatically deducted through the *Facebook Likes* of individuals. As few as around 30 *Facebook Likes* are sufficient to provide some correlation with selected traits, with increased accuracy when more *Likes* are included in the model. This shows the potential of such analyses and models when used with

additional digital records in predicting consumers' psychological traits and other attributes (Kosinski et al., 2013; Matz and Netzer, 2017).

The sources for these analyses not only include data about past transactions and interactions with consumers, search engine data and browsing history but also a wide variety of social media profiles, as well as smartphone sensor and metadata (Matz and Netzer, 2017). The potential of using psychological consumer traits in marketing lies in the possibility of testing the effectiveness of personalised messages and offers (where and how are they successful and in what does the success result?), or to establish precise psychological profiles of the consumers and combine these with already-existing consumer profiles to tailor the communication strategies based on this information.

Thus, many of the uses of big data analytics in marketing aim to gain better knowledge of the consumer and often focuses on uncovering the behaviour, traits and emotions of individuals. Particularly in consumer research, this is seen as a way of improving existing consumer segmentation that is used in companies. Consumer segmentation groups consumers who have similar characteristics and, thus, requirements, regarding how they should be addressed through marketing measures (Arunachalam and Kumar, 2018). The segmentation of consumers has long relied on datafied information about them found in Customer Relationship Management systems (CRM) (Pridmore and Hämäläinen, 2017). Such information may also include knowledge gained through more traditional market and consumer research methods, such as focus groups (Grandclément and Gaglio, 2011). Expert knowledge is also considered to be a contributing factor when creating consumer segmentation (Dubuisson-Quellier, 2010). Still, due to the nature of the knowledge used, based on qualitative research methods, such segmentation is often considered only to be an imaginary conceptualisation of consumers (Ariztía, 2015).

Big data analytics is considered a significant improvement in consumer segmentation practices as the technology can rely on clustering algorithms, which particularly improve the stability of the segments. Big data analytics increases the validity of consumer segments, which means that they represent what they should represent, and continue to do so over time (Arunachalam and Kumar, 2018; Müller and Hamm, 2014). Furthermore, big data analytics has been shown to be useful in providing evidence for the success of marketing campaigns on individual segments (Yao et al., 2014). The increased stability

of consumer segmentation through big data analytics allows for an even more fine-grained grouping of consumers, based on the defining similarities. Nonetheless, ultimately, the aim of big data analytics in consumer research and marketing is to provide a level of detail that goes down to the "segment of one" (Moran, 2014), or what others would call individual consumer profiling (Canhoto and Backhouse, 2007). With one-on-one marketing becoming increasingly the norm, in particular in online environments, the value of big data analytics certainly lies in its potential for creating individual consumer profiles on which businesses can act (Trusov et al., 2016).

2.3.2. Predict your consumer

Besides the building of more precise consumer profiles, the prediction of future consumer behaviour, as well as how it might evolve, is an equally important part of big data analytics in marketing. For example, in the insurance and financing industry, application of big data analytics to psychological traits can be used to predict loan defaults of individual consumers and borrowers (Netzer et al., 2019). But, applying big data analytics to financial and demographic data is not only used to predict if a person will default. Analytics and machine learning technology are also used to infer the psychological traits of loan applicants based on the wording and language they use in their application. What the authors dub a "digital involuntary 'sweat'" (ibid., p. 976) refers to different phrases and words that can be linked to distinct psychological traits. While, as a consequence, this reveals much more about the individual applicant than she or he might want to reveal, the potential of this application of big data analytics and machine learning in different sectors can significantly help in predicting and profiling consumers. It could also be extended to phone and speech analysis, as well as other application areas, such as recruiting, both in human resources as well as education. In general, sentiment analysis and opinion mining, as in capturing individuals' opinions and sentiments through the analysis of their language, speech and texts, is a growing field in big data analytics, particularly popular in marketing settings due to its potential impact on influencing consumer behaviour (Bello-Orgaz et al., 2016; Gandomi and Haider, 2015).

The analysis of individuals is not only limited to the sole prediction of consumer psychological traits but can also include the prediction of her or his psychological states, including emotions, stress, mood or attention. Although this is regarded as more challenging than predicting traits, if these are identified in real time, they enable a

personalisation of communication and targeting strategies based on the psychological state of a (potential) consumer. Identifying and predicting if a consumer is happy, angry, stressed or calm can be extremely fruitful as it allows for immediate targeting with extremely personalised advertisements, offers and messages that specifically take advantage of the consumer's emotional state (Matz and Netzer, 2017).

2.3.3. Target your consumer

By uncovering a more detailed knowledge about consumers through big data analytics in consumer research and marketing, companies aim at better understanding and representing their consumers. The way companies aim at benefiting from these consumer conceptualisations, which can be in the form of detailed segmentation as well as individualised consumer profiles, is through targeting consumers with personalised offers and advertisement (Singh and Singh, 2017; Trusov et al., 2016). The targeting of consumers goes hand in hand with their profiling, as the first is not possible without the availability of the detail found in the profiles (Venter et al., 2014). The gap between the two is the ability to predict consumers, in terms of how they will develop in the near future as well as how they will behave in consumption situations, and the ability to personalise offers and advertisements based on these predictions (Liu and Mattila, 2017). While predictive analytics in marketing is nothing new, the availability of fine-grained consumer behaviour data has allowed companies to make much more detailed predictions regarding future consumer behaviour (Junqué de Fortuny et al., 2013; Martens et al., 2016).

One of the earliest examples of consumer prediction and targeting through the analytics of big data was uncovered through specifically tailored offers sent by the US retailer *Target*. In the early 2000s, *Target* had started to implement big data analytics to analyse large amounts of consumer data and generate predictions based on them. This went as far as predicting whether female consumers were pregnant. Such a forecast was made based on a range of 25 product-variables and served as a baseline to accurately predict the stage of pregnancy of women. The purchase of these 25 products was seen to significantly change in the second trimester of the pregnancy of the consumers. Identifying the start of the second trimester of a pregnancy meant knowing the estimated due date before any potential competitors did. This allowed the retailer *Target* to target these specific consumer-segments – pregnant women and those close to giving birth – at an earlier stage

and get them to buy baby-related products and, subsequently, other products at their stores. The "pregnancy-prediction" score proved to be successful and could make relatively accurate predictions throughout the duration of the pregnancy which, in combination with other information about the consumers, enabled specifically-targeted advertising. The predictive and targeting practices were revealed when *Target* sent out offers to a young girl living with her parents who had not revealed her pregnancy to them. The personally addressed coupons and advertisement for baby clothes and cribs revealed that on her behalf (Duhigg, 2012).

This example dates back more than 15 years and, in the meantime, the opportunities for big data analytics to establish consumer profiles, predict consumer behaviour, personalise offers, and target individual consumers with such offers have increased. These practices have, for example, become an important means of supporting consumer retention (Amado et al., 2018; Ascarza et al., 2018). Instead of only relying on reactive marketing campaigns, in which companies try to regain former consumers, big data analytics allows companies to use proactive consumer retention marketing campaigns. Proactive campaigns for consumer retention have only become possible through big data analytics. Due to a lack of consumer data, as well as a lack of analytical capabilities, the identification of which consumers are at risk of leaving the company, why they might do so, what might be done to counter that and what would the benefits be of these measurements, was not possible through classical data analytics. Ascarza et al. (2018) have provided an extensive example of how to establish proactive marketing campaigns for consumer retention through big data analytics, including what types of consumer data to analyse, which kinds of methods to use and for what purposes. Consumer data on behaviour and individual profiles, as well as consumer segmentation, are substantiated through predictive probability models, as well as machine learning techniques, to predict which consumers could leave the company and why. Importantly, however, the analysis does not stop here, and it further predicts not only when to target consumers and to what extent, but it can also provide predictions of the economic value, and efficiency of these efforts (Ascarza et al., 2018, 73f.). Thus, as in other examples, the process involves the profiling of consumers and predicting how they might behave and change, which informs the personalisation of offers and the targeting of the customers in order to retain them for the company (Martens et al., 2016).

The predictive potential of big data analytics in consumer research and marketing offers many different applications and is not solely limited to the retention of consumers. As we have seen above, sentiment analysis through big data analytics is also a very popular method of predicting a wide range of consumer behaviours, psychological traits or emotional states. It can be used to identify how likely a consumer is to default on credits or payments (Netzer et al., 2019). In the field of retail, consumer profiles and predictions can be generally used to target consumers with personalised advertising (Trusov et al., 2016). In combination with the ability of big data analytics to produce real-time results, these profiles and predictions can be used to target specific consumers in the close vicinity of their stores (Fong et al., 2015). Finally, in e-retailing and e-businesses in general, there lies an advantage in the opportunity to automate many of the processes necessary for e-commerce. In comparison to other areas, the personalisation and targeting of individual consumers at different locations on the web and in different situations can be automated through recommender engines – systems that recommend products and services they would likely buy or consume in that instance (Behera et al., 2020).

2.3.4. The creation of a data double consumer conceptualisation

Big data analytics in consumer research and marketing employs many applications that relate specifically to the consumers.[1] The amount of knowledge that can be gained about consumers is extensive, and it can go on to reach a very detailed level. As a result, the possibilities for organisations to profit from this knowledge about their consumers and to gain economic value by predicting and influencing consumer behaviour appear to be substantial. Looking at big data analytics and its possibilities in consumer research and marketing from a surveillance studies perspective, it very much involves the implementation of what Haggerty and Ericson (2000, p. 606) have termed the *surveillant assemblage*. The surveillant assemblage abstracts "*human bodies from their territorial settings … separating them into a series of discrete flows. These flows are then reassembled into distinct 'data doubles' which can be scrutinized and targeted for intervention*."

[1] There are many other application areas of big data analytics in marketing, as well as other parts of an organisation, such as production, the supply chain, human resources, etc. Since my research is specifically focused on consumer research and marketing, or the conceptualising of consumers, these areas are not covered here. For an overview on other areas of application of big data analytics in organisations, see e.g. Sheng et al. (2017).

The building of consumer profiles, with all the detailed information about the consumers, involves the abstraction of their human body and its reassembly into a virtual 'data double', ready for targeted intervention. However, as I will discuss in the following section, the consumer profiles of the virtual data doubles are not entirely what they might seem. While much of the big data analytics and marketing literature focusses on the possibilities of big data analytics, the limitations and constraints are only rarely discussed. Even less so are epistemological and ontological questions surrounding big data analytics. As a result, in the next sections, I will discuss big data analytics in general, as well as in consumer research and marketing, from a more critical perspective.

2.4. Critical Data Studies – towards an alternative definition of big data analytics

The application of big data analytics in consumer research and marketing aims to uncover new knowledge about consumers. The expectation is that big data analytics opens the possibility of quantifying and classifying almost every single quality and trait of an individual consumer, often assisted by automated systems such as algorithms and machine learning to cope with the sheer amount and diversity of the data and perform the analysis. With this new knowledge, private companies hope to construct new consumer segmentation and profiles, upon which predictions can be made, offers and advertising personalised, and individual consumers targeted. In the mainstream big data analytics and marketing literature, the understanding of big data analytics is that it produces a virtual data double of the consumers (Haggerty and Ericson, 2000). The complex behaviour of consumers, which is distinctive and unpredictable, can now be deconstructed in the form of data, while algorithms and models reconstruct this data into rational and comparable structures, sorting, classifying and segmenting the consumers (Lyon, 2014; Pridmore and Zwick, 2011). The data double serves as an intermediary upon which organisations can act if they want to influence the real consumer.

In the literature, big data analytics is depicted as enabling the creation of an exact knowledge of consumers which helps in uncovering the deepest secrets about them and, ultimately, is perceived as a caesura or a new era for operating companies (Ducange et al., 2018). Big data analytics is seen as introducing a mathematical and scientific method to marketing, steering away from a more subjective approach of social science and market research (Schneider and Woolgar, 2012). However, the ability to exactly produce a data

double of the consumer by creating an exact knowledge of it through big data analytics is often a misconception. This is a misconception based on the alleged objectivity of (digital) data, that, if collected, processed and analysed correctly, will produce an exact knowledge (boyd and Crawford, 2012; van Dijck, 2014). This misconception fails to consider that, in big data analytics, there are many procedures which will affect the outcome of the knowledge collected and limit the "objectivity" of the results. Datafication as the "process of rendering into data aspects of the world not previously quantified" functions within a rationality that comprises certain "norms, strategies, mechanisms and economics" (Kennedy et al., 2015, p. 1f.). There are, for example, the statistical limitations of big data analytics. The results of the analysis are only ever applicable within a limited probability range and are, by default, never an exact representation of *reality* (Kaplan et al., 2014; Symons and Alvarado, 2016). Also, the big data technologies that are responsible for the operations have an influence on the production of the datafied knowledge and affect how accurately or inaccurately data can represent what it should measure. Finally, the organisational setting and the human influence on big data analytical operations also have their effect on the resulting datafied knowledge (boyd and Crawford, 2012).

In the mainstream big data analytics or marketing literature, these problems and limitations are rarely addressed. Most of the articles are based on best-case scenarios regarding the implementation and use of big data analytics in consumer research and marketing, and in which limitations and critical reflections are only addressed in the final paragraphs (Behera et al., 2020; Côrte-Real et al., 2017; Liu and Mattila, 2017; Mazzei and Noble, 2017). If the limitations of big data analytics are addressed more thoroughly, they are largely limited to the barriers organisations might encounter when looking to implement big data analytics, thus not addressing the potential issues with the analytics as such (Alharthi et al., 2017; Brock and Khan, 2017; Cabrera-Sánchez and Villarejo-Ramos, 2019; Erevelles et al., 2016; Günther et al., 2017; Khan and Brock, 2017; Wielki, 2013). Actual articles that address the wider limitations and difficulties of big data analytics are rare in the established literature. In these instances, the focus lies mainly on the operational difficulties of carrying out the analytics. While here, wider societal and ethical questions are also rarely addressed, these articles do, however, address issues of misrepresentation through big data analytics, access to data and data quality, as well as

the generally misleading promise of exact knowledge about consumers or individuals (Bosch, 2016; Gandomi and Haider, 2015; Jagadish, 2015; Wang et al., 2016).

The following sections will rely more on the Critical Data and Surveillance Studies literature, which has already addressed the issues and limitations of big data analytics in the past. This forms an important part of framing the problem that occurs through using big data analytics in consumer research and marketing, particularly to establish consumer conceptualisations that are expected to be a data double of the real consumer, despite not being feasible. I will first start by working towards an alternative definition of big data analytics and elude the issues of big data analytics as a means of creating an objective and accurate knowledge, known as the ideology of dataism (van Dijck, 2014). I will explain what dataism is and show how big data analytics is a sociotechnical process, entangled in human activity and the technology responsible for the analysis, such as the data or the algorithms.

2.4.1. The dataist ideology

The expectation in the marketing literature is that, by using technological tools to get quantified measurements of consumers, big data analytics produces results that are objective, neutral and free from errors. The requirement in organisations and marketing is to have objective and *true* knowledge to be able to make reliable decisions. As, for example, Grandclément and Gaglio (2011) have shown, the acceptance of knowledge by marketers and managers depends on a range of factors, of which one is the perceived reliability and accuracy of the information. When assessing whether knowledge about consumers is reliable, the traditional research methods of consumer research in particular, such as focus groups or panels, are perceived by marketers as only partly reliable. Methods where the consumer is aware of the research are considered to be prone to response-bias and to being subjective (Dubuisson-Quellier, 2010; Schneider and Woolgar, 2012).

With big data analytics as a research method, this subjectivity is expected to be eliminated. Big data analytics transmits the expectation of an objective quantification of consumers and has an aura of neutrality, flawlessness and truthfulness (boyd and Crawford, 2012). The mathematical and scientific approach to a problem – in the form of knowledge about consumers – is perceived as a gateway to overcome the subjectivity and

inaccuracy of previously used consumer and market research methods. As a result, big data analytics conveys the datafied knowledge with a higher level of objectivity, especially contrasted to the more traditional, fallible methods of gaining knowledge about consumers (Gillespie, 2014; Schneider and Woolgar, 2012; van Dijck, 2014).

However, this perception of big data is solely ideological, or a myth, that is not applicable in practice (boyd and Crawford, 2012). The algorithms and models that perform the calculations and the analysis only provide a "*promise* of mechanical neutrality [… and] algorithmic objectivity", which is translated into the datafied knowledge (Gillespie, 2014, 182f.; italics added). This perception and expectation of objectivity and neutrality are what van Dijck (2014, p. 198) has termed the ideology of dataism, which she defines as the "widespread belief in the objective quantification and potential tracking of all kinds of human behaviour and sociality through online media technologies [… as well as] trust in the (institutional) agents that collect, interpret, and share (meta)data culled from social media, internet platforms, and other communication technologies."

The issue that van Dijck (2014) addresses through her conception of dataism closely relates to what other scholars have analysed through the lens of a trend towards (digital) positivism through big data analytics, for example by critical theory scholars such as Fuchs (2017) or Mosco (2017). Similarly to how van Dijck (2014) perceives big data analytics as reducing human behaviour and sociality to mere quantifications, Fuchs (2017, p. 40) similarly criticizes that in social science research such as social media studies, big data potentially leads to an "absolutism of pure [digital, quantitative] methodology" – in the form of digital positivism. According to Fuchs, certain fields of research perform a radical paradigm shift towards digital positivism through their increased reliance on big data analytics. However, by predominantly relying on such hyper-quantified methods, there is a big risk to fail to "connect statistical and computational research results to a broader analysis of human meanings, interpretations, experiences, attitudes, moral values, ethical dilemmas, uses, contradictions and macro-sociological implications" (ibid.).

In another recent study discussing the use of big data in the current COVID-19 pandemic, Di Salvo and Milan (2020, p. 2) also draw a comparison between dataism and positivism. By taking the example of August Comte's "A General view of Positivism", Di Salvo and

Milan show how the definition of positivism by Comte closely relates to van Dijck's (2014) definition of dataism. While for Comte, "positivism brings 'the real as opposed to the chimerical' to the forefront" and contrasts "the precise with the vague" (Di Salvo and Milan, 2020, p. 2), big data analytics through the lens of dataism is the precise and "objective quantification" (van Dijck, 2014, p. 198) of reality, particularly in contrast to the vague and fallible methods of qualitative consumer research. Di Salvo and Milan (2020) also criticize that positivism as well as dataism run the risk of ignoring broader meaning and context.

Furthermore, as Fuchs (2017) discusses through the example of social media research, the drive towards digital positivism and hence also towards dataism, can mainly be observed in areas that have been considered being unreliable and subjective, such as disciplines in the social sciences and humanities (Latour, 2016). Many disciplines and operations have the goal of becoming more quantitatively directed, with the aim of relying on objective methods producing objective knowledge. For this, big data analytics seems like a promising path (boyd and Crawford, 2012; Crawford et al., 2014). In the case of marketing, Layton (2016) refers on the one hand to the managerial and economic practices which are dominant in the field and in which consumption is largely perceived in economic terms. On the other hand, marketing has also strong ties towards social sciences, particularly where the focus lies on consumers as social groups, their behaviour and psychology, micro- and macrostructures consumers act in. This is to show that marketing as a discipline and as a practice lies in between managerial-principles and social science approaches.

This distinction between "marketing-as-management" (Layton, 2016, p. 242) and "marketing as a social science" (ibid., p. 241) is insofar important as it addresses the two different ontologies and philosophical worldviews marketing is exposed to. Marketing-as-management is largely dominated by a realist ontology and tends to adopt a quantitative approach, in which the positivist paradigm dominates. Through this positivist worldview, marketing-as-management practitioners consider that through valid and standardised methods, an objective reality about the market and their actors such as consumers can be depicted. In contrast, marketing as a social science is closer to a relativist ontology and an interpretivist paradigm. This is particularly the case when marketing research considers questions such as the "workings of trust, power and

influence, [or the] understandings of the reach and dynamics of social networks" (ibid., p. 243). However, this does not mean that marketing as a social science cannot be realist and guided by a positivist worldview, as also within social sciences, different ontological and epistemological strands persist.

In marketing, both approaches – the managerial as well as the social scientific – can be observed throughout its history. Nonetheless, the marketing-as-management approach, with a positivist paradigm of objective quantification is what has been dominating the field as of lately (Layton, 2016). Which is also why the implementation of big data analytics and the concomitant dataism can be assumed as a logic continuation of managerial marketing and a tendency towards the worldview of (digital) positivism. Certainly as big data analytics is perceived as gaining information that is not distorted through bias by the consumer or the researcher, but accurate and objective due to the mechanical neutrality of big data analytics (Gillespie, 2014; Wilson, 2008).

The problem with this dataist assumption is, however, that the technological, non-human and algorithmic process of big data analysis cannot be entirely stripped of its human connection (Pridmore and Hämäläinen, 2017), nor can it be entirely neutral or infallible (Pasquale, 2015). Big data analytics are sociotechnical systems that are heavily dependent on human (inter-)action and involvement. There are three major aspects which refute the perceived objectivity and neutrality of big data analytics. These are: (1) the many steps in the big data analytics that require human intervention and rely on subjective assessments, such as in the choice of variables or the interpretation of the results (boyd and Crawford, 2012; Jasanoff, 2017); (2) the data itself, which is prone to error and uncertainty, such as through the manipulations that are necessary to process it for analysis (Agostinho et al., 2019; Busch, 2014; Symons and Alvarado, 2016); and (3) the algorithms *per se* can also include bias and errors as they are produced by humans and reproduce a social world (Mittelstadt et al., 2016).

2.4.2. The sociotechnical of the analysis in big data analytics

Lisa Gitelman (2013) has titled her excellent publication on big data analytics as a cultural phenomenon "Raw Data" is an Oxymoron. This title refers to the necessity to process data to gain information and knowledge since data is always generated and interpreted and is not just a natural resource. Although proponents of the (post-)positivist paradigm

do not necessarily assume that quantitative research is free from interpretation and just a product of facts, big data analytics and its potential for conveying a mechanical neutrality and algorithmic objectivity – in the form of dataism – risks generating a mythology of producing objective facts (boyd and Crawford, 2012; Gillespie, 2014; van Dijck, 2014). The technology appears to remove human subjects from the process of knowledge creation and, thus, remove subjectivity and create an objective truth. However, every step of big data analytics involves individuals. The choice of hypotheses, data, variables, and calculative models need to be determined by researchers. The data needs analysis, at the core of which lies the interpretation of data. While big data can be helpful and accurate in discovering patterns, data analysts specifically need to search for certain patterns, assume linking variables and look for connections. And the researchers are responsible for understanding – and thus interpreting – the results of the analysis (boyd and Crawford, 2012; Crawford et al., 2014).

Interpreting the results of the data analysis appropriately will always depend on the initial theory and hypothesis by which marketers and data scientists operate (Kitchin and Lauriault, 2018). Asking the 'right' questions and including the 'right' variables are crucial in the discovery of connections and patterns in the data. In every one of these steps, different "politics" can occur, as described by Jasanoff (2017, 11f.). There is the *politics of framing*, i.e. the way questions for the data analysis are formulated and the problems are defined. In a marketing context, this can relate to how a problem is perceived and how important it is in order to know more about the consumers. There is the *politics of selection*, referring to how topics and hypotheses are selected for the data analysis; for example, is the focus of the analysis solely on the high-value consumers or are other issues and operations the basis for the analysis? And finally, there is the *politics of non-knowledge*; this is the knowledge that big data analytics does not address or is not considered by the data analysts or marketers. While it is difficult to address the unknown, particularly when there are so-called "unknown unknowns", it still should be an important part of the analysis and ignoring it will influence the resulting datafied knowledge.

In terms of the practice of doing big data analytics, there is a further important issue relating to the interpretation of the data, which is the pitfall of equalling correlation with causation. While, through the analysis, patterns might be uncovered between variables, the actual existence of these patterns is not given but might just be the outcome of having

such a large amount of data and variables included in the analysis (Symons and Alvarado, 2016; Zuboff, 2015). For example, while a correlation between shoe-size and spending behaviour of consumers might be an outcome of data analysis, this does not mean that shoe-sizes influence how much consumers are willing to spend. Although this example might be rather obvious, in many situations the correlation between variables in big data analytics might be more difficult to spot.

Finally, although big data analytics promises a reduction of risk and uncertainty, it will never be possible to eliminate errors (Agostinho et al., 2019; Symons and Alvarado, 2016). Also, since big data analytics is based on statistical models and calculations, error rates and confidence intervals mean that a certain inaccuracy and uncertainty will always remain (Kaplan et al., 2014). As Busch (2014, p. 1736) mentions, in some scientific disciplines, such as in high energy physics, a considerable amount of time is spent analysing and addressing errors. In other disciplines, such as the social sciences and marketing, error analysis is rarely addressed. However, as big data analytics offers "a mechanical corrective, a concrete and technical method to analyse, detect, eliminate, and neutralise human error" (Agostinho et al., 2019, p. 430), the interpretation of errors in big data analytics in marketing should not be ignored, but rather examined.

2.4.3. The sociotechnical aspects of the data in big data analytics

The entity of *data* is often considered to be solely technical and, thus, neutral, which ignores its sociotechnical entanglement. First and foremost, data requires manipulation and cleaning before being ready for use for big data analytics (boyd and Crawford, 2012). This is a process which involves subjective decisions concerning why certain data should not be included – such as it not leading to the expected results upon testing and, thus, being assumed to be erroneous (Ariztía, 2018). Another issue that is often overlooked in terms of data is that of *lossiness* (Busch, 2014, p. 1732). *Lossiness*, as defined by Busch, means that the collection of data does not necessarily register all the important cases, and that certain information can be lost. This can include extreme outliers that are ignored during the data collection, such as marginalised individuals that only have limited access to the internet (Agostinho et al., 2019), or certain aspects that are difficult to operationalise into measurements and cannot necessarily be calculated (Busch, 2014).

A major contributing factor of lossiness in data lies in the omnipresence of standards. Through standards, definitions are established of what counts as the norm and what lies outside of it. While standardisation is perceived as a neutral practice, it is often a highly politicised one (Espeland and Sauder, 2007). There are many examples that show how entire populations are not accounted for because they are considered to be out of the norm of measurements. Criado Perez (2019) has, for example, dedicated a whole book to showing how the typical middle-aged, white male is defined as the norm in almost every situation of daily life. As a result, this norm is being designed into everyday practices, leading to women encountering discriminatory and biased situations on a daily basis. Just to list a few examples, this can be in terms of road safety, because car safety belts are not designed to secure the female body as there are almost no female test-dummies (ibid., p. 186ff.); in healthcare because females are significantly underrepresented in clinical trials (ibid., p. 195ff.); or in public transport, which is mainly designed for the typical (often male) commuter and ignores the needs of care-givers, who are mainly female (ibid., p. 29ff.). Although this daily discrimination does not specifically relate to the use of big data analytics, it shows how, through time, lossiness in data can result in the emergence of these issues as they are simply not captured in data – for a variety of reasons – and are not addressed. The proliferation of big data analytics and the increased reliance on data as a means of standardisation has the potential to aggravate this issue (Krenn, 2017a), as I will discuss more specifically in the next chapter.

Besides the misrepresentation of individuals or entire populations through this lossiness in data collection, this can also lead to the effect of other phenomena or populations being over-represented and their importance being amplified or others disproportionately represented (Agostinho et al., 2019; Busch, 2014). As a result, careful attention needs to be paid to what the narrative is that is being told through the data analysis. As data in itself, as well as through its interpretation, is rarely able to represent certain phenomena in their entirety, it risks creating a narrative that is only “true” through big data analytics (Ball et al., 2016). This can, for example, be through the use of algorithms and data analytics to predict re-offenders in the US Justice system, in which the algorithm fixates the narrative of who will be a re-offender, often without the possibility of redressing this decision (Sanders et al., 2015; Sanders and Condon, 2017). A similar example is the algorithm that is currently being tested in Austrian Job Centres for the profiling of job-

seekers, which also risks setting the narrative of what type of job-seekers can be catered to (Allhutter et al., 2020).

Data can be further influenced by the sampling strategies that are often necessary in the selection of what data to include and what not, implying there are still subjective decisions that are made as to what data should be used for the analysis (Busch, 2014; Kaplan et al., 2014). These decisions can depend on the expected outcome of the analysis and can, for example, be based on whether the dataset for the analysis should include all the consumers of a certain retail chain or only those that purchased a certain product (Busch, 2014; Symons and Alvarado, 2016). Due to the sampling, there is always a risk of having errors and biases in the data, the same as with traditional statistical methods. These include sampling biases, a bias in the scope of the recorded information, lack of representative data, outdated data, missing (important) variables, errors in measurement, etc. (Kaplan et al., 2014).

2.4.4. Sociotechnical algorithms in big data analytics

The last thing to consider is the sociotechnical nature of the algorithms themselves, which are an important part of the functioning of big data analytics. These algorithms are prone to suffering from bias and discriminatory practices in their decision-making processes (boyd and Crawford, 2012; Lee and Björklund Larsen, 2019; Mittelstadt et al., 2016). Biased and unfair discrimination in digital technology is surely not a new development. Friedman and Nissenbaum (1996) described, more than two decades ago, how computer and algorithmic systems can include bias. They considered that bias may occur in algorithms when they "systematically and unfairly discriminate against certain individuals or groups of individuals in favour of others." Furthermore, unfair discrimination suggests that the system "denies an opportunity or a good or (…) assigns an undesirable outcome to an individual or group of individuals on grounds that are unreasonable or inappropriate" (ibid. p. 332).

Algorithms work on previously defined characteristics and learn from these rules and the data in the dataset. This is independent of whether they are used for the sorting of data through classifying and clustering algorithms, for the analysis of data through identifying association rules, for correlations or networks in data sets, or for predictions. As a result, the bias, discrimination and ideology are often unintentionally encoded into the algorithm

and the analysis (Dourish, 2016; Feldman et al., 2015; Lee and Björklund Larsen, 2019; Sanders and Sheptycki, 2017; Sivarajah et al., 2017). Mager (2012) has shown how capitalist ideology is inscribed in search engine algorithms, prioritising consumption-focused links in their search results. She has identified three central actors which contribute to this inscription-practice: the engineers as the people who are the architects of the code; the "website providers, who create websites and link the connections the search algorithms need to index, rank and display results" (ibid. p. 775), and the users of the search engines, who leave a trail of their data and who automatically provide feedback-loops.

Algorithms can only learn based on the initial input, the data, and the feedback they get, all of which are based on set rules of learning. In this way, the biases of the creators can trickle down on to the algorithms, and further onto the datafied knowledge. This can be observed in many areas where algorithms are applied and can target individuals differently. Sweeney (2013) has shown how online searches of racially associated names – names being statistically more often assigned at birth to black or white babies – trigger different advertisements by Google AdSense. The black-identifying first names resulted more often in personalised advertisements suggesting arrest records of the person than for white-identifying first names. The reasons for this discriminatory behaviour of the advertisement algorithm are difficult to assess and can only be assumed. As with the ideology in the search engine algorithm (Bilić, 2016; Mager, 2012, 2017), there are multiple agents that influence the behaviour of this one: (1) the advertiser who provides the ads and, ideally, wants the highest click results from the ads they provide; (2) Google as the platform providing the advertisement possibilities and also generating a profit by high click rates; and (3) the users giving feedback by clicking certain ads over others. The algorithm thus learns over time what advertisements to prioritise to generate more clicks and to align with "the financial interests of Google, as the ad deliverer, [and] with the advertiser" (Sweeney, 2013, p. 34).

Lee et al. (2019) have also proposed the consideration of algorithms as sociotechnical systems which only work through their relationship with society, technology and nature, and which are influenced by them. The advantage of looking at the relationality of algorithms lies in the ability to perceive them much more in their plurality instead of as a single entity. This also helps in explaining the multiple normativities and biases that can

be included in the algorithms, and in the analysis of big data (Grosman and Reigeluth, 2019; Lee and Björklund Larsen, 2019). In addition, as potential bias and discrimination through algorithms become much more difficult to detect, assessing algorithms more broadly and relationally potentially allows for multiple ways of addressing these issues of bias and discrimination. These can be through judicial measures in the form of privacy and anti-discrimination laws, as well as ethical assessments and design processes (Mann and Matzner, 2019; Mittelstadt et al., 2016).

Another issue with algorithms is the obfuscation of their functioning, which makes it almost impossible to check the validity of the claims made by algorithm-based decisions. In particular, in the private sector, where decision-making processes are often considered business secrets, using big data analytics adds an additional layer of secrecy (Pasquale, 2015). Increasingly, the reasoning behind algorithmic decisions and choices being hidden not only from those impacted by the decisions – in this case, the consumers – but also from the ones who are supposed to be making the decisions – the marketers and advertisers. The algorithm works as a black box in which the design of the programmers of the algorithm, as well as the process by which the data input is transformed into output, is not visible, and only the successful output of the device is of interest (Latour, 2000; Pasquale, 2015). While the algorithms cluster data and individuals, assist or produce decisions, or predict events and behaviour, the non-transparency is seen by the ones deploying them as a necessity to guard against competitive advantage and any threat to intellectual property (Pasquale, 2015).

Algorithms mostly work silently as an intermediary. Only when the algorithm does not deliver the intended results, when the desired output is no longer guaranteed, does it inherit a more prominent role, which requires a dismantling of the black box (Latour, 2000). Opening the black box of algorithms is, for individuals, next to impossible. If a consumer is denied a new data plan for her or his mobile device due to credit scoring, the decision is generally done algorithmically. For the individual, knowing on what basis and according to which rules the decision was made to deny the plan is not possible without extensive enquiry. And even with this enquiry, the companies and organisations using the algorithms do not need to comply and reveal their decision-making processes. Active secrecy and obfuscation are thus hindering the understanding of algorithms (Pasquale, 2015).

Furthermore, there is no guarantee that the functioning of the algorithm is actually understood upon it being rendered transparent. If algorithm developers refer to algorithms as “magic black boxes” and their work as a “black art” (Thomas et al., 2018, p. 2f.), the technical opening up, e.g. by publishing the source code of the algorithms or of the rules, will probably not contribute to a better understanding of the decision-making processes amongst lay individuals. Also, due to the constant changing of the algorithms, which is an inherent attribute there to help them learn and improve their decision-making capabilities, the transparency can only be temporal. As such, algorithms are not only obscure but also “malleable” as they are “easily, instantly, radically, and invisibly changed” (Gillespie, 2014, p. 179). Consequently, there are inherent power imbalances regarding big data analysis and algorithmic decisions-makers. Not only can individuals not steer and determine what data is collected (or even be aware of data collection practices), they are also not able to influence the purpose of its usage and its processing (Zwitter, 2014).

From the perspective of the end-user of the datafied knowledge, such as the marketers basing decisions on algorithmic data analysis, the obfuscation of big data analytics through the use of algorithms creates the issue of epistemic opacity in relation to how new knowledge through big data analytics is produced (Symons and Alvarado, 2016). The epistemic opacity in big data analytics happens because of the algorithms as the process between the data input and the datafied knowledge is black-boxed and cannot be understood by the end-users. In many instances, it is impossible for marketers to check whether the analysis was correct, whether the algorithm operated as intended and whether the newly created datafied knowledge exactly represents what it should be representing.

While the considerations surrounding epistemic opacity are largely of a philosophical nature, it is important that they are considered when looking at the application of big data analytics. When the analysis and, subsequently, the datafied knowledge is responsible for deciding on access to products and services for individual consumers, or the individual pricing of these products and services, it is necessary to have an understanding of the nature of the knowledge production process and, especially, about the possibilities of errors occurring within this process (Symons and Alvarado, 2016). Many examples have shown that big data analytics and the underlying algorithms are not infallible, are prone to bias and discrimination and that there is a demand for careful consideration when

making generalisable claims. For this research, it opens up the question of how big data analytics emerges as a reliable source of knowledge about consumers, other market actors, or the market in general, that is trusted and used in marketing situations (Gillespie, 2014).

2.5. An alternative definition of big data analytics

All these issues do not mean that quantitative methods or big data analysis are as subjective as qualitative methods. Both paradigms of positivism as well as interpretivism have their advantages depending on the context, the object/subject of study, and the type of interpretation that is aimed at. We can say, solely through the basic statistical concepts of reliability and validity, that the results produced by big data analysis are more generalisable and less subjective. However, this does not mean that the results are entirely objective and neutral. As the examples above show, big data analytics should not be considered solely a technical process; it should be seen as a sociotechnical assemblage. The datafied knowledge that is being produced cannot be automatically considered as being factual and objective but requires critical scrutiny, like every other kind of knowledge. However, through the ideology of dataism, there risks being a misconception about how the process of big data analytics functions and the kind of knowledge it produces. Relying on dataism risks leading us to believe that every aspect of life and the individual can be objectively quantified and measured and that we should trust in the institutions and technologies that help achieve this objectivity.

As a result, I would like to introduce an alternative definition of big data analytics, which can complement the definition presented above, it being the combination of techniques and technologies for the (automated) analysis of exhaustive, complex and unstructured datasets in real-time, with the aim of revealing statistically unidentified connections and patterns in the datasets, creating new knowledge that assists in decision-making. Through its role as a research method in many different fields of application, big data analytics has wider societal implications and epistemological qualities and is, therefore, much more than just a technical entity. In order to better address the sociotechnical assemblage of big data analytics, I would complement the definition with a definition provided by boyd and Crawford (2012, p. 663). According to them, big data analytics is “the interplay of:

- *Technology*: maximising computation power and algorithmic accuracy to gather, analyse, link, and compare large data sets.
- *Analysis*: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims.
- *Mythology*: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy."

By having this twofold definition of big data analytics – the technical one presented in the first section of this chapter and the second presented here by boyd and Crawford (2012) – a better understanding of big data analytics can be provided, which addresses both its technical and its societal side. This is particularly important as I will focus further on the use of big data analytics in consumer research and marketing and address the practice of conceptualising consumers, as well as their prediction and targeting, more specifically under this new twofold definition.

2.6. Profiling the data double and the politics of classifications

Having addressed the general issues concerning big data analytics and the ideology of dataism, it is worth looking once again more specifically at its use in consumer research and marketing settings. As discussed in the previous sections, big data analytics in this area is specifically used to gain a better knowledge of the consumers in order to profile them better – be it in segmented or in individualised profiles – and subsequently predict their behaviour, or how they might change in the near future. This new kind of knowledge is then used to personalise offers and advertisements and target consumers accordingly (Pridmore and Hämäläinen, 2017). Specifically, the profiling of consumers can be described as the aim of creating a virtual data double of the consumer (Haggerty and Ericson, 2000) as big data analytics is assumed to produce accurate and true knowledge about the consumers.

The dataist assumption of "truth, objectivity and accuracy" (boyd and Crawford, 2012, p. 663) sees further amplification in its use for profiling consumers when it is combined with the general desire to classify the world or, as in this case, market actors, such as consumers. As Fourcade and Healy (2017c, p. 286) have noted, "scores and classifications are dual to one another", and both are based on an assumption of apparent

‘natural’ boundaries of the world and society, which can be measured and categorised. In consumer research and marketing, consumers are classified based on these measurements and categorisations, as if they are grounded on distinguishable and apparent natural traits. However, as with the practice of big data analytics, categorising and classifying individuals and consumers is an active practice that involves power, politics and subjectivities (Fourcade and Healy, 2017b [2013]; Hacking, 1986; Krenn, 2017a).

The coupling of individuals with distinguishable categories can happen in many ways. They can be self-labelled by individuals, labelled externally, as is often done in marketing situations, or can be based on natural traits. These classifications, being decisions on the sorting of characteristics and individuals, are socially constructed and spatially and temporally limited (Bowker and Star, 2000). Relying on the literature of the sociology of commensuration (Bourdieu, 2005; Bowker and Star, 2000; Espeland and Sauder, 2007; Espeland and Stevens, 1998; Fourcade and Healy, 2017b [2013]; see for example: Hacking, 1986; Roscoe, 2016), I will discuss in the following sections how, in society as well as in marketing, classifications of individuals are based on apparent ‘natural’ boundaries, with related assumptions concerning the objectivity and neutrality of quantified measurements, as well as trust in the veracity of the results of its result (Krenn, 2017a). In particular, with big data analytics in marketing, the results of these classifications are expectations of categorising consumers into exact conceptualisations of a virtual data double. However, as I have discussed with big data analytics, these expectations are based on misconceptions, and the practice of classification is not a neutral undertaking. Instead of creating data double consumer conceptualisations, these conceptualisations are turned into postulations, expected factual representations of the consumers, without real evidence of their existence.

2.6.1. Measuring and classifying as an everyday practice – in marketing and in society

In marketing, knowledge about consumers is often used to classify and order consumers and markets and is a central aspect of marketing strategies, particularly in the managerial conceptualisation of marketing. Even prior to the implementation of big data analytics, consumers were categorised or segmented into groups based on similar preferences, behaviour, needs, etc., towards which marketing strategies were tailored and targeted (Venter et al., 2014). Marketing has relied on data and knowledge about consumers, the

market setting, environment and its competitors since its early stages (Pridmore and Hämäläinen, 2017). With the emergence of digital data, ordering practices have increased significantly as an ever-increasing stream of digital consumer data also requires significantly more standardisation and clustering of this information. But big data analytics, in conjunction with datafied knowledge, have also provided the means to create even more precise categorisations, leading to increasingly tighter classifications of consumers (Fourcade and Healy, 2017b [2013]; Krenn, 2017a).

Measuring, categorising, and classifying are common activities in society, and are not limited to marketing. For example, the education system is heavily structured by metrics and rankings to compare and assess the 'success' of institutions – a ranking system which is often responsible for the allocation of government financial contributions. Also, censuses and demographics, which are used to compare nations and regions, are based on previously established categories. And as individuals, we calculate and compare on a regular basis "such potentially incomparable values as career and family, [… or] freedom and commitment in love" (Espeland and Stevens, 1998, p. 316). Due to them being omnipresent and found in almost every situation of life, the processes behind the creation of these classifications are often not visible and run in the background. However, looking more specifically at how these categorisations are created and consumers ordered and grouped, we can see that this is a practice that involves politics and power (Bowker and Star, 2000). Decisions are actively taken on how these classifications are made and what characteristics should be used to build them. In marketing, for example, categorisations – in the sense of segmentation or individualised profiles – are actively established by the marketers and their organisations (Grandclément and Gaglio, 2011; Krenn, 2017b). These categorisations are based on measurements of certain characteristics, of one or multiple common traits of individual consumers, and are aimed at reducing the complexity of the information gathered and making it comparable (Espeland and Stevens, 1998; Fourcade and Healy, 2017b [2013]).

2.6.2. The politics of classifying

It is important to look at how commensuration and classifications function, what their requirements are, who decides on the boundaries of the categories and the characteristics of the classifications, and what expectations groups and individuals have when classifying the world. In many cases, classifications are established top-down by market

professionals, imposing (knowingly or unknowingly) categories upon individuals (Hacking, 1986). One of the more common examples of classification practices in market or economic situations are the credit rating systems (Krenn, 2017b; Lauer, 2010). In these, individuals are ranked based on a wide range of variables to assess their creditworthiness. Creditworthiness generally defines the probability of individuals failing or succeeding to pay their credit rates or other recurring fees. The variables define the credit ratings and include, amongst other economic factors, income, assets and previous or current credits. Non-economic factors are also included to measure creditworthiness and help establishing these classifications. These factors include socio-demographic data such as age, education, marital status or the street where you live (Fourcade and Healy, 2017b [2013]). They can also involve softer metrics, such as psychological traits, gained through digital data or derived from the wording of credit applications (Netzer et al., 2019).

In marketing, the classification of consumer qualities is always "market-derived *and* market-oriented" (Fourcade and Healy, 2017a, 22). This means that in order to establish consumer classifications, the characteristics of the consumers that will be measured are derived from what is important for the market and oriented towards profit-making and economic value. Consequently, creating consumer classifications and conceptualising consumers focuses almost exclusively on the marketable qualities of the consumers. The measurements of consumers are focused on what knowledge provides the necessary insight that increases the economic value for the company (Fourcade and Healy, 2017a; Krenn, 2017a). This means that there is a risk of consumers being reduced to these characteristics, so that only their economic value matters and all their other qualities are entirely dismissed.

Furthermore, these conceptualisations of consumers are the result of debates within the individual organisations and within marketing teams. Despite the assumption of categories being based on natural boundaries, the boundaries are still based on negotiations, particularly regarding the definition of their standards and limitations (Bowker and Star, 2000; Gieryn, 1999). In marketing settings, the practice of classifying consumers depends a lot on these negotiations regarding the boundaries of the categories, and what type of consumers to include and exclude from certain classifications (Ariztía, 2014; Desroches and Marcoux, 2011). Consumer classifications involve the creation of imaginations of how the real consumers might look and act (Desroches and Marcoux,

2011; Sunderland and Denny, 2011). Classifications depend on who has more "authority to speak on behalf of" the consumers (Simakova, 2016; [2013], p. 33), which leads to slippery concepts (Sunderland and Denny, 2011) that are not created as one-off events but need constant re-negotiation and re-configuration (Dubuisson-Quellier, 2010).

With the emergence of big data analytics, the process of classifying consumers shifts considerably (Fourcade and Healy, 2017b [2013]; Krenn, 2017a). Consumer data is increasingly available to marketers, and it is used for much deeper and more precise analyses compared to classical data analysis and traditional market and consumer research methods. This makes unprecedented predictions concerning consumer demographics and behaviour possible, creating an additional incentive for marketers to rely on big data analytics to classify consumers (Kerr and Earle, 2013). One of the issues with big data analytics as a means of classification is not only that it obfuscates the sociotechnicality of big data analytics itself but it further obfuscates the politics behind the classifications, and solidifies a practice that has already been hidden. Furthermore, the use of big data analytics to support the classification of consumers provides such classifications with an even greater veil of neutrality and reliance on natural boundaries due to the ideology of dataism discussed above (Krenn, 2017a; Pasquale, 2015; van Dijck, 2014).

As an example, the classification mechanisms through big data analytics are heavily dependent on the classification algorithms involved in the processes (Gillespie, 2014; Matzner, 2016). Due to the amount of data involved in big data analysis, manual processing and analysis of the data are not feasible. Instead, in big data analytics, algorithmic processes take over the work of analysing the data, creating clusters and segmentations and sorting and categorising the economic actors (Pridmore and Hämäläinen, 2017). As opposed to classical data analytics, the characteristic of big data and the analytical algorithms is that they do not look for already established categories. Categorisations are, instead, derived from within the data; relationships within the data are discovered (Matzner, 2016). Algorithms become responsible for selecting what is deemed important in the data, as well as contribute to the datafied knowledge. Through the automatic analysis of data, decisions about what information is deemed as valid and what is not are transferred from human actors to (apparent) non-human algorithms (Gillespie, 2014, p. 168).

Through this process, big data analytics helps produce new consumer profiles and segmentations and highlights previously unknown relationships between characteristics, behaviours, traits, products, etc. (Pridmore and Zwick, 2011; Zwick and Denegri Knott, 2009). The categorisations produced through the analysis of large quantities of data result in a comparison of actual consumer behaviour with a "standardised" and desired behaviour, enforcing norms, ideologies and beliefs held by marketers and including rules which are deemed desirable. In this way, normativities in the form of "how one should live one's life" are "readily encodable in data". Big data analysis can promote and reinforce normativities in society, which can contradict individual subjectivity, what they experience and want to experience (Ball et al., 2016, p. 70).

While consumer data and insight are the sources big data analytics use to establish consumer classifications, the analysis is nonetheless reliant on previously performed classifications. Algorithms need to learn their decision-making from decisions that were made before (Fourcade and Healy, 2017c). It, thus, must be decided what categories and classifications are to be considered, on what criteria decisions should be made and what data is deemed relevant for the analysis (Gillespie, 2014). The same applies to the improvement of the procedures behind big data analytics. Only by relying on previously made decisions that have yielded positive outcomes can the analysis be improved. For the functioning of big data analytics, (self-)validating feedback loops[2] are a core component (MacKenzie, 2006). The feedback loops are a necessity for eliminating false results, as well as preventing false results from occurring in the future and increasing the accuracy of the predictions of algorithmic models. However, the increased accuracy can also be the consequence of an adaptable consumer. Consumers need to make themselves "algorithmically recognisable", meaning that they must adapt their behaviour to the (assumed) functioning of the algorithm and of big data analytics in general, to obtain a desired result (Gillespie, 2014).

This also means that, very often, market segmentation and classification can lead to the solidification of existing social hierarchies, as well as the creation of new social differentiations (Allhutter et al., 2020): "Methods of tracking and scoring […] data and

[2] In algorithms, the feedback loops are only partially self-validating as, especially in the beginning, individuals contribute to the program-code and the data.

metadata of consumer behaviour directly affects stratification" (Krenn, 2017a, p. 12). The social implications of these boundaries and classifications, especially being of a discriminatory nature, are a major concern. Market classifications not only "see" the consumer in a specific way, but they also act on the consumer and, as a result, have the effect of consumers seeing themselves in the way the classification portrays them. These extensive conceptualisations of the consumer can result in individuals gaining or losing access to certain markets (Allhutter et al., 2020), products or services, as well as individual price-making (Pasquale, 2015; Sweeney, 2013). Mismatches can happen in that the wrong product is attributed to the wrong categorised individual, which mostly leads to annoying situations. But mismatches can also be more severe: for example, when individuals are wrongfully denied access to market situations, or goods supplier benefit while the customer is disadvantaged. The decisions behind the operations are not transparent, being hidden in algorithms and categories, and they generally do not allow for the results to be contested (Fourcade and Healy, 2017a). This is particularly important to consider, as marketing relies on consumer conceptualisations that are significantly simplified. With categorisations and classifications being based on characteristics that are rationalised to their main marketable qualities to make them more malleable for marketing operations, the conceptualised consumer never can be an exact and holistic copy – or data double – of the individual.

2.7. Conclusion – creating a postulated consumer conceptualisation

To recapitulate, the problem that has emerged with the implementation of big data analytics in marketing is that big data analytics is increasingly used to gain new, better and more precise knowledge about consumers. This involves knowledge about the consumers socio-demographics, as well as their behaviour, in order to produce better consumer conceptualisations in the form of more detailed segmentation, in addition to individualised profiles. These consumer conceptualisations – which can be described as virtual data-doubles of the consumers – are used to predict future consumer changes and their behaviour, and personalise offers and advertising accordingly to target the consumers.

These consumer conceptualisations are depicted as factual copies of the real consumers in the established big data and marketing literature as their conceptualisations are based

on objective and quantified measurements of consumer qualities. However, as I have discussed, the critical literature strongly disapproves of this description. Instead, many authors argue that both the creation of datafied knowledge and the measurement and categorisation of consumers into classifications – thus the conceptualisation of consumers – are inherently subjective processes. These processes include a range of deliberate and involuntary decisions regarding how to collect, process and analyse data, as well as how to measure, categorise and classify the consumer. Instead of viewing big data analytics as a neutral, mechanical procedure that produces exact datafied knowledge (Gillespie, 2014), big data analytics should be considered a sociotechnical assemblage. Involving many different actors, technologies, norms and politics, the resulting knowledge – while potentially being more precise and detailed than other types of knowledge – cannot be an exact representation of reality. Datafied knowledge is still dependent on statistical limitations, involves subjective decision-making and includes technical and human errors.

As a result, questions arise whether the outcome of the processes and practices of big data analytics in marketing actually can be a virtual data double of the consumer, one that factually represents the real individualised consumer. Or whether the outcome is rather a *postulated* consumer conceptualisation in the sense that the conceptualisation is only perceived to be self-evident and factual, without any evidence of its existence? This question is also reflected in the research question presented in my introduction, which asks:

> How and why do marketers conceptualise consumers differently when using big data analytics in comparison with traditional market and consumer research methods?

To answer this question, in the next chapter, I will discuss the theoretical framework of my research, addressing how similar questions have been researched in the literature to date, elaborate on what approaches are missing in these discussions and introduce the concept of the co-production of knowledge (Jasanoff, 2006c), which should help us gain an understanding of how such questions can be answered.

3. Theoretical framework – The complexity of knowledge in marketing

This chapter sets out the theoretical framework which underpins my research. The central question of this research addresses whether marketers conceptualise consumers differently when using big data analytics in comparison with traditional market and consumer research methods. This question has its foundations in two closely interrelated disciplines. On the one hand, the discipline of Critical Marketing Studies and here in particular through a cultural materialist perspective. In this, the focus of research lies on marketing as a practice in a market that should also be considered a sociotechnical assemblage – similar to what has been discussed in the previous chapter with big data analytics. On the other hand, my research relies on the theories of knowledge production, especially those that have emerged in Science and Technology Studies, such as the concept of knowledge co-production (Jasanoff, 2006b, 2006c).

In the first part of the chapter, I will first focus on how the discipline of Critical Marketing Studies contributes to my research. Here, two theoretical branches will be discussed which both have a predominantly critical perspective on mainstream marketing theory, and both having their foundation in social science research (Zwick and Cayla, 2011). One strand of Critical Marketing Studies is based on a critical theory approach, considering "marketing as a form of power that acts (…) on consumers" (ibid. p. 7). The other, cultural materialist strand focuses more on material and social practices of marketing and address some of the issues around consumer-oriented knowledge and marketing that are also of interest for this thesis. Having its foundation in Actor-Network Theory, the cultural materialist strand of Critical Marketing Studies considers markets and marketing to be sociotechnical assemblages, involving different actors, technologies and normativities (Araujo et al., 2010). I will progress with the cultural materialist strand of Critical Marketing Studies as the theoretical framework of my research, being more suitable to research marketing and consumer and market research as a practice in change through big data analytics. In this last section on Critical Marketing Studies, the concept of markets and marketing as sociotechnical assemblages will be explained further.

In the second part of the chapter, I will add how the concepts that have been developed more specifically in the discipline of STS can help in addressing more systematically the

questions related to (datafied) knowledge about consumers, that have emerged. Specifically, the idiom of the co-production of knowledge, introduced by Jasanoff (2006c), will prove to be helpful as a theoretical framework for addressing the epistemology and the ontology behind knowledge production in general, as well as big data analytics for knowledge about consumers and the corresponding consumer conceptualisations.

The co-production of knowledge addresses, on the one hand, how the creation of knowledge is entangled in social, material, normative and cognitive procedures, in which neither the social nor the natural have primacy but, instead, are interdependent. On the other hand, the concept of co-production addresses the processes behind the stabilisation of knowledge through its credibility and epistemic authority. The stabilisation of knowledge depends communication, translation, and portability, which also includes how it is perceived and used. In particular, the latter can contribute to how new consumer conceptualisations can emerge and replace existing ones. These concepts will be explained and put into a marketing context through existing research from Critical Marketing Studies, where similar questions of how knowledge about consumers and their conceptualisations are produced, have already been addressed.

Finally, I will turn my focus more specifically back on the use of big data analytics in marketing. Here, I will discuss some of the more recent approaches to researching both big data analytics and the co-production of knowledge about consumers. However, while some of these studies have started to include research on big data analytics in marketing, the main gap in the (critical) literature still concerns how the co-production of datafied knowledge contributes to the conceptualisation of consumers.

3.1. Critical Marketing, self-governmentality and the sociotechnical assemblage of marketing

One goal of Critical Marketing Studies is to address marketing more broadly and critically, while considering it to be a practice that is deeply entangled in social, political, natural and technological processes, and through which companies can exert power on consumers. The two different theoretical underpinnings of Critical Marketing Studies – critical theory and cultural materialism – have some overlap, both critically engaging with the practice of marketing. Yet, there are some major differences in the focus of research

which need to be addressed. As I will discuss in detail in the following sections, critical theory approaches to marketing put the emphasis on "the manner in which marketing acts on consumers (…), producing a certain culture of self-government (…) that promotes techniques of self-care, self-improvement, and self-responsibility" to its consumers (Zwick and Cayla, 2011, p. 7). Its focus lies quite strongly on the consumer subject that is acted on and exploited through marketing practices, where the consumer takes a significant role as an actor in marketing practices. In contrast, a cultural materialist approach to Critical Marketing Studies considers marketing in terms of its "materialisation, (de)stabilisation, and (re)qualification" (ibid., p. 6) and as a practice that is entangled with the market, its products, its surroundings as well as its consumers – in short the whole Actor-Network of marketing (Araujo et al., 2010). This means that, on the one hand, practices in marketing have a big influence on the consumers while, on the other, consumers have a big influence on the practices of marketing. Which means that marketing should be considered a performative practice. Understanding this entanglement of marketing and the consumers is important if we wish to better understand the way consumers are conceptualised in marketing practice.

3.1.1. Critical Marketing turning inside marketing

A systematic critique of the mainstream marketing theories has emerged from within the field social science research from the 1990s onwards under the label of Critical Marketing Studies. One commonality of this research programme is that it has shifted its gaze to "*inside* marketing" to understand how marketing functions in practice (Zwick and Cayla, 2011, p. 7). Looking at marketing in this way means focusing on the practices, the knowledge and the different actors involved. The aim of the Critical Marketing approach is to better understand the effects of marketing practices, such as the manipulation of consumers or the consolidation of ideologies (Callon, 2010).

Critical Marketing research raises concerns regarding the narrow-mindedness of the traditional managerial orientation of marketing research, which often sticks to a positivist approach of acquiring objective knowledge through quantitative-based empiricism and a lack of theories (Arndt, 1985; Saren and Svensson, 2013; Simakova, 2016; [2013]). Traditional marketing often assumes a universal notion of the market and their actors, which are seen as constant, abstract and neutral, and, by doing this, it neglects the instability of the market (Zwick and Cayla, 2011). For Critical Marketing Studies, the

market, the consumers and, especially, their boundaries and the fluidity of these boundaries are an important constituting factor of marketing and vice versa (Callon et al., 2002; Zwick and Cayla, 2011). Yet, as mentioned, there are different approaches towards how to critically address the practices of marketing, which not only shift the lens of what to research but subsequently also can lead to differing conclusions.

3.1.2. Marketing, critical theory and the self-governmentality of the consumer

As discussed in the introduction of this chapter, Critical Marketing research using a critical theory approach has its focus on the way marketing installs a necessity of self-governmentality of its consumers, promoting and encouraging consumers "to fashion him- or herself as an autonomous and voluntary agent in the production of economic value" (Zwick and Ozalp, 2011, p. 237). Through marketing, the consumer is put into the role of producing (economic) value him- or herself, leading to a blurring of the lines between production and consumption. Several scholars have, as a result, coined the term "prosumers" and "prosumption" to denote how through contemporary capitalism, consumers are increasingly put into value-creation. This can be seen in the physical space, best exemplified through the self-service restaurants like McDonalds where consumers need to order their food at the counter, carry it to their table and also clean up after having finished consuming. The concept of presumption can also be seen in the virtual, where certainly social media sites have merged the consumption of content and the production of value through data processing (Fuchs, 2011; Ritzer, 2015; Ritzer and Jurgenson, 2010; Zwick, 2015).

This approach of Critical Marketing also addresses and criticises the traditional assumption of the free choice of the consumer, often adopted by managerial marketing research. As Saren and Svensson (2013, 374f.) argue, the lack of a *sovereign consumer* becomes apparent when looking at the language used in marketing literature. Marketers aim at targeting, penetrating, and capturing markets and consumers. This leads to the assumption that the supposed free will of the consumer is not only illusionary but is something which is not desired in the marketing world. Instead, the consumer is exposed to influencing and manipulating powers, generally exerted by marketing strategies and activities (Sunstein, 2015; Wertenbroch, 2015b). A power which is ultimately not solely directed towards consumers, but also at other market actors, such as competitors, governmental institutions, supply chains, the media, etc. Consequently, the goal of

marketing is not to identify consumer behaviour and needs and then adapt their commodities accordingly, with freely-acting consumers on the receiving end. The goal is rather to imagine an ideal-type consumer and their connections to others and trying to manufacture consumers accordingly, identifying their behaviour and influence their choices, controlling their behaviour (Tadajewski, 2010a, 2010b; Wertenbroch, 2015a; Zwick and Denegri Knott, 2009).

Particularly with the emergence of social media, making database marketing an important source for knowledge about consumers, the role of the consumer in the creation of his or her data and the exploitation of this data has been analysed thoroughly by critical theorists. In this, some researchers such as Fuchs (2014) (or in a more simplified version Zuboff (2015)) have focused on the exploitation of individuals' user data through the lens of Marxist theory, in which data extraction and processing is seen as an immaterial and unpaid labour. Although users might be remunerated through other means, such as the consumption of certain products, their data is extracted, commodified, and put to value, ultimately leading to an exploitation of the consumer. Others, such as Zwick and Denegri Knott (2009) or more recently Charitsis et al. (2018) have argued that in the era of "surveillance capitalism" (Zuboff, 2015), the consumer's individual data collected through social media is not what is of (economic) value. Platforms, such as Google or Facebook, establish value from this data only by processing, manipulating this data into profiles and most importantly, putting it into relation with other consumers. Thus value from consumer data can only be extracted through creating consumer profiles or "manufacturing consumers" (Zwick and Denegri Knott, 2009) and relating them to other consumers. This requires that consumers are enabled to be self-empowered and self-governed, in order to produce the data by which total control through database marketing is made possible in the first place (Darmody and Zwick, 2020).

Finally, in Critical Marketing's focus on the consumer, the discipline does not only reflect on the marketers' role but also aims to analyse the agency of consumers in the market and marketing process, specifically how forms of resistance and activism emerge and influence marketing. This can range from conventional complaints and boycotts to active acts of grouping and campaigning against brands, companies and consumption in general. Thus, as an emerging field within marketing research, Critical Marketing Studies sees its raison-d'être in critiquing the prevailing concepts of the field, which often can be reduced

to a simplification of the market and consumption, an obliviousness to the impact of marketing on the actors involved in the wider process, and the agency of the consumers (Saren and Svensson, 2013).

3.1.3. Marketing and the consumer as an Actor-Network

Moving to the second strand of critical research into the complexity of marketing – a cultural materialist perspective – and the close connection it has with the consumer can be largely attributed to Michel Callon and his application of the concepts of Actor-Network Theory to the study of markets. One of the main assumptions of ANT is that of a "radical indeterminacy of the actor" (Callon, 1999, p. 181), meaning that nothing concerning the actors is static and predetermined. As such, the motivations behind actions, the compositions of actors, their competencies, and how they come together are all indefinite. This approach makes actors and actions a vague concept: difficult to define and open to many interpretations. But it also challenges the prevailing distinctions in the social sciences, specifically the one between the social and the material.

In ANT, the inclusion of both human and non-human actors helps us to understand social action and social relations more thoroughly as a part of sociotechnical assemblages (Çalışkan and Callon, 2010; Latour, 1991; Simakova, 2016; [2013]). Framing markets as sociotechnical assemblages means that they feature human and non-human actors which are coming together through necessity and need to establish relations for a certain period to perform a market transaction. The sociotechnical assemblages of markets describe the extensive range of actors and actions that are involved in marketing, such as the different professionals, their competencies, their knowledge, their expertise and experience, the tools and technologies that are used, the way knowledge is transmitted between the marketers and the teams they operate with, as well as the subjects of study themselves, such as consumers, competitors, etc. (Callon, 2010).

Within marketing, the role of the marketers is to research how consumers function and act in market situations in order to adjust their strategies, their production or their design (Callon and Muniesa, 2005). Marketing, as a part of the sociotechnical assemblages of markets, is seen as being located "between producers and consumers" (Cochoy, 1998, p. 195), and specific to its market contexts. Marketing is also subjected to the heterogeneity, instability and complexity of the sociotechnical assemblages (Araujo,

2007). This means that for marketing and marketers, it is not sufficient to focus solely on the consumer but to also keep track of the other human and non-human actors of the market, the "networks of attachment" of the consumer (Araujo, 2007; Callon and Muniesa, 2005; Cochoy, 2010), thus adopting a more materialist approach towards Critical Marketing in comparison to a critical theorist approach (Zwick and Cayla, 2011).

3.1.4. The performativity of the sociotechnical assemblages of marketing

An important attribute for cultural materialism in Critical Marketing Studies relates to the sociotechnical assemblages of markets and marketing and their performativity. The performativity of markets and marketing is concerned with theories about the market, as well as marketing knowledge as "forms of technical-economic reasoning" actively shaping how the market functions (Kjellberg and Helgesson, 2016; Roscoe, 2014, p. 195). In terms of the knowledge companies have about their consumers, this means that marketing does not simply discover consumption and consumer behaviour, but it is performative through acquiring marketing knowledge and its subsequent behaviour (Cochoy, 2010). With the purpose of conceptualising the consumers and taking action based on this, knowledge about consumers and other market actors, as the outcome of market and consumer research, is inscribed in the product, in the service, in marketing strategies and in the market itself (Cochoy, 1998; Roscoe, 2014). Particularly in recent years, researchers following the Callonian programme have started considering the performativity of markets (Çalışkan and Callon, 2009), analysing how markets, those involved and the actants included, are shaped and performed by marketing, in retail (Azimont and Araujo, 2010; Cochoy, 2010; Dubuisson-Quellier, 2010; Hagberg, 2010), the housing market (Ariztía, 2014), or the dating industry (Roscoe and Chillas, 2014), or how the widely-used practice of market segmentation is performative (Venter et al., 2014). In this sense, marketing theories and practices do not simply *discover* consumer preferences and behaviour but actively shape them.

The performativity concept of marketing allows for the adoption of a broader unit of study when researching marketing. As Cochoy (2010) argues, by solely focusing on the consumers, marketers tend to lose track of the other actors – several mediators are involved in the market process, including both human and non-human actors. It is not solely a matter of identifying a consumer's preferences and trying to advertise and sell to her or him a relevant product. Instead, there are "many other actors and actants in between

(…) the distribution process, the distributive equipment, and the 'distributed agency' that make consumers move, goods flow and supply-side actors make money" (Cochoy, 2010, p. 30). Opting for the pragmatic approach of performativity (Overdevest, 2011), it allows us to look at marketing and markets by not only *seeing* supply and demand but by recognising that there are many other agents involved in the process. They all contribute to the practice of individuals consuming products and services, products being shipped around the world and sellers, producers and marketers increasing their income (or not) (Cochoy, 2010). And finally, it shows us that the consumers are actively involved in shaping marketing, markets and consumption. This performative framework for the topic of marketing requires a more holistic approach to researching its practices.

Looking at marketing in its entirety, and particularly their deep entanglement with the consumer and other actors of the market, has generated extensive questions in the cultural materialist strand of Critical Marketing research, regarding the problems and implications of marketing. Considering the sociotechnical assemblages of markets aim to describe "the various entities that (…) enact economic calculations and shape consumer behaviour" (Cochoy et al., 2016, p. 5), it is part of the concept to address the wide range of agents included in transactions. As an example, Cochoy (2010) has researched how specific marketing practices perform market transactions not by focusing on consumers nor on advertisements but on what he calls the "distributed agency" (p. 30). With this term, Cochoy describes the spatial assemblages of supermarkets, the distributive processes of goods, and the tools available for distribution, all of which influence not only the flow of goods but also of consumers (through the aisles of the supermarkets). Cochoy's empirical research is based on issues of the trade journal *Progressive Grocer* between 1929 – 1959, a professional magazine focusing on smaller independent groceries. In his analysis, he specifically looked at how the journal presented the ways stores were modernising to its targeted readers, the store owners. Cochoy describes the role of materiality in the form of the shelves that were channelling goods and consumers, but also checkout counters, shopping carts, etc. These not only had the effect of modernising entire grocery stores in the US (and probably other countries) but also changed consumer behaviour. Consumers could no longer just enter the store, order their products, pay and leave. Instead, they were 'forced' to walk through all the aisles of the grocery store to get what they wanted to buy and, ideally, would see and buy other products they might not have otherwise thought

about (Cochoy, 2010). This example shows that markets are not only sociotechnical assemblages but also performative practices that are not neutral.

Research into marketing from this critical – cultural materialist – perspective is interested in these broader processes of the market, often in combination with a critical stance on the classical marketing research, and has a strong focus on what exactly marketing *in practice* does, how it is entangled in a network of actors, materials and norms and what its effects on the consumers and the market itself are (Araujo et al., 2010). In the same sense as the social studies of science started looking into the processes of laboratories and scientific fact-making in the late ’70s and early ’80s (Latour, 1988; Latour and Woolgar, 1986), so have researchers in the last decade turned their interest to the processes in the market and marketing. This has led to a wide and heterogeneous variety of research topics, focusing on different parts and processes of the marketing collective, and how they perform with regard to the markets, the consumers, the products, and society at large.

One of these strands of cultural materialist Critical Marketing Studies is interested in the production of consumer-oriented knowledge in marketing processes, and its performativity on the consumers (Ariztía, 2014, 2015, 2018; Desroches and Marcoux, 2011; Grandclément and Gaglio, 2011; Sunderland and Denny, 2011). As my research is similarly interested in the production of consumer-oriented knowledge in marketing through the use of big data analytics, I also opt for a more cultural materialist approach to Critical Marketing. Furthermore, in the next section, I will introduce the concept of co-production of knowledge in detail, which aims at better researching how knowledge – here, concerning consumers – is created and stabilised, and entangled in social, normative, natural and political processes. The empirical examples from Critical Marketing Studies will help to elaborate on the concept of co-production and show its applicability to research marketing and consumer research practices.

3.2. Researching the co-production of consumer-oriented knowledge

3.2.1. Introduction – the idiom of co-production

Consumer-oriented knowledge, as one of the major contributors of marketing and consumer conceptualisations, is established through classical research methods, as well as through big data analytics. As such, for researching into how and why big data analytics affects the conceptualisation of consumers in marketing, in which potentially multiple consumer conceptualisations emerge, it is helpful to look specifically at how datafied knowledge about consumers is created. To research the production of knowledge in consumer and marketing settings, I will turn to the concepts developed in Science and Technology Studies and the social studies of sciences. This discipline started as laboratory studies in the 1980s (Latour and Woolgar, 1986) and has been interested, since its conception, in how science is being done in practice and knowledge is being created, negotiated and used. Over time, the discipline has progressed to address many other fields where (scientific) knowledge is produced. In the development of the fields of STS, multiple concepts have been brought forth that address the epistemology and ontology in science and other points of knowledge production.

In the late 1990s and early 2000s, the discipline of STS experienced a so-called *market turn*, in which several STS scholars started to show an increased interest in economic markets and all their components, and critically examine how they function and what their effects were (Simakova, 2016; [2013]). As I have discussed earlier, the seminal studies by Michel Callon (1998, 1999; Callon et al., 2002; Callon and Muniesa, 2005) have contributed key concepts of STS to the critical study of markets and economics, such as Actor-Network Theory. The thick descriptions that have been produced by ANT researchers certainly contribute to an understanding of the manifold entanglements that are present in marketing practices. However, while ANT addresses important sociotechnical aspects of marketing, other aspects, such as how consumer-oriented knowledge is produced and, particularly, how it is stabilised and made credible, are often absent and can better be addressed by the idiom of co-production of knowledge introduced by Jasanoff (2006b, 2006c).

The main argument of the co-production of knowledge is that, in comparison to the social constructivist approach to science studies, co-production gives no primacy to the social order or the natural. Instead, with the co-production of knowledge, we should consider

the natural and the social being produced together because “the ways in which we know and represent the world (…) are inseparable from the ways in which we choose to live in it” (Jasanoff, 2006c, p. 2). Knowledge cannot be an exact representation of reality as it is always influenced by “social practices, identities, norms, conventions, discourses, instruments and institutions” while, at the same time, influencing these practices, norms, discourses and politics as well (ibid., p. 3). Knowledge production is constantly entangled in and performative on cognitive, social and normative practices, as well as the materiality and naturality of the world.

The idiom of co-production was initially established for research into scientific knowledge creation and its entanglement with cultural and political assemblages (Mager, 2017), but it has since expanded into other domains. The aim here is to introduce the concept of co-production into the critical research of marketing. At a time when marketing aims to become more scientific and objective through the use of big data analytics, focussing on the managerial and economic principles of marketing, the co-production of knowledge will help to untangle the social practices, the politics, the material and the normative, behind the processes. How knowledge and knowledge production is being legitimised in marketing is also influenced by the social and political context and by identities, prospects and ideologies created by marketers, and is rightly subject to boundary-work about what counts as acceptable knowledge and what not (Gieryn, 1983, 1999; Simakova, 2016; [2013]).

Empirical examples about the practical process of knowledge co-production in marketing settings show this complexity and messiness (Sunderland and Denny, 2011). Knowledge about the consumers aims to better understand the consumers at whom big data analytics is targeted, hoping to gain an exact knowledge of them (Darmody and Zwick, 2020). However, looking at how consumer-oriented knowledge is being established, research shows that this process is not straightforward. There are examples of how such knowledge about consumers emerges from different teams within a company or is outsourced and, as a result, sees different definitions of the consumer emerging (Sunderland and Denny, 2011). Other examples show that there are different forms of contributing knowledge and expertise that are regarded as being more credible over others, trusted to be more accurate, thought to better represent the ‘real’ consumer or stand closer to the expectation decision-

makers have regarding the consumer (Dubuisson-Quellier, 2010; Grandclément and Gaglio, 2011).

For my research, the advantage of the concept of co-production is that it does not only highlight the sociotechnicality of the practice of how knowledge is created. The co-production of knowledge addresses not only epistemological questions but also addresses ontological questions (Jasanoff, 2006b), both of which contribute to answering my research question. The epistemological – or interactive – strand of co-production (ibid., p. 28f) provides an explanation for *how* consumers are conceptualised differently through big data analytics. How does big data analytics contribute to the knowledge organisations have about their consumers, in which the process of producing that knowledge is by no means a neutral, singular practice? The ontological or constitutive strand (ibid., p. 22f.) provides an explanation as to *why* consumers are conceptualised differently through the use of big data analytics since different kinds of knowledge about them emerge, are made credible and are all used to create different representations or realities about them.

In the following sections, I will start to explain more specifically the interactive strand of co-production, which focuses on the stabilisation and authority of knowledge. This explanation will be underpinned by examples from Critical Marketing Studies and will help to address the question of how different kinds of knowledge emerge in marketing, with some being preferred over others. In the second part, I will address the constitutive, ontological strand of the co-productionist idiom. Here, the focus will be on why ontologies should not be considered single but multiple entities, meaning that, within the same setting, several ontologies can emerge and co-exist (Law, 2008, 2017; Mol, 1999). These multiple ontologies can be attributed to the way knowledge is translated and stabilised, and can help explain why multiple conceptualisations of one and the same consumer in marketing settings emerge. This will be supported by empirical examples from Critical Marketing Studies.

3.2.2. The epistemology of consumer-oriented knowledge

It is important to first address the epistemological questions that the concept of co-production raises when it focusses on how knowledge is being established while being entangled in practices, norms, instruments, and institutions. Furthermore, the epistemological also addresses how knowledge is stabilised while continuously being

enmeshed in these different influences. In consumer research and marketing, this means that not only knowledge about consumers is affected by these different procedures, but that there are also different kinds of knowledge which emerge, become stabilised and are made credible due to these persisting influences. Thus, the epistemologically-oriented interactional approach of co-production addresses the ways "facts about the natural world are confronted (…) by problems of social authority and credibility" (Jasanoff, 2006b, p. 29).

The main questions here revolve around how science or consumer research establishes knowledge, which is entangled in social and natural settings and, in which, new kinds of knowledge constantly need to be (re-)organised and compared with other kinds of knowledge and realities (Jasanoff, 2006a, 2006b). The epistemology in any kind of setting is deeply dependent on contexts, and knowledge is always situated in practices. This means that understanding knowledge co-production does not follow a linear path, but relies on complexity due to the constant necessity of referring to context (Jasanoff, 2006a). The complexity of the production of knowledge lies in its embeddedness in the social, in norms, technologies and institutions. Knowledge co-production is embedded in social practices, in which natural realities and social and individual interpretations of these realities are closely intertwined. It is embedded in norms, paradigms, conventions, and standards, that – as we have seen in the previous chapter – can be the result of natural boundaries as well as being socially constructed. Knowledge is embedded in instruments and technologies, which help co-produce knowledge, but also render it mutable and transportable. And it is embedded in institutions, organisations and states that organise and define how knowledge is established and communicated, and what kinds of knowledge are deemed credible (Jasanoff, 2006c).

Regarding the epistemology of knowledge, Jasanoff (2006b) particularly discusses the relationships the co-production of science or knowledge has with the state, politics and society. This is not only because science accounts for many political decision-making processes but, vice-versa, modern states are often the reason for the emergence of new scientific disciplines, be that in the social sciences and humanities or in hard sciences, such as the environmental sciences. Scientific "facts" and knowledge are deeply interlinked with political context and assumptions about the roles in the state and society. They also "serve – and shape – the modern state's desire for specific forms of order,

control and reassurance" (ibid., p. 33). I would, furthermore, argue that this interlinking should be supplemented by the private or corporate world, whose many processes depend on scientific evidence and knowledge, while it simultaneously drives, finances, and influences many scientific endeavours. The private sector not only participates in publicly-funded research (such as the Human Genome Project (Collins et al., 2003)) but is also increasingly involved in universities or it funds its own external or in-house research groups for different reasons, be it product development or consumer and market research.

The embeddedness and interlinking relationship of knowledge with politics, as well as private companies, can be shown through the example of consumer research and marketing. If we look at how knowledge about consumers is being perceived in marketing, there are different concepts that emerge, depending on the context as well as the worldview that is dominant in the respective marketing tradition. In the more traditional, managerial marketing context, which is oriented towards the worldview of positivism, consumer-oriented knowledge is perceived as a "non-obvious understanding" of the consumer upon which the organisation can act and, by using it, "has the potential to change" the behaviour of the consumer (Laughlin, 2014, p. 76). Ideally, consumer-oriented knowledge should be actionable as it requires follow-up action to enable a change in consumer behaviour. From a positivist perspective, Laughlin (2014) considers consumer data to be the substance of knowledge about consumers, around which data analytics, database marketing and consumer research operate. If we turn to a slightly different context within marketing and shifting towards a more interpretivist worldview – advertisers describe consumer-oriented knowledge as being far less schematic and precise as Laughlin (2014) does, but rather as something abstract, almost mystical. Advertising professionals – to involve a more creative part of marketing – refer to knowledge about consumers as a 'truth' or a feeling about the consumers that needs to be uncovered as it exists deep inside them and is something which they are not aware of (Ariztía, 2015; Bauman and Lyon, 2016).

These different worldviews of what consumer-oriented knowledge should be or do can have an influence on how knowledge about consumers is being produced. In the process of knowledge co-production, a range of actors are involved that affect the resulting knowledge. In the example of Ariztía (2015), these are mainly the clients of the

advertising firm, who have already set the focus of what the consumer-oriented knowledge should entail. For this, a brief which contains an initial framework of the targeted consumers, based on a set of socio-demographic and contextual criteria is used. A similar approach is described by Grandclément and Gaglio (2011), in which a brief is used prior to the commissioning of focus group studies, and it serves as a way of specifying the client's desired consumers. What, at first, is still a vague definition of possible consumers' needs to become a detailed description so that participants for the focus group can be selected accordingly. The brief serves as a technology for standardising the consumer description and provides information about the aim or the objectives of the campaign. It is already the result of internal research and development on the client side. The brief, which sets the direction of the production of knowledge about consumers, is the outcome of multiple negotiations and translations of the different demands of different teams.

In these negotiations, the definition of the consumers that are conceptualised by managers or the marketers needs to be translated into a definition of the consumers for the study designers. In the advertising setting, the brief needs to translate the consumer definition from the commercial/managerial form into a more creative form, in which the initial qualification of the consumers (by the client) is enriched with additional information and framed to fit the creative work of the advertisers (Ariztía, 2015). Although this is only a small section of how knowledge about consumers is being co-produced, it gives a good first impression of the range of actors and techniques/technologies that are involved in the practice within the same setting and context. The co-production of consumer-oriented knowledge is a process in negotiation, by professionals and teams doing the consumer research within the company, within the advertising agency, or within the company producing the good or service.

3.2.3. Stabilising consumer-oriented knowledge – epistemic authority and credibility

Apart from its focus on the embeddedness of knowledge in the social, normative and material in the creation of knowledge, the interactional approach of co-production puts emphasis on the stabilisation of that knowledge enmeshed in these influences and, by extension, on how individuals determine the authority and credibility of knowledge, who is to be trusted and on what grounds. It is about how epistemic authority is established in

science and knowledge production (Hardwig, 1991; Origgi, 2004, 2007), in which "doing science" merges into "doing politics", as many effects are involved to make knowledge trusted and believable (Jasanoff, 2006b, p. 29). Within institutions, the decision on what is considered scientifically right and wrong can be contested depending on ontological and epistemological views, and unpacking these practices and views contributes to an understanding of how knowledge is being stabilised in these institutions (Jasanoff, 2006b).

The focus on how knowledge is made credible and trustworthy in the process of stabilisation of that knowledge closely connects to the concept of epistemic authority which, on the one hand, is attributed to individuals who are in a position of authority. The knowledge they communicate normally tends to be trusted and believed more than the knowledge others communicate (Hardwig, 1991; Origgi, 2004). On the other hand, and closely linked to the first, epistemic authority also relates to other factors contributing to the credibility of and trust in knowledge. Here, the way knowledge is communicated and stabilised plays an important role (Jasanoff, 2006c; Origgi, 2004).

As new forms of knowledge and new methods of knowledge production emerge, the legitimation of knowledge and the establishment of its credibility need to be re-enacted. New forms of knowing, such as through the shift from more traditional forms of knowledge production to big data analytics, often depends on how research is portrayed internally and externally as a solution to the problem in question. In this process, of portraying knowledge and its methods of production, a simplification process takes place that is required to make knowledge malleable, transportable or displayable (Ezrahi, 2006; Jasanoff, 2006b). As Ezrahi (2006, p. 255) explains, this is "connected with the shifts from knowledge to information", which, essentially, reduces knowledge to its core information. This makes knowledge more transportable, objective and conceals "the interpretive layers and normative commitments" that underlie its corresponding knowledge (ibid., p. 257). Taking the example of big data analytics, in most situations, it does not make sense to communicate or present the entire analytical operation. Instead, selections are made by the data analysts as to what to communicate further. In a sense, data analysts not only produce datafied knowledge but they also produce a certain story relating to that knowledge (Jasanoff, 2017). This is something that is often overlooked, but which is, however, crucial as the communication of knowledge is a "creative and

richly interpretive process" (Origgi, 2004, p. 69), one which can influence the trust and credibility of the knowledge and, as a result, the sense-making of individuals.

In the marketing context, this process of stabilising knowledge, making it credible and establishing epistemic authority can also be observed. An example of an ethnographic study by Dubuisson-Quellier (2010) at a research and development and marketing department of a French food company can illustrate this. First, Dubuisson-Quellier discusses how the process of co-producing consumer-oriented knowledge is influenced by social, normative and material aspects. Descriptions of consumer preferences for new products are influenced by quantitative surveys as well as panel discussions, by secondary data as well as tailored market research, by tasting tests for consumers as well as by experts within the company. All these lead to constant definitions and redefinitions of the targeted consumers, their preferences, and the associated products: in short, to a wide range of knowledge about consumers.

The interesting aspect Dubuisson-Quellier (2010) discusses is that, depending on the method used to discover the consumer preferences, there is a different impact of how this knowledge is negotiated and interpreted within the organisation, and how it can contribute to the production of consumer-oriented knowledge. Quantified consumer research is especially preferred as it requires little interpretation and is easily circulated among and understood by other departments and teams. Qualitative data, however, is often confronted with scepticism. In response to this, the author shows that there are different forms of knowledge that are made more (or less) credible, which may depend on the ease of communicating and circulating it. Another example from Dubuisson-Quellier's (2010) research concerns the tastings organised by the company, where targeted consumers can try (new) products. The answers provided by the consumers are generally considered difficult to interpret due to the openness and qualitative nature of such tests. Interestingly, tastings by experts from within the company – heads of all the departments and the chief executive officer – are regarded as an indispensable part of the qualification process of the products and consumers. These tastings have several aims and involve the experts taking on multiple roles. With the possibility to assist in the production process, define a purchasing policy, or contribute to the market positioning, the professionals both provide technical feedback while, at the same time, provide input as consumers (Dubuisson-Quellier, 2010).

The processes described by Dubuisson-Quellier show that not only are many different actors involved in the production of insight but also that not every actor and every type of knowledge holds the same credibility when it comes to what is considered acceptable knowledge for contributing to marketing. In this example, quantified knowledge, as well as expert knowledge, are regarded as more credible types of knowledge and hold more epistemic authority. As a result, these kinds of knowledge are more easily stabilised in the organisation and seen as more legitimate, whereas the insight provided by the consumers directly is often ignored or considered illegitimate. The latter is perceived to be more subjective, thus difficult to interpret and less reliable. Social practices and identities – in the form of hierarchies and expertise – have an influence on the credibility of the knowledge.

Another example, by Grandclément and Gaglio (2011), investigated the use of focus groups to establish consumer-oriented knowledge and consumer segmentation. In this, the focus groups are described by marketers as an ideal method to observe and know the consumers "in the flesh" (ibid. 2011, p. 88), and the groups are seen as a credible source of knowledge. The reason why consumer-oriented knowledge from focus groups is preferred is that marketers are not physically present in the room where the research takes place but are hidden behind a mirrored window. The expectation is that this distancing and the observations provide an opportunity to see how consumers behave with minimum intervention from marketers and to give an impression of objectivity to the research – despite the artificiality of the situation of a focus group. The goal of this practice was to gain credible knowledge about consumers, for "marketers to 'step into consumers' shoes" (ibid. 2011, p. 87), allowing them to construct a description of the consumer that mirrors them. This example also shows how the materiality – in this case, the mirrored window – can contribute to the stabilisation and epistemic authority of a certain kind of knowledge.

3.2.4. Representing reality, representing consumers – ontological questions

Turning to the ontological questions that are addressed in co-production, the focus here is on how representations of the world and of 'reality' are produced, as well as taken up, by others (Jasanoff, 2006b, p. 41). The ontological, constitutive strand of co-production goes back to the initial work of Bruno Latour's laboratory studies. In this, he not only argued that science, technology and society are constantly co-produced but that the dualism of nature/society is constructed and does not actually exist (Jasanoff, 2006b;

Latour, 1988, 1993). This, ultimately, led to his development of ANT and the equivalence of human and non-human actors. In the anthropological sense of Latour, it is the thick description by analysts, by which the hybridity of nature and culture, the co-production of science and society, will eventually be made visible again, which uncovers what renders the stability in the networks of actors (Latour, 1993). Latour's provocative ideas, however, gave little information about why scientific facts, knowledge, communities and networks are stabilised: why some kinds of knowledge are contested in some contexts while others are not, or how ideologies, beliefs, values and history come into play in the science-making process (Jasanoff, 2006b). Nonetheless, Latour's work has initiated a range of complementary ideas that go further than being just a description of scientific processes, trying to include an explanatory account, as well.

Much of the recent work about co-production elucidate Latour's constitutive idea of the process by contrasting the production of science and society to the role of representation. Using the analogy of the nation-state and the work by political scientists Benedict Anderson and James C. Scott, Jasanoff concludes that not only the manner of how science is produced is important, but also the representation and communication of knowledge. A modern nation state can be partly seen as a political community that is imagined and its members are influenced by its powerful symbolic representation. The same applies to scientific networks that are successfully held together by a common set of ideas and beliefs, and are more than just the totality of their *actor-networks*, as described by Latour (Anderson, 2006; Jasanoff, 2006b, p. 25f.). James Scott's analogy (1998) goes in the same direction. Scott describes how policy planning generally requires a representation of the population that is *idealised* and *standardised*, in short, that made them commensurable, in order to adapt the planning accordingly. Jasanoff (2006b, p. 26f.) argues that scientific processes often work in the same way: simplifications and standardisation for classifications purposes are a necessary tool for making scientific measurements (see also Hacking, 1986).

These standardisations – in policy planning as in science – are influenced by normative and cultural elements of both planners and researchers, inherently impacting the measurements, commensurations and outcome. Representations of the natural and the societal play an important role in Jasanoff's (2006b, p. 41) idiom of the co-production. She considers the "making" of representations one of the core instruments of co-

production in stabilising what is known, as well as how it is known. Thus, analysing how these representations come to be is another important aspect of the co-productionist idiom. Here, questions that are addressed in relation to the making of these representations include how they are influenced by the historical, political and social, what is used to inform representations, and how these representations are taken up by others.

From a consumer research and marketing point of view, the ontological questions of co-production Jasanoff (2006b) highlights are equally adaptable. The empirical examples of consumer-oriented knowledge co-production in marketing show how perceptions and representations of products, markets, competitors and consumers are constitutive for the stability of the networks the marketers are involved in and engaging with. In the example of the French food company, the experts are defining the tastes of the consumers and stabilising the characteristics of the products intended for sale – at least for a certain period and within this specific context (Dubuisson-Quellier, 2010). Furthermore, the way in which the conceptualisations of the consumers are established in marketing requires a constant transformational process in which minimalist perceptions of consumers are created to be targeted in advertisements (Ariztía, 2015). As in policy-planning and in scientific communities, it is to be assumed that in this (semi-)scientific community of market research and marketing, normativities, beliefs, ideologies and history contribute to these processes, granting marketers "the power to create new ways of "reading" people" (Jasanoff, 2006b, p. 28). As Ariztía (2015, p. 151) explained, marketers create "an imaginary collectivity of consumers" through the assistance of knowledge about consumers.

However, the imaginary collectivity or representation of consumers do not necessarily accurately represent reality. Instead, they are created for certain purposes and contexts. For example, marketers are not necessarily interested in knowing what the consumer looks like in their entirety but are rather interested in knowing how consumer-oriented knowledge can serve to achieve marketing goals (Grandclément and Gaglio, 2011). Similar to what we have seen in the discussion on the commensuration and classification of consumers (Bowker and Star, 2000; Fourcade and Healy, 2017b [2013]), the description of the consumer is always only a partial representation of their entirety and can never be a mirror of reality. This artificiality is what Grandclément and Gaglio (2011, p. 89) mean when they refer to the description of the consumer by marketers as a

“construction of the consumer”: a construction of the consumer that is heavily entangled in practices performed by different actors, depending on techniques and materials, as well as a knowledge that is negotiated and weighed regarding its credibility. In the end, the construction of the consumer is solely an imaginary consumer conceptualisation – distilled and limited to just a few defining marketable traits.

Furthermore, the research by Dubuisson-Quellier (2010) also shows that different stabilisations of consumer-oriented knowledge and different consumer conceptualisations emerge during the process of consumer research or during R&D. The author calls this process the qualification of consumers and products. In the organisation she researched, the co-production of consumer-oriented knowledge did not produce one single qualification of the consumer but, instead, multiple qualifications were produced, both local and temporal. These qualifications need to be reproduced and re-defined over and over again, which means that the stabilisation of consumer-oriented knowledge is constantly ongoing. And, simultaneously, this means that the representation of the consumers is not a single entity but, rather, multiple and temporary representations emerge that all co-exist in the same way Mol (1999, 2002) has described multiple ontologies.

The co-productionist idiom becomes the more interesting for research when the process of sense-making in question, in this case consumer-oriented knowledge co-production, undergoes changes, innovations are introduced and the boundaries of epistemic authority change. Marketing, as well as research in general, might currently be in the process of one of the biggest changes in recent years. The introduction of big data analytics has changed, and will probably further change, the process of consumer research in marketing. Big data analytics has the potential to considerably transform the way knowledge about consumers is co-produced and used as a basis for decision-making (Erevelles et al., 2016). In times when big data analytics and data-intense science is seen as the “fourth paradigm for science” (Hey et al., 2009), it has the potential of becoming an innovation which might impact the stabilisation of what we know and how we know it, and thus of how identities, institutions, discourses and representations are made (Jasanoff, 2006b). As Law and Urry (2004, p. 397) note, “different methods produce different and often inconsistent results”, which also means that “different research

practices might be *making multiple worlds*" and creating multiple realities (Law, 2008; Mol, 2002).

The transformation of consumer research towards a big data-centred method thus demands a closer analysis looking at how different results might be produced. As I have discussed in the previous chapter, marketing as a discipline is both influenced by managerial, positivist worldviews as well as interpretative worldviews that are more attributed to the social sciences (Layton, 2016). Big data analytics, stemming from computer science and requiring a strong mathematical understanding is a research method that – methodologically – is connected to a realist ontology, objectivist epistemology and positivist paradigm (see also Moon and Blackman, 2014). Upon increasingly using big data analytics as a method for consumer and marketing research, it is to be expected that the dominant worldview within marketing further moves towards one of digital positivism – with the risk of drifting into an ideology of dataism (van Dijck, 2014). This raises thus the question that I address with my research, asking how and why the co-production of datafied knowledge, through big data analytics, leads to a different kind of knowledge, producing different kinds of consumer representations?

3.3. Researching the co-production of consumer-oriented knowledge through big data analytics

While the previous examples I have included have largely focused on how knowledge about consumers is co-produced through qualitative and quantitative consumer research, the focus on big data analytics has been of little concern so far, with some notable exceptions. Sunderland and Denny (2011) have, for example, shown how data available at organisations establishes a predefined conception of the consumers, which leads to further qualitative and quantitative research on them. In this procedure, this already predefined conception of the consumer is stabilised through establishing consumer segmentation upon which marketing measures are taken, similar to what has been described by Grandclément and Gaglio (2011).

Pridmore and Lyon (2011) have discussed more specifically how marketing and, in particular, loyalty marketing has at its core not only the collection of vast amounts of data but also uses this data to construct or assemble "digital representations of consumers" (ibid., p. 115). These representations include their socio-demographics along with their

needs and desires or behaviour. While not yet fully embedded in the big data analytics literature, the authors show how CRM practices, which are at the core of one-to-one marketing, may be based on large consumer databases, serving to assemble consumers into categories. Consumer data is a means of managing and sorting individuals. Pridmore and Lyon (2011) stress the political dimensions of these marketing practices, which not only deeply invade private everyday-practices of individuals but also have the potential of reinforcing an idealised conceptualisation of how consumers ought to be. Consumers not acting accordingly risk being excluded from participating in consumption practices.

Turning more specifically to the practical use of big data analytics to create consumer-oriented knowledge, the existing research is sparse, with some notable exceptions. Most of the earlier research on the use of big data analytics for the co-production of consumer-oriented knowledge has focused on questions concerning why an increased desire for the implementation of big data analytics emerges in organisations – if at all – and how it changes marketing decision-making (Quinn et al., 2016) or consumer segmentation practices (Pridmore and Hämäläinen, 2017). Similar to how focus groups hold the "place of quasi-absolute supremacy among qualitative methods in market research" due to leading to a more objective kind of knowledge about consumers (Grandclément and Gaglio, 2011, p. 87), so has the availability of digital data created an ever-increasing demand to salvage this knowledge.

As Quinn et al. (2016) discuss, big data analytics changes not only the way in which knowledge about consumers is created but, overall, leads to a demand for ever more knowledge about consumers and their behaviour. However, despite these promises and requirements, the implementation of big data analytics to enrich consumer-oriented knowledge is not an easy process. Instead, big data analytics leads to a disciplinary crisis in marketing due to the difficulties that emerge from handling these large amounts of data and the technological development surrounding it (Quinn et al., 2016). Issues emerge with people in organisations not knowing how to process data and gain the knowledge about the consumers they require or want out of the analysis. Often, the problem lies not in having access to the data, but rather in knowing what to do with it, how to work with and interpret it, which has been described as "paralysis by analysis" (ibid., p. 2117).

Changes in marketing settings are not adopted quickly and there is a discrepancy between theory and practice (Pridmore and Hämäläinen, 2017). In theory, big data analytics is depicted as leading to a higher and better kind of consumer-oriented knowledge. In practice, however, while marketers might use big data analytics for this, a lot depends on the initial hypothesis to segment consumers, which is often based on existing conceptualisations of them as both Grandclément and Gaglio (2011) and Sunderland and Denny (2011) have discussed. Segmentations of consumers are not so much emerging out of big data analytics but, rather, are the production of segmentations that analysts are trying to translate into consumer databases (Pridmore and Hämäläinen, 2017). Thus, while big data is expected to bring forth great opportunities in gaining knowledge about consumers and other market actors and informing marketing strategies, the handling also brings forth great difficulties due to organisations being "technical ill-equipped [and having] a lack of understanding that is limiting how creatively the data can be used in practice" (Quinn et al., 2016, p. 2123). As a result, big data analytics and digital consumer-oriented knowledge creation are often outsourced, causing a separation of the consumer-oriented knowledge from the marketing team. In such circumstance, marketers are less and less in control of what kind of knowledge about consumers to gain, where to focus on and how to use this knowledge.

Both these studies are interesting in so far as they were among the first to look specifically at the implementation of big data analytics and its use for creating consumer-oriented knowledge and for segmenting consumers. However, as discussed initially in Chapter 2, when defining big data analytics, the technical development progresses quickly, with not only the capabilities of the technology changing, but also its potential level of implementation in organisations. Looking at later research on this topic, such as by Ariztía (2018) or Darmody and Zwick (2020), we can see that the acceptance of big data analytics in marketing seems to have progressed further. While some of the issues presented in the previous research have persisted, such as the shift in control over knowledge about consumers from marketers towards data analysts, new issues have emerged. Darmody and Zwick (2020), for example, argue that the narrative in digital marketing revolves around the use of digitalised and highly detailed consumer conceptualisations to provide more relevant offerings to those consumers. While the highly detailed consumer conceptualisations are still considered to be an "epistemological vision" (ibid., p. 9), the

conceptualisations will not be an exact representation of the individuals but rather will be a representation of how marketers hope they will turn out to be. Through the nudging and manipulating effects of marketing, it is not so much about knowing what the consumer is doing, but rather about shaping her or his behaviour.

Finally, Ariztía's (2018) empirical example partially reveals some of the issues of co-producing datafied knowledge about consumers and the subsequent conceptualisation of consumers in marketing settings. In particular, the study shows how the mundane practices behind these operations reveal epistemological and ontological politics (Mol, 1999), and how practitioners decide on what are considered to be relevant data and variables for analysis. These mundane practices are particularly revealing when practitioners are cleaning the initial datasets as this is when major decisions are made regarding what will be included and what excluded, involving a transformation of the dataset. In this phase, decisions can be made regarding the data, such as leaving out half a database, not because it better represents the real consumer, but rather because it contributes to the practice of analysis – and, thus, is more "a process of internal problem-solving" (Ariztía, 2018, p. 10).

The construction of a workable database for big data analytics revolves around recurring failures and errors. This means consumer databases require continuous maintenance and repair to produce consumer-oriented knowledge. These moments of repair, the way data analysts resolve errors in the databases, are moments at which critical decisions are made in terms of what is included and excluded from the database, while also providing a moment to reflect on previously-made decisions. As a result, the data that emerges from the database and is used in big data analytics does not do so because it is an exact representation of reality or the best copy of reality. Instead, that data is the result of a plethora of different operations by carried out by different people, and it is the data that works best in these operations.

3.4. Conclusion

What all these examples have in common is that they help us to better understand the practice of co-producing consumer-oriented knowledge and they particularly show how the epistemology behind consumer-oriented knowledge is influenced by many sociotechnical assemblages, as well as how the use of knowledge about consumers serves

to create specific and multiple ontologies about the consumers that are distant from the real consumers. And while some examples have started to look at how big data analytics contributes to the co-production of consumer-oriented knowledge, so far, no research has delved deeper into the epistemology of datafied consumer-oriented knowledge, nor how that datafied knowledge further contributes to the stabilisation of knowledge and the conceptualisation of consumers.

As shown, the concept of co-production can help here. With markets and marketing, as well as big data, already considered to be sociotechnical assemblages, the process of producing knowledge should be considered in its entirety, entangled in social, normative, cognitive and technological procedures. Furthermore, as Jasanoff (2006b) has highlighted, the practices of stabilising knowledge, through epistemic credibility and authority (Jasanoff, 2017), reveal much about why certain kinds of knowledge are preferred by its user over others, how different worldviews in marketing shift depending on the methods used to gain consumer and market knowledge, and how this leads to different representations of "realities". This might help explain why traditional consumer research is being considered an unreliable source of knowledge about consumers, only capable of producing an imaginary consumer conceptualisation, as well as if and why big data analytics is being seen as the new method of acquiring exact knowledge about consumers, potentially producing a data double consumer conceptualisation, as expected in the mainstream marketing literature. And finally, the co-production also helps to address whether the risk of relying on a postulated consumer conceptualisation in marketing is considered. In summary, the concepts of co-production help to answer the research question of:

> How and why do marketers conceptualise consumers differently when using big data analytics in comparison with traditional market and consumer research methods?

4. Methodology

4.1. Introduction

In this chapter, I will discuss my methodological approach and present the research design that I have developed as a means of answering my research question. I will start by laying out the philosophical perspectives, which are founded in the discipline of Science and Technology Studies. In the second section, I will discuss the process of developing my research design. The third and fourth sections include my methods for data collection and data analysis. Here, I will explain the choices I have made, providing a rationale for my empirical research. Finally, the last section gives an overview of the ethical considerations required for my study.

4.2. Philosophical perspectives – epistemological and ontological choices

Empirical research is always guided by one's philosophical perspectives in terms of the *ontology* as "what belongs to the real, the conditions of possibility we live with" (Mol, 1999, 74f.), and the *epistemology* as how knowledge is created within the reality or realities of the world (Moon and Blackman, 2014). These philosophical perspectives set the path for the research question and the choice of methods for the empirical research. In the following sections, I will explain my methodological choices in more detail, according to the guidance found in Science and Technology Studies and Critical Marketing Studies. These disciplines themselves feature enquiries into the practices behind doing research in marketing and aim to answer questions on the *epistemology* and *ontology* which lie at the foundation of these practices. As I have discussed in the previous two chapters, the discipline and practice of marketing is influenced both by managerial and economic principles, as well as approaches found in the social sciences. Hence, why there is the potential of having two contrasting ontologies – realist and relativist – and with this two contrasting research paradigms – positivist and interpretivist – lying at the core of consumer and market research. This also means that by researching the epistemological and ontological underpinnings of consumer research, realising one's own approach towards the matter of research is of importance and should be reflected upon (Neuman, 2014; Woolgar and Lezaun, 2013). It is of special importance to highlight the core differences between opting for an STS approach and a more classical sociological approach, such as social constructivism. While both are connected as STS, originating

from social constructivism and the sociology of scientific knowledge (SSK), upon developing STS as a discipline, some crucial differences between the two approaches have emerged (Law, 2008).

4.2.1. The theoretical framework of Science and Technology Studies and Critical Marketing Studies

STS examines the discovery of knowledge in scientific research from a social science perspective. The discipline has expanded and developed greatly and has been influenced by far more than just sociology, but also philosophy, anthropology, semiotics, post-structuralism and feminism. The most important influence in STS has been, however, its 'technological turn' (Woolgar, 1991), which also marks a major difference from the discipline of sociology and, specifically, the social constructivist theory. As Law (2008, p. 634) notes, the notions of 'the social' as well as of 'construction' have disappeared as analytical categories in STS due to this turn to technology. Instead of focusing on macro-sociological grand narratives, STS has been (and still is) interested in the praxis of scientific and technological undertakings, often in the form of case studies. This focus leads to considering primarily the relational and processual aspects of science and technologies, in which the social has been put on the same explanatory level as the material/technical. Thus, unlike the social construction (of knowledge/technology), where the social is seen as a constituting force in the creation of knowledge and technology, STS sees the social and the technological in a continuing process of relationships which "are enacted, enacted again, and enacted yet again" (Law, 2008, p. 635). Or, to put it in the terms of the co-productionist idiom, STS considers science and technology as a process of cognitive, material, social and normative procedures and settings, in which neither the social nor the natural are given primacy (Jasanoff, 2006c).

With STS focusing on how science is done in practice, the discipline has a strong focus on the epistemology behind these practices. It is a focus on the different representations of realities by individuals involved in the practices of science and technology and how these representations are enacted and performed in a social and material way. It is about "doing realities" (Law, 2008; Woolgar and Lezaun, 2013). This focus is crucial for my thesis and the research into the epistemology of consumer-oriented knowledge. With big data analytics, the practice of doing consumer realities, of enacting and performing consumer representations or conceptualisations, is experiencing a significant change in

method and methodology. As other researchers have addressed already, big data analytics brings forth epistemological problems (Floridi, 2012) that require closer scrutiny and a better understanding in the form of empirical research (Agostinho et al., 2019; Ariztía, 2018).

However, the focus by STS on science in practice and on the different representations of the realities is not solely epistemological but raises important questions about the existence of realities, which brings us to the ontological. The view of ontology is an empirical or practical view, with not so much attention paid to 'what there is', but rather on how different realities are enacted (Woolgar and Lezaun, 2013), and seeing reality as a multiplicity (Law, 2008; Mol, 2002). Instead of only being concerned with "whether representations of reality are accurate" (Mol, 2002, viii), as is the case with a sole focus on the epistemology, the reference to ontology asks questions about how realities are shaped and enacted through "mundane practices" (Mol, 1999, p. 75).

A consequence of the ontological multiplicity is that, in a continuous process of enactment and re-enactment of realities, these realities compete against each other, where "some will be preferable to others" (Law, 2008, p. 636). Especially here, where the topic of the research question revolves around the method of doing consumer research, a central concern of STS is on the role of "methods (…) (as) routinised practices that do reals and representations of reals" (ibid. p. 638). Considering how in research, different ontologies can dominate depending on historical, cultural and/or organisational developments, different methods will be favoured, leading to the resulting knowledge being trusted more than other kinds of knowledge. With big data analytics having its roots in the field of computer science that is dominated by mathematical and informatics logics, the ontology, epistemology, and philosophical worldview prevailing in that field will also transcend into the practice of using big data as a method for consumer and market research.

In marketing and in consumer and market research, both a more positivist as well as an interpretative paradigm can be found, as I have discussed in the previous chapters (see also Layton, 2016). Considering that marketing-as-management is the more prominent approach towards marketing, in which a more realist ontology and positivist worldview are already in place, the emergence of big data analytics as a novel research method in this field will be crucial to observe and research. Certainly, as big data analytics has the

potential of leading to an ideology of dataism as positivism in its most radical form. In this, STS provides the theoretical and methodological tools to research the effects of big data analytics on the ontological and epistemological underpinnings of marketing. A philosophical approach which is critical for understanding the role of big data analytics practices as a means of creating consumer representations through which multiple realities of the consumers are enacted in marketing practices.

4.2.2. A pragmatic turn

While much of the research on the practices in marketing and consumer research is influenced by the general concepts of epistemology and ontology from STS, there is, additionally, an overlap with American Pragmatism. Notably, regarding the emphasis on the practical endeavours of marketing and consumer research, a connection between Critical Marketing Studies, STS and Pragmatism becomes apparent (Chakrabarti and Mason, 2015; Davies, 2015; Overdevest, 2011). Pragmatism as a social theory emerged in the late 19th and early 20th century in the US, largely led by Peirce (1878), James (1897) and Dewey (1929). To a similar extent as in STS, Pragmatism considers the experience or action of individuals as a crucial concept for organising and 'understanding' reality in the sense of understanding what occurs in the world (Chakrabarti and Mason, 2015). Pragmatism considers that (individuals') practices are constitutional for the world and should also be made a focus of research – as many STS approaches do (Morgan, 2007). In this, Pragmatism recognises that knowledge is not a representation of independent facts (in the realist sense) nor entirely constructed, but rather "co-authored products of agents and their natural, social (and technological) contexts" (Dorstewitz and Kremer, 2016, p. 5), an approach which is closely linked to Jasanoff's (2006b) idiom of the co-production of knowledge.

There are, however, some differences between the approaches which should be acknowledged here to avoid a mix-up and a blind interchangeability between both philosophical perspectives. From a Pragmatist view, the epistemology of individuals, in terms of how they make sense of the world and how they create knowledge, is always "issue-driven" (Overdevest, 2011, p. 536). This means that actions mainly arise from problem-solving, in that problems lead to necessary adjustments. Individuals are required to constitute knowledge (or reconstruct knowledge) to make those adjustments. In this, as Pragmatists argue, the actions will create re-actions, upon which individuals need to

constantly adapt their actions and "come to terms with new and different aspects of reality" (ibid.). Although, here again, the similarities between STS and Pragmatism shine through (such as the various aspects of reality or the multiple realities), within STS, the focus on practices is by no means solely issue-driven or aimed at problem-solving. While they are analytically interesting – from an STS perspective too – they are not considered to be the sole driver of sense-making. Instead, STS has a strong focus on the constituting and stabilising practices of science and knowledge production (Jasanoff, 2006b) as they provide insight into the stabilising forces of change.

The contribution of the Pragmatist paradigm for my thesis lies in its approach towards empirical research, and, specifically, how it addresses methods and methodology. As Pragmatists put the practices of epistemology and action in the foreground, they reject the requirement of choosing between realism and relativism as there exist different versions of reality. This means, for Pragmatist researchers, that the questions of the research itself should be the main driving point, where equal attention is paid to epistemology, methodology and methods, without having a "worldview", in the paradigmatic sense, deciding on the methodological choices (Morgan, 2007). This has contributed considerably to an increasingly research "problem"-driven methodology. Instead of focusing on deciding what worldviews account for what research methods, researchers should stress what the research problem is about and, from this, try to find as many ways as possible and use as many methods as necessary to acquire sufficient knowledge to understand the problem (Creswell, 2014). In short, empirical research should be driven by context and methods chosen by 'what works', instead of by philosophical positions (Greene et al., 1989; Moon and Blackman, 2014).

With the focus on using the method that works for the specific issue of the research, Pragmatism is often seen as the main theoretical approach for mixed-methods research designs (the combination of qualitative and quantitative methods). This does not mean, however, that Pragmatism necessarily requires a mixed-methods approach. Nonetheless, it argues for a diversity of methods to understand a given problem, for which also triangulation can be suitable. Concerning the present research, a within-method triangulation, along with data-triangulation, can contribute considerably to the understanding of the application of big data analytics in consumer research and

marketing, providing a more holistic image of the matter in question, as I will show in the next section (Kuckartz, 2014; Yin, 2014).

4.3. Developing the research design – facing difficulties and turning them into possibilities

After having done a thorough literature review and having formulated my research question, the next step in my research process was to develop my research design. In the next sections, I will discuss the steps and reasoning of this process. The development of my research design has a twofold purpose. First, it serves as a preparation for the upcoming research since it requires a thorough reflection on the different steps of the research and enables anticipation of foreseeable problems. At the same time, the developed research design also serves as a guideline and handbook during the research phase (Yin, 2014). During my empirical research and data collection, it was helpful to be able to refer to my research design and check whether the research I was doing was still in line with what I had planned in the first place.

This is not to say that the research design is a rigid construct that should be followed rigorously. Depending on the circumstances and the development of the empirical research, the research design can and should be adaptable to accommodate potential changes that might occur – which can be of a practical, as well as theoretical, nature. This also means that, although an initial research question has been set out which guides the development of the research design and the conducting of the empirical research, this research question can also be adapted to a certain extent (Maxwell, 2009), as was the case in this research. An initial research question, as well as an initial research design, were developed, and both of which have been adapted over time. This development of my research design will be described in the following sections, and I will give a rationale for the steps taken, in particular where changes and adaptations occurred. I will, furthermore, describe the sources of the data that have been used.

4.3.1. Developing the research design:

As case studies are not just one of the main research approaches of Science and Technology Studies (Law, 2017), but also for researching the implementation and effects of information systems in organisations, my choice for doing case study research on the co-production of datafied knowledge in marketing was made early on (Darke et al., 1998;

Hughes and McDonagh, 2017; Kuckartz, 2014; Yin, 2014). Further advantages of case study research are that it is a holistic, exploratory approach towards the research subject or the unit of analysis, relying on a triangulation of multiple sources of data and multiple methods of data collection (Kuckartz, 2014; Yin, 2014). Additionally, it allows for research that is situated within a "real-life setting", producing a more rigorous understanding of the case in its practical application (Gerring, 2004; Harrison et al., 2017).

While different forms of case study design exist – single case study, multiple case study, cross-case study (Yin, 2014) – I initially opted for a single case study, based on methodological and practical reasons. From a methodological point of view, the study is designed as an inductive, explorative piece of research (Hartley, 2004). As mentioned in the previous chapter, critical research into the implementation of big data analytics in marketing is sparse, as is research on the impact of different kinds of knowledge on the conceptualisation of consumers. As such, an explorative research design in the form of an in-depth case study was considered to be a good approach to get an initial understanding of the practices and issues surrounding the use of big data analytics for customer-oriented knowledge in organisations (Harrison et al., 2017). The unit of analysis – i.e., the case – involved the process of co-producing datafied consumer-oriented knowledge through big data analytics, which covered the entire process of doing big data analytics, as well as stabilising and using the datafied knowledge.

From a practical point of view, researching the practices of data analytics for marketing purposes can be difficult, as it can entail reviewing sensitive and strategic operations of the company. Furthermore, the research would entail wider access to the company, with many interviews with employees, access to documents and records, as well as participant observation. Not only would getting such a level of access in multiple organisations be challenging but also the resources, in terms of time and manpower, would make such research difficult to complete within the timeframe of the PhD. These were the main considerations for aiming for a single in-depth exploratory case study in one organisation.

While planning the research design, considerations regarding the possible difficulties of researching organisations and their implementation and use of big data analytics for marketing purposes had been established beforehand. These included the creation of an

access negotiation strategy plan, which entailed briefs and information material on my research, as well as specific approaches for identifying and contacting gate-keepers to the research sites (Eyben, 2005; Johl and Renganathan, 2010). Recruitment for the case study was done through multiple channels. This involved attending industry conferences and speaking with representatives of companies, where I presented my research and tried to recruit them for a first interview. These representatives either where head-marketers or heads of data analytics or data science teams. Other recruiting channels were personal contacts, who had connections in organisations in general or with people working in the area of data analytics and/or marketing. Finally, possible gatekeepers were identified and approached through professional social networks, such as LinkedIn. Despite this broad approach, the access negotiation process was even more challenging than expected.

4.3.2. Adapting the research design

The initial idea of having one large in-depth case study was followed for a long time, despite the multiple setbacks of not acquiring more thorough access to the organisations. The method of attending industry conferences and recruiting potential candidates proved fruitful initially. Many participants showed a lot of interest in the research upon initial approach, and contact details were shared easily. However, the step of gaining further access was often a lot more difficult. In many cases, it proved difficult just to get a first interview going, despite an initial confirmation for an interview at the conference. Even though several emails were exchanged, and information material was provided, the responses often faded after potential candidates were asked for a concrete interview date. In other cases, were initial interviews were scheduled and conducted, the contact broke down after this first interview. Regardless of an apparent interest in the study, further access to the organisation was denied. Similar experiences resulted from the other recruiting methods, where often the contact, through third parties, was made but actual interviews or further contact to organisations often failed to be established.

This created the necessity of making several adaptations to the research design, while still allowing for a robust empirical and explorative study. A first approach was to extend the research beyond a single case study. Since negotiations with several organisations were ongoing, it became apparent that the main obstacle was the necessity for in-depth access. As such, in some of these situations, access for several interviews within one organisation proved to be easier. This meant that I had to change to a multiple case study approach.

The main change was to not ask for in-depth access in each case, but to arrange for just one or two interviews to be conducted. These interviews were treated as key-informant interviews, which are interviews with "strategic informants" – people who "as a result of their personal skills, or position within a society, are able to provide more information and a deeper insight into what is going on around them" (Marshall, 1996, p. 92). Key-informant interviews have a long tradition in the German-speaking branch of the SSK by Alfred Schütz, Peter Berger and Thomas Luckmann, in which the strategic informants are seen as *experts*. The classification as an expert mainly relates to the participants' specific expert knowledge relating to a professional area or an occupation where this expert knowledge is used in practice (Littig, 2009). Meuser and Nagel (2009) describe the expert as being in a special position due to having access to a kind of professional knowledge, which closely relates to Marshalls (1996) description of the key-informant.

The contribution of the key-informant to my research lies in her or his expert knowledge of the processes of co-producing datafied knowledge, as well as the use of that knowledge to conceptualise consumers. While at first, it might appear that the expert knowledge solely relates to the technical expertise in the domain of big data analytics or marketing, there are two other aspects of expert knowledge which are important here. On the one hand, there is the process or practical knowledge, which relates, for example, to practical operations and day-to-day interactions with big data analytics or the datafied knowledge. On the other hand, the expert knowledge also expands to the interpretative knowledge of the key-informant, thus her or his subjective perception of the operations of big data analytics in their organisation (Bogner and Menz, 2009). Overall, key-informants have access to a highly specialised kind of knowledge, highlighting their relevance as participants for my research and the case studies (Bogner and Menz, 2009; Marshall, 1996). Key-informants are able to provide an in-depth view on the use of big data analytics, the co-production of datafied knowledge, as well as the conceptualisation of consumers in their organisations. All this can be done while not focusing on a single organisation or case anymore. A further advantage of opting for key-informant interviews was that another recruiting method was enabled for single interviews, largely through the assistance of a university-led professional network.

These changes still allowed to conduct explorative case studies, as the unit of analysis and the context of the case – the process of co-producing datafied knowledge embedded

in marketing settings – remained the same. The in-depth character of the case study might have decreased slightly as the research was no longer being conducted within one organisation. Additionally, some sources for data collection could not be pursued, such as participant observation in the organisation in the form of work-shadowing. As the theoretical foundation of my research relies on STS and ANT, in which in-depth case studies with participant observation are a key method to research the practices of interest, not being able to rely on these methods and sources of empirical data proved to be a challenge at first.

One measure that was taken to counter the lack of participant observation data on the practices of doing big data analysis was to rely on key-informant interviews. As explained, their strategic role within their respective organisations enables the participants to provide important insight into the practices of using big data analytics, and the wider processes that are involved. As a result, detailed descriptions around the practice of implementing big data infrastructures, analysing big data, communicating datafied knowledge as well as utilising this knowledge for consumer conceptualisations were created.

Moreover, the adaptation of the research design led to conducting multiple smaller case studies, in contrast of having only a single case as initially planned. This also helped to compensate for a lack of participant observations, as this opened my research to compare several cases, offering an understanding of diverging practices regarding the implementation of big data infrastructures and using big data analytics as a means for analysing consumer data and creating conceptualisations (see also Hughes and McDonagh, 2017; Stake, 2005). Thus, although the in-depth character of the research has diminished slightly without the thick description resulting from participant observations, conducting and analysing multiple cases has compensated this by adding a more diverse view on the unit of analysis.

Finally, another source for data collection was enabled due to the adaptations to the research design, which further compensated the lack of participant observation data in the case studies. As I was visiting industry conferences as sites to recruit participants for my research, I started looking at opportunities to also use these conferences as fields of study. Many of the topics that were being discussed there appeared to be relevant to my research.

Additionally, many of the participants I was planning on recruiting were partaking at these conferences to gather information, possibly looking for big data analytics applications, and to network. It thus seemed to me that the topics that were being discussed there also had to have an influence on the participants in my research.

This had led to me including conferences as sites of research through relying on the concept of field-configuring events (FCEs) (Lampel and Meyer, 2008a), which has been an emerging method in the last decade, especially in the domain of Science and Technology Studies and organisational studies. Field-configuring events are conferences, events and exhibitions that target a certain group of specific fields, disciplines or organisations and play an important role in exchanging knowledge on specialised topics. These events often assemble leaders in the respective fields and have an influential role in the setting out and defining of the near future of the field (Lampel and Meyer, 2008a; Schüßler et al., 2015). FCEs are considered forms of organisations that emerge temporarily in a certain location and have their own social-material component (Garud, 2008). Researching FCEs can be especially fruitful regarding emerging trends, technologies, and organisations. Relying on a sociology of associations (Callon, 1999; Latour, 2005; Law, 2007), the FCE is a great site to follow and to use to reconstruct how associations are created that enable new technologies or organisations to emerge (Garud, 2008).

According to Lampel and Meyer (2008a, p. 1026), there are six characteristics that define conferences and events as FCEs:

- "*FCEs assemble in one location actors from diverse professional, organizational, and geographical backgrounds.*
- *FCEs' duration is limited, normally running from a few hours to a few days.*
- *FCEs provide unstructured opportunities for face-to-face social interaction.*
- *FCEs include ceremonial and dramaturgical activities.*
- *FCEs are occasions for information exchange and collective sense-making.*
- *FCEs generate social and reputational resources that can be deployed elsewhere and for other purposes*."

Examples of FCEs are conferences of the United Nations or specialised industry conferences, such as the former CEBIT Expo for the computer industry or the Geneva

Motor Show for the automotive industry (Hardy and Maguire, 2010; Lampel and Meyer, 2008a; Lange et al., 2014). In marketing, as well as in the data analytics and data science industries, several such events are organised throughout the year in Scotland and in the wider UK, and I had already attended some. As a result, the participant observation at FCEs was included as an additional source for data collection, contributing to a broad collection of different types of data, providing a wide variety of perspectives on the same issue and through this compensating for the necessary changes in the research design.

4.4. Data Sources

With these changes, a very robust triangulation of data collection methods and data collection sources was pursued. All these sources provided an in-depth view of the process of co-producing datafied knowledge about consumers and other market actors. Overall, the data corpus consists of 15 single key-informant interviewees. As mentioned, the recruitment of these participants was diverse and through many channels. Additionally, some of these interviews were recruited and conducted through the wider Big Data Surveillance (BDS) project, of which my research is part and to which it contributes. In that part of the BDS project, a similar kind of research was conducted which focused on the implementation of big data analytics in marketing, although with a different theoretical approach. The interviews conducted in that part of the project had an emphasis on the procedures behind implementing big data analytics in organisations and dealt with some questions relevant to the co-production of datafied knowledge, as well as the conceptualisation of consumers. The interviews did not cover every topic of relevance for my research but did provide a significant contribution to my data corpus.

All the participants in my research were in diverse positions and industries in which the conceptualisation of consumers significantly contributed to marketing practices and the sale of products and services of their organisations. Several participants were from consultancy firms, in which they were responsible for digital marketing projects or big data analytics projects with their clients. Thus, while these participants where operating on a B2B basis in their function as a consultant for other companies, their tasks consisted in advising these companies on big data analytics or marketing practices in relation to their consumers – and thus on B2C. As such, the content of the interviews was oriented to consumer and market research methods in a B2C environment. All key-informants,

their position or role within their organisation, the relating industry sector, as well as the duration of the interviews are listed anonymised in *Table 1*.

Interviewee code[3]	Position	Sector	Duration
Key1_entertainment	Digital Insights Manager	Entertainment / Media	45:30
Key2_publishing	Analytics Manager	Publishing / Journalism	54:36
Key3_travel	Head of Digital Experience and Digital Marketing	Travel / Tourism	35:04
Key4_insurance	Data Analytics Consultant	Insurance / Consultancy	1:08:01
Key5_consulting	Marketing and Customer Behavioural Change	Consultancy	35:26
Key6_consultant	Consultant in data analytics	Digital Marketing	55:11
Key7_consultant	Consultant in data analytics	Digital Marketing	27:31
Key8_consultant	Consultant in data analytics	Digital Marketing	50:23
Key9_consultant	Consultant in social media analytics	Data Science Consultancy	34:28
Key10_consultant	Consultant	Software Provider	51:14
Key11_consultant	Consultant	Data Science Consultancy	42:31
Key12_consultant	Consultant in data analytics	Consultancy	32:58
Key13_consultant	Consultant in data analytics	Consultancy	46:44
Key14_consultant	Consultant in data analytics	Consultancy	39:54

[3] This code will be used throughout the findings chapter when quoting individual interviewees. This allows for an easier attribution of the quote with the anonymised interviewee.

Key15_publishing	Head of Marketing technology	Publishing / Journalism	49:28

Table 1: Key-informant interviewees

Additionally, some contacts ultimately yielded smaller case-studies, which focused on the implementation and use of big data analytics for marketing in organisations. Also here, two case studies resulted from the interviews conducted through the wider BDS project. All the case studies were not conducted as in-depth case studies, as was initially set out in the research design. Instead of using multiple data collection methods, I relied on interviews and documents in these case studies. Nonetheless, through these changes, additional three smaller case studies were conducted, providing a great insight into the processes and practices of co-production and consumer conceptualisations.

The advantage of conducting small case studies in addition to the key-informant interviews, was that these case studies provided multiple perspectives on one phenomenon within the organisation. Each organisation and their implementation of big data analytics was treated as a single case. This meant that within each organisation, multiple participants in different positions and organisational departments were able to recount their view on one phenomenon within that case: how in their organisation, the implementation of big data analytics has influenced the co-production of consumer-oriented knowledge and what its effects were on the practice of conceptualising consumers. This has helped to partially compensate for the lack of in-depth insight into the case studies. In particular from an analytical point of view, the multiple small case studies have allowed for a cross-case analysis and thus a comparison of procedures and practices big data analytics for consumer and market research in different organisations.

A detailed list of these case studies can be found in *Table 2*. In some of these case studies, a wide range of industry documents were shared, which were also reviewed regarding their potential contribution to the data corpus. Although only few actually contained relevant material for my research, the ones that did were also used as data sources. A list of these documents can be found in Appendix 1.

Case	Interviewee Code	Position	Sector	Duration
F_beverage	F1_beverage	Marketing Manager	Beverage Industry	43:35

F_beverage	F2_beverage	Head of Digital	Beverage Industry	39:57
F_beverage	F3_beverage	Business Relationship Manager	Beverage Industry	44:24
I_travel	I1_travel	Data Scientist	Travel	49:09
I_travel	I2_travel	Director of Growth	Travel	39:12
I_travel	I3_travel	Head of Marketing technology	Travel	42:58
I_travel	I4_travel	Former COO	Travel	45:06
J_travel	J1_travel	Managing Director in R&D for marketing automation, data science and e-commerce operations	Travel / Loyalty Scheme	59:01
J_travel	J2_travel	Principle Data Scientist	Travel / Loyalty Scheme	51:15
J_travel	J3_travel	Data Scientist in marketing automation	Travel / Loyalty Scheme	40:19
J_travel	J4_travel	Managing Director in Product Design (J4)	Travel / Loyalty Scheme	50:01
J_travel	J4.1_travel	CTO in Product Design	Travel / Loyalty Scheme	50:01
M_entertainment	M1_entertainment	Digital Sales and Marketing Manager	Entertainment	25:53

M_entertainment	M2_entertainment	International Marketing Manager	Entertainment	43:01
M_entertainment	M3_entertainment	Ad Operations & Analytics Manager	Entertainment	31:16
M_entertainment	M4_entertainment	Business and partner development – mobile Apps	Entertainment	27:32
L_travel	L1_travel	Head of Analytics	Travel / Transportation	55:37
L_travel	L1.1_travel	Data Analyst	Travel / Transportation	55:37
L_travel	L1.2_travel	Data Analyst	Travel / Transportation	55:37
L_travel	L2_travel	Regional Commercial Manager	Travel / Transportation	45:39
L_travel	L3_travel	Data Analyst in Customer Experience Management	Travel / Transportation	44:17
L_travel	L4_travel	Business Analytics in Revenue Management	Travel / Transportation	35:02
L_travel	L4.1_travel	Manager in Revenue Management	Travel / Transportation	35:02
L_travel	L5_travel[4]	Head of Customer Experience Management	Travel / Transportation	Approx. 45 min.

Table 2: Case Studies

[4] Initially, this interview was planned to be conducted through an internet video call. However, due to unforeseen circumstances, this was changed at the last minute by the interviewee and the interview was conducted via telephone. As this was not planned, the settings of my telephone were set-up incorrectly and the recording of the interview was unsuccessful. As a result, a transcription was made based on memory minutes and notes.

The total interview sample size for my thesis consists of 39 interviews: 15 in the form of single key-informant interviews, and 24 interviews which are part of the five case studies. All the interviews lasted in average for 45 minutes, ranging between just under 30 minutes to a bit over one hour. The fluctuating duration of the interviews could be attributed to the situation that some of the interviewees were only able to allocate half an hour for the interviews, particularly as these interviews often were conducted during working hours. Nonetheless, despite some outliers, the duration of the interviews was sufficient to cover all the necessary parts of the interview schedules I had designed and gave enough room for the participants to talk and discuss about the implementation and use of big data analytics in their organisation and its contribution to co-producing consumer-oriented knowledge. Certainly, when put into relation with the sample size, the interviews provided a lot of information power for the analysis of my empirical data (Malterud et al., 2015).

Although there is no specific formula designed to specify a required sample size for a piece of qualitative research, even less prior to starting with the data collection (Sim et al., 2018), the selection of participants needs to be based on certain criteria. In terms of which participants to interview, multiple criteria were taken into consideration. From a practical point of view, the primary question concerned who was considered to have key or expert knowledge and was able and willing to provide information on the co-production of datafied knowledge for consumer conceptualisations. In particular, the last part is of importance as, with sensitive and commercial information at stake, not every participant would be willing to share information about their operations (Gläser and Laudel, 2009). The selection for contacting potential participants was largely made based on their position within their organisation. I specifically addressed data analysts or data scientists in middle management as they would most likely have the relevant knowledge about the practical operations of big data analytics, as well as about the decisions leading to these operations. The same choice was taken for the end-users of datafied knowledge – the marketers. In cases where other individuals from an organisation were involved, I aimed at building the contact to the relevant positions in their organisation.

In addition to the interviews, the empirical data was enriched with a total of seven industry conferences, of which five are counted as FCEs as they were attended with the purpose of collecting data, besides recruiting participants for my research. Although the

conferences had slightly different topics, in general, they all related partially to big data analytics and marketing. The Big Data Scotland 2017 conference specifically addressed big data analytics technologies, whereas the Digital Transformation Conference (2018 & 2019) focused on the strategies and processes of digital change and on the implementation of digital technologies. The Technology for Marketing 2018 Conference and Exhibition covered a range of digital technologies that were in use at the time in marketing, ranging from e-commerce technologies to big data analytics for consumer profiling, personalisation and tracking. DigitExpo 2019 was a large exhibition with some conference presentations on a broad range of digital technologies in general. Intelligent Automation 2019 addressed the topics of machine learning and artificial intelligence in different situations – not only specific to marketing. Data Fest 2019 was a hybrid of an academic and an industry conference and exhibition, focusing on the use of digital data across a broad range of fields, amongst them, marketing. It was also one of the few events which critically and academically discussed the problems that occur when operating in a datafied environment – be that in organisations or relating to an entire society. All these FCEs further contributed to the rich data corpus by supplying participant observation minutes, conference notes, memory minutes of discussions with attendees, extensive industry documents and pictures.

The choice of these conferences, which are listed in Table 3, also had practical reasons besides having topics of interest. As all except one took place in Edinburgh and were mostly free to attend for delegates, their attendance was resource-friendly – both timewise and moneywise. Only one conference was in London and so travel and accommodation had to be organised and paid for out of a budget which was included in my project. One conference – the Data Summit – had an entrance fee of £365, which was covered by my funding.

Field-Configuring Event	**Duration of Participant Observation**	**Topic**	**Location**
Big Data Scotland 2017*	1 Day	New trends in big data analytics	Edinburgh

Digital Transformation 2018*	1 Day	Implementation and development of digital technologies in organisations	Edinburgh
Technology for Marketing 2018	2 Days	New trends in (digital) technologies for marketing	London
DigitExpo 2018	1 Day	Exhibition and Conference on the newest trends in digital technologies	Edinburgh
Digital Transformation 2019	1 Day	Implementation and development of digital technologies in organisations	Edinburgh
Data Summit / Data Fest 2019	2 Days	Conference on the advantages and difficulties in a data-driven world	Edinburgh
Intelligent Automation 2019	1 Day	New trends in Artificial Intelligence and Machine Learning	Edinburgh

Table 3: Field-Configuring Events

4.5. The data collection procedure

My data collection consisted of a triangulation of different data sources and methods for collecting that data. The main data collection method involved semi-structured, qualitative interviews. Additionally, more ethnographic research methods in the form of participant observations and conversations were used for the FCEs. In the following sections, I will expand on these methods and explain my procedures and reasoning for collecting data to answer my research question.

* These events were visited prior to the change of including conferences as a source for data collection, and contributed only a little to the corpus of data that was used in the analysis.

4.5.1. Key-informant interviews

Most of my empirical data was the result of conducting key-informant interviews, either as individual interviews or as part of a case study. As mentioned above, the participants of the key-informant interviews are regarded as strategic informants who have an expert knowledge of their professional practices due to their position (Bogner et al., 2009; Bogner and Menz, 2009; Littig, 2009; Marshall, 1996). Conducting interviews with key-informants is not so different from other types of qualitative interviews but there are, nonetheless, a few considerations that need to be accounted for. The key-informant interview is a relatively structured form of qualitative interview in which the interview schedule serves as an important guideline. The schedule helps to ensure that all the questions of relevance are included – particularly during the initial sessions. However, as with the key-informant interviews, the schedule had to cope with deviations – be that in a change in the order of the questions, or an accommodation of other topics introduced by the interviewee, as well as ad-hoc questions. This is why the approach is considered a semi-structured interview (Gläser and Laudel, 2009).

Overall, my interview schedule – which can be found in Appendix 2 – was developed in such a way that it included the operationalisation of my research question (Brinkmann, 2018), provided a useful guide to conducting the interviews, while leaving room for interpretation. This allowed for other themes and questions to be added that may have emerged during the first round of interviews (Alvesson and Ashcraft, 2012). In my research, this included, for example, the topic of changes in skills through the implementation of big data analytics. This topic was initially not included in my interview schedule but was added later as participants were talking about the issues and some interesting aspects emerged. Besides the operationalisation of my research question, the questions for my interview schedule were informed through my research literature (King, 2004), particularly that of Critical Data Studies and Critical Marketing Studies, as well as the theory of STS and the co-production of knowledge.

Another important consideration when developing the interview schedule related to the artificial setting of the interview, which was significantly different from regular communications settings (Gläser and Laudel, 2009). Due to the artificiality, a lot of emphasis had to be put on the opening question(s), as this formed the basis of communication between the participants of the interview. For this, I first started with

explaining the research once again, as well as the general settings of the interview – such as the guarantee of anonymity, that the participants were free to leave the research at any time or provide a non-answer, and that the interview would be recorded to ease the load of transcription. This was followed by a very open introductory question(s) about the participants' roles in their organisations and the use of big data analytics in their organisations in general. The idea behind this opening was that the interviewees could start the interview by talking about something general and that the focus of the interview would be placed entirely on them. Depending on the role of the interviewee, the following is an example of these questions:

> "*So, to start, could you tell me about your role in your organisation? What is it that you do in your work, what current projects are you working on? How does big data analytics contribute to your daily work?"*

After the opening questions, different topics were addressed in the interview, which largely related to the research question and to the literature and theory. The first major topic revolved around the process of co-producing consumer-oriented knowledge through traditional market and consumer research methods, as well as the contribution of this knowledge to the conceptualisation of consumers. The second major topic addressed the implementation of big data analytics and was the core of the interview session (Alvesson and Ashcraft, 2012). The topics for these questions included the procedures of implementing big data analytics in the operations of co-producing knowledge – thus in producing as well as communicating and stabilising knowledge. In this topic, the questions also asked about the differences in operations before the implementation of big data analytics. This was followed by inquiries into whether any unforeseen difficulties in operations or uncertainties in knowledge had been experienced and, if so, how these were addressed. The concluding part of the interview schedule – and, ideally, of the interview – aimed at asking about future trends and plans for big data analytics in marketing, in their organisation and in general. The final question or statement always aimed at leaving some space for the interviewee to ask any questions which might have come up during the interview.

In terms of the practicalities of the interviews, there were several aspects that needed to be considered. The first concerned the setting of the interview. Face-to-face interviews

were the preferred method, as they promoted easier interaction with the interviewee and allowed for an observation of gestures and facial expressions, all of which are essential parts of the communication and can be subject to interpretation (Christmann, 2009). However, as almost all the interviewees and cases were not located in Scotland or in the UK, pragmatic reasons of time and money meant that only a small n number of face-to-face interviews could be held. As a result, only one of the single key-informant interviews could be conducted in person and, from my case studies, only the L_travel case study consisted mainly of in-person interviews, with six out of eight interviewees. None of the other interviews could be conducted face-to-face.

To be able to maintain visual contact with the interviewees during the distance interviews, internet video-messenger services were chosen as the second method, which proved to be a reliable alternative (Gläser and Laudel, 2012). Although it required relying on a technical functioning of the devices and the connection, the interviews were not much different than actual in-person interviews. Unfortunately, not every interviewee was willing to use or could rely on internet-based video-messengers services. In the case of L_travel, for example, due to company policy, the interviewees were not allowed to use these video-messenger services. This meant that the two remaining interviews from that case had to be conducted via telephone, as were two other interviews.

Telephone interviews are specific and require alternative preparation as the communication is only verbal. This involves, on the one hand, the questions having to be formulated very clearly and precisely, so as not to confuse the interviewee (Christmann, 2009). On the other hand, there is the difficulty of interpreting pauses in speech as it is not always clear whether the respondent is still thinking about an answer or has concluded her or his answer already. Besides depending on thorough preparation of the interview and the questions, I tried to overcome the latter issue by leaving enough time after asking the questions and after getting the initial answers. Even though this led, in some instances, to awkward pauses where the interviewee was waiting for the next question, in general, this technique contributed to the participants providing more elaborate answers. Overall, both the video interviews and the telephone interviews passed without any major problems. Some technical problems occurred from time to time, but these had no lasting impact on the quality of the interview, nor on the quality of the recording. That is, except for one telephone interview where the discussion could not be recorded. As this interview

was initially planned to be a video-interview but was changed at the last minute to be conducted via the telephone, the recording device was not properly set up, and no recording could be created.

In terms of conducting the interview, careful attention was paid to the general practices of qualitative interviews, such as being flexible, listening actively, and not "assuming that the answer to a question is so obvious that it need not be asked" (King, 2004, p. 18). However, the status of the interviewee as a key-informant or as an expert added another layer that needed to be considered in comparison with traditional qualitative interviews. This involved first that the topics were actually addressed to the specific expertise of the interviewee (Gläser and Laudel, 2012). This proved more difficult in some of my interviews as the profession of the interviewee was not always made absolutely clear, particularly when the participants were referred to me through other contacts. In one instance, I was informed that the participant was doing data analytics in marketing, while, in fact, she was partially responsible for B2B-marketing in her consultancy. Her main task of consultancy in big data analytics for behavioural change in consumers was still significant for my research, but it did create some confusion at the start of the interview. Additionally, the expert interview also requires that the interviewer has some expertise on the topics of discussion. The reason for this is so that the interviewee is not put in the role of being solely a basic information provider, but a situation is created in which wider discussion is possible (Gläser and Laudel, 2012). Therefore, it was important for me to get acquainted with the basics of big data analytics, as well as of digital marketing.

4.5.2. Critical reflection on the key-informant interviews

As the key-informant interviews are the main source of data for my research, it is vital to critically reflect on the data collection process. The first step concerns the assignation of who counts as a key-informant or an expert (Littig, 2009). Although the final decision was always mine as a researcher, the assignation of this role had already happened earlier. When recruiting potential participants at the industry conferences, their expertise had largely been ascribed by the conference organisers, in the case where they had been invited as panellists or presenters. Similar to the co-production of knowledge in organisations, this social and normative assignation is something to be wary of as it can contribute to how knowledge was co-produced for my research.

In the slightly different context of the interviews, aspects of knowledge co-production became especially noticeable. Due to the nature of the topic potentially addressing sensitive information such as commercial interests or surveillant practices, I always started the interviews by guaranteeing the participants full anonymity. This not only meant that their names would be anonymised but also all their affiliations, such as the name of their organisation, the specific sector they worked in, as well as the location they were working from. Furthermore, I made sure they were aware that they could, at any point, choose not to answer any questions and that my research was not aimed at accumulating any strategic or competitive organisational information. Some participants had no problems sharing a lot of information on their operations, even if it concerned potentially more problematic issues, such as the criteria they utilise to target consumers, or from where they collect the data. However, other participants were rather reluctant to talk about sensitive or critical topics on the record. For example, one of the single key-informant interviewees was very careful in answering my questions and deliberated a lot longer about her answers than others did. In some situations, she also specifically mentioned that she would rather not continue talking about these issues on the record.

Having these issues does not mean that the information provided by the participants was not useable or relevant. Their accounts provided an insight into their view of the operations of co-producing datafied knowledge in their organisations and the conceptualisations of consumers. However, this means that the findings of my research are not generalisable for a large population, although this does not limit the validity of my research (Maxwell, 2009). As the interviewees had contributed immensely and provided interesting insights into the operations at their organisations, with a lot of "information power" (Malterud et al., 2015), this left me with enough "rich data" for the analysis to establish valid findings (Maxwell, 2009, p. 244). This rich data is furthermore improved through the triangulation of research methods and data sources, particularly the participant observation at FCEs.

4.5.3. Participant observation

Besides the key-informant interviews for my cases and the single interviews, I also relied on participant observations as a data collection method at FCEs. While in some FCE studies, researchers have relied on an extensive array of official documents and notes, which are publicly available (Hardy and Maguire, 2010), this was not the case for the

industry conferences I visited. As hardly any documents were publicly available, the research of the FCEs required a physical presence. Since I was visiting the conferences anyway for recruiting participants, a participant observation was the most obvious choice. My role was that of a participant as an observer (Brannan and Oultram, 2012), which meant that, on the one hand, I fully participated in the event. This involved attending the conference talks, visiting the exhibition stalls, as well as networking and talking to the delegates of the conference. On the other hand, I also disclosed my role as a researcher and the purpose of my visit: to recruit participants for my case studies and interviews as well as carrying out a participant observation.

The opportunity of researching FCEs in person bears multiple advantages. First, the data has not been pre-selected and edited. When compared to the research of historical FCEs, which solely relies on documents and archival records, participant observations are the more flexible data collection method. Furthermore, as Lampel and Meyer (2008a) also note, the preparation for the empirical research of FCEs is also easier in comparison to the preparation of other sites of research. As the conferences and events are usually organised long in advanced, there is always some information material available, such as lists of participants, speakers, schedules, etc. Participation in the FCEs and observation of the activities taking place meant that I, as a researcher, could plan my visit accordingly and chose where to put my focus during the participant observation. Often, I could make an initial selection of participants I wished to talk to and could inform myself about their roles and their organisation.

Another advantage of being at the FCEs in person meant that I could react quickly to the environment and the subjects or objects of interest, as well as adapt my focus of research, if necessary. This mainly concerned the conference talks and workshops, which were part of the events. Often, these were happening simultaneously and a choice had to be made regarding which of the talks to attend. These decisions depended solely on the description of the talks that were provided beforehand. As a result, in some instances, these were either not as informative or the description and the content of the talk did not align. For data collection purposes, this meant that I could, if needed, switch rooms and attend other talks that were also relevant for my research.

Finally, as the FCEs is a social event which is strictly bound in time, location and participants (Lange et al., 2014), it eases the focus on what to observe and consider as important for inclusion in the research. As I mentioned, I could prepare in advance what to prioritise during my stay at the FCEs, and only sometimes had to adapt to changing circumstances. This made the participant observation a lot easier as the method can be relatively challenging. A significant difficulty of the participant observation is that there is potentially always an activity to observe and the risk of missing out on important events and activities always remains (Waddington, 2004). This risk is obviously a little bit lower at the FCEs. However, as FCEs may only last for a couple of hours, or two days maximum, there is the problem that one is limited to that specific instance of data collection, and there is no possibility of going back to the field of research after the event has finished. Overall, I still considered that the boundaries of the FCEs made the participant observation a little bit easier.

4.5.4. Critical reflection on the participant observations

Some difficulties arose during the participant observations, that should be considered here, which mainly concerned my note-taking. With participant observations, note-taking is the main method for collecting and recording data and, as such, the process is extensive and essential (Brannan and Oultram, 2012). My notes included descriptions of the participants, the events taking place, the actions, the settings, and the timing and sequence in which all this took place. Furthermore, I took notes of my feelings and emotions and made sure I was aware and recorded emerging hypotheses and interpretations of the situations, as is advised by Waddington (2004). During the talks, discussions and chats, I made sure to take notes, as these interactions contained a lot of interesting and relevant material. This meant that I was taking notes throughout the entire length of the FCEs, often for six to seven hours. Particularly in the second half of such a day, it was difficult to stay as focused as I had been at the beginning; this could also be observed in my note-taking. Events that occurred later in the day were not recorded in as much detail as the early day events, meaning that my notes were not as complete as they could have been. However, due to the large number of FCEs I visited, I was still able to collect a lot of data during my participant observations and the missing data was not impactful.

Another difficulty that occurred was my dual role in doing a participant observation and recruiting participants. This dual role meant that I often had to switch between “recording

device" and "recruiter". Certainly, after a few hours at the research site, it was increasingly difficult to stay focused on both roles and, at certain FCEs, this resulted in me being reluctant to continue trying to recruit participants later in the day. As I had visited two FCEs before starting my participant observation, I had a point of reference and knew that this loss of focus was not as prevalent when I was only recruiting. Nonetheless, this does not mean that my note-taking or my recruiting could be considered unsatisfactory. My empirical research has profited significantly from including FCEs as a source of data, and it certainly increased the depth of my findings.

4.6. Data analysis

During my data collection phase, I had already started to create an electronic repository for my data in order to manage my research data, as well as ensure that it was secure on the online server of the university. This included my field notes from the FCEs, my recordings of my interviews, and the full transcriptions I created afterwards. Also, some of the documents I had collected during my empirical research were included in my data repository.

For the data analysis, I opted for a thematic analysis, which serves as a method for "examining the perspectives of different research participants, highlighting similarities and differences and generating unanticipated insights" (Nowell et al., 2017, 2). In thematic analysis, a theme is seen as "a *pattern* of shared meaning, organised around a core concept" (Braun et al., 2019, p. 845). As thematic analyses are widely used in qualitative research, different approaches and concepts have developed over time, which often differ regarding the epistemology and ontology of the researcher (Maguire and Delahunt, 2017). My approach for thematic analysis is what Braun et al. (2019) have termed the reflexive approach. In a reflexive thematic analysis, it is important to consider the patterned meaning in relation to its context, similar to the way Jasanoff (2006b) considers the co-production of knowledge in its respective context. This means that is relevant to reflect on what role the participants are taking when they are talking about certain issues. In my interviews, this was particularly apparent when the interviewees switched from their roles as data analysts or marketers, into the role of consumers themselves.

From a practical point of view, a reflexive thematic analysis is always "the *output* of coding", which is "an organic and open iterative *process*" in which the codes "evolve throughout the coding" (Braun et al., 2019, p. 848). In this case, a code can be described as "a word or short phrase that symbolically assigns a summative, salient, essence-capturing, and/or evocative attribute for a portion of language-based or visual data" (Saldana, 2013, p. 3). Themes are not defined before the analysis is carried out, nor are the interview questions or interview themes regarded as the initial themes of the analysis (Maguire and Delahunt, 2017). It is also important to consider that the themes are not simply summaries of the data. Instead, as themes emerge from an iterative coding of the data, as a researcher I am constantly interpreting the data (Braun et al., 2019). Here, the data can be at the level of one or multiple sentences from the interviews, or only of parts of the sentence. Sometimes, the data can also be at the level of single words or sub-units of words or morphemes, depending on the depth and further methodological grounding of the research (Froschauer and Lueger, 2003). In my analysis, I mainly remained at the sentence level and did not venture too deeply into the interpretation and investigation of the meaning of single words. Before starting with the coding and the development of the themes, Braun et al. (2019) suggest that the analysis starts with familiarisation of the data. This can best be done by reading the transcripts once or twice while making notes of any first interpretations (Maguire and Delahunt, 2017).

Finally, the changes that occurred through the adaptation of my research design, moving from an in-depth, single, case study to multiple smaller cases studies and key-informant interviews, were also reflected in the analysis of my data. The small case studies were analysed individually. This means that each case started with the same codes that were deducted from the literature and the initial review of the data, before being coded and themed separately. Opting for this approach served to maintain the case-specific view and gain insight that might only relate to individual cases. Only after the individual cases were analysed separately, these were merged together and combined with the codings of the key-informant interviews, FCEs and documents, allowing to compare the findings, refine the codings and further develop the themes. While I have followed the cross-case analysis approach described by Stake (2005), I have not opted to proceed as schematically as he proposes, such as establishing matrixes, rating themes across cases, establishing individual case reports. As my cases were rather small, with some only comprising three

interviews, my analysis and comparing of the cases largely was done through note keeping in my research diary.

4.6.1. The coding process

The core analysis of a thematic analysis involves the coding of the data (Braun et al., 2019). For my coding, I took orientation from Saldana's (2013) coding manual, which provided excellent guidance. Before starting my coding, I made a list of questions I would ask while reading my data. These are questions concerning what the participants were trying to achieve with the actions they were describing, what the meanings behind these actions was and how the participants characterised these things (cf. Saldana, 2013, p. 21f.). Furthermore, I included questions and topics which I considered significant for my research question, such as how participants described how they conceptualised consumers.

I started my coding using the Computer Assisted Qualitative Data Analysis Software (CAQDAS) MAXQDA Plus 2018, which I later updated to MAXQDA Plus 2020 while still conducting interviews and visiting FCEs. My coding approach was both inductive and deductive (Braun et al., 2019), meaning that I started with three broad codes which had resulted from the familiarisation of the data, as well as from my literature review and theory. These deductive codes were the large categories, or parent codes, of "producing and stabilising knowledge", "conceptualising and segmenting consumers" and "difficulties and uncertainties".[5] During the coding process, a fourth and a fifth major category were added, inductively, in the form of "the role of data", as well as "practices".

To achieve the in-depth character of my research and the case studies, as well as to comply with the explorative approach, I adopted several coding strategies in a simultaneous coding method. My main coding strategy was to apply descriptive coding, which means that the code describes the key topic of the segment that is being coded (Saldana, 2013). To further analyse the meaning of my data, I coded the segments further into process codes and evaluation/versus codes. Process codes are specific codes of action and were chosen to denote the practices in my data. As the focus of my research relies a lot on the practices of co-producing datafied knowledge, process codes play an important role in unpacking these practices. This is what led to the "practices" category to be developed as

[5] My full coding system can be found in Appendix 4.

one of the five main categories. In this category, three parent codes of practice were chosen to describe the three different roles of my participants and their related practices: the practice of data analytics, the practice of marketing and the practice of insight management. Each of these parent codes has corresponding sub-codes (or children) (Saldana, 2013) to provide a deeper analysis of the wide range of different practices that are involved with these roles.

Evaluation and versus codes are considered by Saldana (2013) to be two different forms of coding; however, I used both together as a single coding strategy. Evaluation codes can be described as assessments about certain programmes, initiatives or changes. Versus coding is used to code a direct binary or dichotomy comparison that is used by participants. I have used both as one form of coding and applied this coding strategy to the research concerning the different methods of co-producing knowledge about consumers, in which there is a comparison between the traditional market and consumer research methods and big data analytics. Here, participants evaluated their efficiency and accuracy using both methods. It should be noted that, at times, my interview questions specifically asked the interviewees for such an evaluation.

The coding of my data was a lengthy process due to the amount of data that was being analysed, as well as the in-depth insight I was trying to obtain. I coded my data in two cycles. The first cycle was used to develop the codes by reading through my data and, subsequently, adding more and more (sub)codes to my categories and parent-codes, as well as attributing segments to my codes. After I had finished a first round of coding, I had a lot of codes and segments, while simultaneously having a lot of overlapping codes. The overlap of codes is something that cannot be avoided as segments can have multiple meanings. During the second round of coding, I focused specifically on these overlapping codes in order to merge similar codes and create a more robust and concise code system. In the end, my code system consisted of 94 codes and 3743 coded segments. This was without the “practice” category. Although this category provided interesting insights into the practices surrounding big data analytics for consumer-oriented knowledge, it contributed little to the development of the themes.

4.6.2. Developing themes

As mentioned, although coding is important for a thematic analysis, it is essential to note that the coding process is only one part of it. The next step entailed identifying the themes which had emerged from the codes (Nowell et al., 2017). Although I had already developed some initial categories for my coding, these could not automatically be considered to be my themes. Braun et al. (2019, p. 854) emphasise that, for the construction of the themes, it is important to refer back to the research question as "good themes are those that tell a coherent, insightful story about the data in relation to the research question." Referring to my research question, there are two major points or categories that where addressed in my empirical research and which were also covered by the thematic analysis. The first one being the conceptualisation of consumers and the second being the role of consumer and market research method used – or in other words, the co-production of consumer-oriented knowledge.

These two major categories also were reflected in my coding, where two of my five main codes were "consumer conceptualisation" and "producing and stabilizing knowledge." This led to my thematic analysis further proceeding from these two categories, taking into account the other categories I had established during my coding and looking for shared patterns of meaning within these codes and subcodes (Braun et al., 2019). This led to parent codes being split and included in different themes. For example, one parent code that had emerged was that of general "difficulties and uncertainties". Through the thematic analysis, it became clear that most of these difficulties and uncertainties, the participants were referring to, either related to the "process of co-producing knowledge", or to "uncertainties in knowledge" in general. In a similar way, the parent code of "the role of data" with many of its subcodes, had a shared meaning with how participants tried to make the case in favour for datafied knowledge, and thus significantly contributed to forming the themes of "datafied knowledge vs traditional knowledge about the market and consumer" and "communicating, translating, and stabilising knowledge".

As shown in Table 4, the two main categories of "co-producing knowledge" and "conceptualising consumers" thus each subsume 4 respectively 3 unique themes that emerged from my empirical data. The category of co-producing knowledge includes on the one hand a theme around "the process of co-producing knowledge" – through traditional consumer and market research means as well as big data analytics. On the other

hand, this category also includes themes that relate more to the stabilization of knowledge, through communication and translation, the comparing of different kinds of knowledge, and how uncertainties in knowledge are dealt with, but also not dealt with. Considering the theoretical underpinning of my research, these themes largely relate to the epistemological strand of the co-production of knowledge (Jasanoff, 2006b). The other category, "conceptualising consumers", includes three themes covering the different ontologies of the consumers that emerge as a result of different consumer and market research methods being used: the imagined, the data double, and the postulated.

Co-producing knowledge	**Conceptualising Consumers**
The process of co-production	The imagined consumer
Datafied knowledge vs traditional knowledge about the market and consumer	The data double
Uncertainties in knowledge	The postulated consumer
Communicating, translating, and stabilising knowledge	

Table 4: Thematic analysis of the empirical data

Finally, the writing-up of the findings serves as a "final test of how well the themes work, individually in relation to the dataset, and overall" (Braun et al., 2019, p. 857). Writing up my findings was an important step to further develop and refine the themes that had emerged from the data. For my thesis, this, for example, meant that one large category, which I had termed "the role of data" was initially considered as a theme in its own right, as it was a useful contribution to my findings. However, after writing a first draft and consulting with my supervisor, it became increasingly clear that this theme was better suited embedded in the other themes. This exemplifies why writing up is considered an essential part of the data analysis.

4.7. Ethical considerations

As my empirical research included the involvement of human participants, it was important to consider the ethical implications of my research before, during and after the research, and their participation in it. Even if the research being conducted appears non-threatening and, thus, of no ethical concern, participants can have issues with certain

aspects. The first step to mitigating this was to apply for ethical approval, which was granted by the University Teaching and Research Ethics Committee (UTREC) prior to my engagement with the empirical research (see Appendix 3 for the Ethics Approval Letter).

Another approach I took to minimise the risk of ethical concerns with my participants was the full disclosure of my research, both when recruiting them for my research and prior to conducting the interviews. I had prepared research information sheets which detailed the purpose of my research, the approach I was taking and how the information they would provide was to be used – in my thesis as well as in future publications. While conducting my interviews, I always reserved the first minutes before starting the recording to explain my research and ask for consent to record and transcribe the interview. In this, I took care to explain that the transcriptions would be entirely anonymised so that anything they said could not be traced back to them or their organisation. While this was acceptable to all my participants, and all agreed to be recorded, some preferred not to talk about certain aspects of their operations on tape. This is why I reserved a few minutes after the official part of the interview, after I had switched the recording device off, to allow participants to talk about things they had not wanted to mention earlier and which would not be included in the data collection. Moreover, all the participants signed the university's consent form, which detailed the ethical considerations and provided an additional form of information.

During my participant observation at the FCEs, I also tried to fully disclose my role as a researcher – before attending the event as well as at the event itself. Here, a difficulty I encountered was that, although I tried to be open about my role as a researcher, I was not able to do so at every instance or with every participant I interacted with. While not trying to be a covert observer at the FCEs, sometimes the situation did not allow me to be open about the research. This was the case when walking through the exhibition halls at the Technology for Marketing conference, which included over 250 exhibitors and well over 10,000 participants. As Brannan and Oultram (2012) have noted, while being open about the intent of the research is necessary and ethical, this might not always be practicable, especially at large events, as was the case here. In such situations, I tried to make sure I was not interfering with the participants' privacy. For example, in some instances, I

overheard interesting conversations but the situation did not allow me to disclose my role as a researcher, which meant that I did not take note of these conversations.

Finally, another step I took to ensure the ethical handling of my research was that I aimed to be as transparent as possible about my procedures. This entailed uploading and publishing my data in the online data repository – Pure of the University of St Andrews – after the completion of my thesis. This layer of transparency presents, on the one hand, the opportunity to review the findings with the corresponding data. On the other hand, It also permits other researchers to further use that data for future research and the potential transfer of findings. Besides ensuring open and ethical conduct, this approach also has the benefit of increasing the validity of my research (Maxwell, 2009).

To conclude my methodology chapter, the empirical research of the co-production of knowledge for the conceptualisation of consumers has produced a range of interesting findings through the triangulation of data collection methods and data sources, as well as a reflexive thematic analysis. In the next two chapters, these findings will be presented and discussed in order to provide an answer to my research question.

5. Pursuing the data double – Conceptualising consumers in data-driven organisations

5.1. Introduction

The conceptualisation of consumers is a common and long-standing process in organisations. A wide range of studies have been conducted looking at the process and the politics behind these practices. These studies, however, have focused more on the traditional, classic consumer research practices which rely on classical data analytics, surveys, focus group research, expert panels, experiments and the like, in which the consumer conceptualisations tend to be established through the segmentation of consumers into different groups (Dubuisson-Quellier, 2010; Grandclément and Gaglio, 2011; Sunderland and Denny, 2011). The research here extends this by further looking at how consumers are conceptualised through the practices of big data analytics, examining how these conceptualisations are created, might emerge differently and, to some extent, lead to different outcomes based on these conceptualisations. In this chapter, I will focus on how consumers are conceptualised in the participating organisations and how these conceptualisations tended to be different, depending on what research methods were used. This largely addresses the first part of the research question. As I will show here, there is not only a – somewhat expected – difference in how the participants conceptualised consumers comparing the results emerging from traditional types of market and consumer research methods and big data analytics. There is also a disjunction between the expectations the participants had regarding datafied consumer conceptualisations and their practical implementation.

Upon explaining the practice of conceptualising consumers through big data analytics, the participants compared the practice with how it was done before the implementation of big data analytics. Conceptualisations based on traditional market and consumer research methods – quantitative as well as qualitative – were often discussed by the participants with an air of discontent. The main criticism was that such conceptualisations are imprecise, vague, too broad and not actionable. These traditional conceptualisations, which I would call "imagined consumers" based on the descriptions provided by the participants, are largely perceived as outdated, and the development of big data analytics has been welcomed as an improvement.

With the increased availability of digital data and the possibility of processing this data, more advanced means of analysing data and creating better, or more accurate, consumer conceptualisations seem to have emerged. These consumer conceptualisations can be described, by relying on Haggerty and Ericson's (2000, p. 605) concept of the data double, as the process of "abstracting human bodies from their territorial settings, and separating them into a series of discrete flows [that] are then reassembled in different locations as discrete and virtual 'data doubles'". This description of the data double translates well into how the participants described their consumer conceptualisation based on big data analytics. In particular, the possibility of abstracting details and information of individuals and reassembling them into virtual 'data doubles' was portrayed by the participants as being a distinct improvement on the classic conceptualisation of consumers.

However, looking more specifically at the procedures of big data analytics in practice, there is a disconnect between the expectation of the data double and its implementation. Based on how participants described the processes behind conceptualising datafied consumers, while all of them were trying to build data doubles of their consumers, it appears as if none of them were actually in the state of doing so. There are multiple difficulties that emerge, and participants in organisations often go to considerable lengths to make big data analytics work. Inaccuracies in datafied knowledge are persistent and, because of this, a disconnect between the actual consumer and the data double consumer conceptualisation emerges. Still, participants often act on these consumer conceptualisations as if they were an exact one-to-one copy of the real consumer – a data double. This disconnect is what I would call "postulated consumers", conceptualisations which are expected to be representations of consumers as an exact one-to-one copy of their real counterparts without clear evidence that they are.

In the following sections, I will discuss how these three different consumer conceptualisations – the imagined, the data double and the postulation – were described by the participants and, importantly, how these conceptions are influenced by traditional and datafied knowledge. The first conceptualisation is based on traditional forms of knowledge and can be used to describe the 'old' way of segmenting consumers. The second is the data double, which relates to the use of datafied knowledge for the creation of these conceptualisations, but which is only based on expectations and visions of the

participants and not found in practice. And the third is that of the postulated consumer. This is the data double in practice, described by participants as the exact copy of the consumer, while just being an approximation, with uncertainties and errors included, and upon which participants still tend to act as if it is the data double. Finally, I will conclude this chapter by discussing how the use of the postulated consumer conceptualisation has the potential of affecting the real consumer.

5.2. Conceptualising consumers – the imagined consumer

The imagined consumer relates largely to the classical ways of conceptualising consumers, based on the basic consumer research methods. These can be created quantitatively through classical data analytics and data sourced using basic loyalty programmes and surveys, but also qualitatively through data collection methods such as focus groups or experiments. Here, the participants talked mostly about the more typical consumer segmentation, in which individual consumers are grouped in segments based on certain commonalities in demographics and behaviour. In a way, the participants described how conceptualising consumers is a process in which they are trying to imagine what their different consumer segments or groups approximately look like and how they potentially behave – hence, why I have called the conceptualisation, *imagined consumers*. Imagined consumer conceptualisations are established through a mix of demographic and behavioural information. These two kinds of information are generally described as equally important in establishing segmentations and each provides a unique piece of information for the imagined conceptualisation.

Consumer demographic information represents the basis of consumer segmentation in the form of age, gender, income, location, race, education, mobility, etc. Demographic consumer information is traditionally described as being collected through surveys or is included in different databases within an organisation, such as being part of a loyalty scheme. When discussing existing customers of their organisations, the participants mentioned the large availability of demographic information which has enabled the creation of segmentations up until now. Demographic information was portrayed as essential by the participants as a large majority of the marketing messages and strategies they used were specifically tailored for specific groups or individuals. This can, for example, relate to a product being more specifically designed for elderly people, or

because it is assumed that different groups respond to messages in different ways. As an example from the entertainment industry shows, demographic information may constitute a key piece of information for establishing segmentation and addressing the segments accordingly:

> "*... so, we have a [company]-internal, let's call it market research bible [...] where our team specifically defined target audiences. Thus, based on age and intensity of music consummation, where we classify people demographically. Thus, also gender, income, [...] with different demographic criteria they are classified. And these groups then also have their own names, so we actually address them in these terms, when we are talking for example, [...] young families, they have a specific name, [...] or elderly [...]*". [M2_entertainment, Pos. 24 – translated from German]

Consumer behaviour, on the other hand, targets, more specifically, the way consumers act in relation to certain products and services, and this information has often been uncovered through focus group research and experiments. For example, as the same interviewee from the entertainment industry explained, they would – and still do, partially – ask individual consumers in focus groups about their daily routines, how they would consume music, what type of music, etc., after which they would have them listen to different music tracks and rate them according to their taste. This information serves an important role in marketers better knowing the different consumer segments of their organisation, the different consumer behaviours, and their relationships with and perceptions of the products.

However, in many instances when talking about these traditional consumer segmentations, participants would address the approximate nature of these segmentations. They were perceived as being imprecise, rather vague and not directly applicable to the individual consumers. Although these consumer segmentations were based on actual knowledge about the consumers, resulting from consumer research, some participants referred to the knowledge as simply being "assumptions", which were being used in the construction of imagined consumer conceptualisations. As the following quote from two interviewees in the travel industry shows, the use of such broad assumptions was perceived as not being specific enough. Instead, the interviewees rather preferred to have

knowledge about consumers available that was specifically applicable to the individual passengers:

> L4: *"Yes. Yeah. A lot of assumptions are made whereas, we need all of this to be able to make something that's a little bit more clear, for everyone."*
>
> I: "*So basically currently the assumptions are not based too much on... or are not quite accurate or?*"
>
> L4.1: *"I think they're just quite wide assumptions so it's a limited range of people, or just buying type, or for as it's, we used people who purchased within a certain time to departure to differentiate masses of things. So, they're quite broad brush rather than very specific to that single passenger."* [L4 & L4.1_travel, Pos. 142-147]

The participants often mentioned the difficulty of working with these imagined consumer conceptualisations and, in particular, the use of broad and imprecise consumer segmentations was perceived as being anachronistic. In some of the organisations – for example in the travel or the entertainment industry – participants described how operations were increasingly moving into the digital sphere, with the practice in marketing shifting more towards individualised consumer targeting. Thus, in many instances, being able to track consumers on their journeys, harvest the information they leave throughout their online presence and, most importantly, individually profile and target them is seen as a necessity. It should be noted that profiling and targeting are not seen as being feasible through the classical means of having unspecific consumer segmentations or imagined consumer conceptualisations.

Even in the case where advanced information already has contributed to the conceptualisation of imagined consumers – through the implementation of classic data analytics and basic digital market research – the participants still regarded the conceptualisations as inefficient. While imagined consumer conceptualisations might provide an indication of the direction of the marketing strategy, or of some of the nuances marketers might apply, they are rarely seen as giving the necessary information about the targeted consumers. The imagined consumer is seen as an outdated method of acting on consumers, with the traditional consumer insight being considered inaccurate and the

resulting conceptualisations not being actionable, consequently limiting the possibilities of marketing.

The following quote of a digital insights manager in the entertainment industry provides a good summary of the perceived problem with the imagined consumer conceptualisation. The expectations of what marketing should be doing seem to have changed. In her view, marketing now requires the ability to apply advertising and promotion strategies on an individual level instead of on large segments. This means that the existing practices and procedures are, by and large, outdated and in need of change:

> "*I think the outputs have become different, though. The expectation that a segmentation that is non-identifiable, and by that I mean I can't apply it as an individual customer level. I can't just… I can't create a like-for-like profile assuming that every man in his 50s, who owns a car and earns more than 30 grand is gonna be the same.*" [Key1_entertainment, Pos. 31]

Simply relying on "assumptions" and "approximations" regarding consumers are not sufficient anymore. The concept of segmentation in the form of imagined consumer conceptualisations is still in use in certain organisations that were part of the research, such as in the case of M_entertainment. Nonetheless, as we can see through the accounts of the participants and the presentations and discussions at the FCEs, the emergence of big data analytics and datafied knowledge about consumers creates the expectation of having the ability to individually conceptualise consumers.

5.3. Conceptualising consumers – the data double

With the changing environments in which certain participating organisations are operating, including the increased shift of their business and their consumers into the digital sphere, the expectations of marketing operations seem to have changed. Certainly in some of the case studies, such as in M_entertainment and L_travel, their organisations had only recently, as in the last two to three years, started to transition to becoming a so-called data-driven organisation, which essentially means that decisions in many areas, including marketing, are made based on previously analysed data. One of the main expectations regarding these changes in marketing is the emergence of an ideal-type of consumer conceptualisation, which can best be described as the data double, in the form of a "*uniquely identified consumer*", as a data scientist in the travel industry described it

[J1_travel, Pos. 31]. Regarding this, participants often mentioned the necessity of profiling, personalising, and targeting individual consumers on a granular level, so that, ideally, their actions and behaviour can be better anticipated. With the imagined consumer conceptualisation being seen as inflexible, inaccurate and limiting, the data double is perceived as a more practical and actionable conceptualisation for use in marketing, it being a type of consumer conceptualisation that is expected to be obtained through big data analytics.

Similarly, as with the imagined consumer, datafied consumer information is still divided between demographic and behavioural consumer information. The participants described how the availability of digital data had improved the knowledge regarding both types of information. As some participants explained, the possibility of relying on large data firms, such as Experian, to gain more (demographic) information about their consumers permits them to "*enrich* (…) *spotty information that you've got* [Key15_publishing, Pos. 26]. Still, the participants largely put their focus on the possibility of uncovering more detailed knowledge on consumer behaviour when talking about big data analytics, as the following participant, an analytics manager, elaborated. Although she also considers that demographics contribute to the building of individualised consumer profiles, the advantage is seen in the availability of more detailed insight into the behaviour of consumers:

> *"Now, we do have a trials kind of model. And in that instance, there is more focus on profile because we do know that people with a kind of certain demographics, but also certain behaviours as well, are more likely to convert and so. They would monitor that sort of thing and certain demographics are indicators of kind, of quality or field or something. But I'm not aware of anything around kind of demographics recently that has been like a big change in perception or strategic shift. It will be more around behaviour, I think."* [Key2_publishing, Pos. 39]

Generally, the participants considered the more advanced – datafied – consumer-oriented knowledge to include consumer behaviour, their different interests, their interactions with the organisation, and with other relevant social or group networks. For example, the emergence of big data analytics has been described as enabling the possibility of recreating entire consumer journeys. This can include all customer interaction from the

moment the customer starts to interact in one way or another with an organisation, through the consumption process, or any other kind of interaction between the customer and their organisation. More 'sophisticated' consumer-oriented knowledge is not only limited to assessing consumer behaviour but can also include wider consumer interests, perceptions, where consumers are located in the digital sphere, how they can be reached with advertising, how they can communicate more effectively, at what instance of the product release different types of consumers are more responsive, etc. Furthermore, the participants noted the increase in information on the general market or the environment their organisation was operating in, which also has been enabled through the emergence of digital information.

All in all, many participants mentioned having access to "*dozens or maybe hundreds of bits of information*" [Key4_insurance, Pos. 80], and now being able to provide a "*holistic view of the consumer,* […] *a holistic view of the market performance,* […] *a holistic, competitive view*" [F1_beverage, Pos. 129]. Furthermore, big data analytics not only provides the possibility of better understanding the existing customers of an organisation, but also the ability for organisations to analyse what they do not yet understand. As the following account from one of the interviewees shows, this can be about their existing customers, but can also relate to non-customers or potential customers of their organisation:

> "*…we model against you know our current subscribers or interest groups and you know apply big data sources to them so that we can find... So, that we can actually find other people out in the ecosystem of the worldwide internet and to also bring in*." [Key15_publishing, Pos. 8]

It can be seen that the participants were aware how big data analytics was able to improve knowledge about their and potential future consumers in their organisations, allowing them to measure consumers in a way they had not been able to before. Besides the quantity of information that big data analytics had made available to them, there was another improvement that the participants considered to be a crucial outcome of big data analytics. Instead of relying on declared behaviour by the consumers themselves – as is often reported through the use of traditional market and consumer research methods – they could now rely on consumer information that had been collected automatically,

sometimes without the consumers' knowledge, and aggregated together. This is often considered to be closer to reality than the declared behaviour:

> "*So, through mining data sets, we can actually get a real handle on actionable insights, like what they actually do, rather than what they say they do.*" [F2_beverage, Pos. 42]

The goal of the data double is to create personalised consumer conceptualisations, one-to-one profiles that accurately represent the individual consumer. The participants considered that only through the availability of datafied knowledge, could one-to-one profiles that personally identify consumers have become possible. The individuality of these profiles – in comparison to the broader imagined consumer conceptualisations – was seen as enabling the participating organisations to establish data doubles, as the following quote from a digital insights manager in the entertainment industry shows:

> "*What it does do is, because the nature of big data is digital, it enhances the, the amount of data that we have about individuals, and that is individually identifiable. Which means that, on a one-to-one level, your experience is better. Yeah, because we're getting away from this idea of grouping people based on, a spiral* [unclear] *set of like characteristics which won't apply to probably just under the majority in that group. So, at an individual level, it has vastly improved the customer experience because big data, as I said, by its nature, is digitally identifiable down to a one-to-one level.*" [Key 1_entertainment, Pos. 59]

Besides the personalisation of individualised consumers in the form of a data double, the participants referred to the potential of anticipating consumer behaviour through the datafied knowledge they possessed about them. Big data analytics has refined ways of predicting how consumers might behave and how they might react to specific marketing measures. Due to the vague nature of the imagined consumer, the precise anticipation of consumer behaviour had not been perceived as feasible before. The targeting of consumers and their behaviour has not only become more refined through big data measures but has been projected into the future, where big data analysis allows the calculation and anticipation of consumer behaviour, which organisations can act on, as the following quote from a participant in the insurance industry shows:

> *"So, what would happen is, you would say, for example, give me an algorithm, give me a mathematical algorithm, that models THIS particular behaviour. And one example that we were interested in doing for the insurer, for (company), was to look at things like: How can we predict whether this person will renew in 12 months' time? […] So, with millions of records what we did is, we pulled together the ability to, to sort of submit to this modelling tool, you know dozens of variables and learn from the data and say, what do these people you know what does it look like when someone, when someone does this. So, in a typical case, you might say: well I've got all this historical data some of these people cancelled and some of them didn't. What can I learn about the people that did and that didn't? What differentiates them? Now I've got a model for it. [I: I see what you mean, yeah yeah]. Now, now I get a new customer. I'll run them through the model, and it'll tell me how likely they are to renew."* [Key4_insurance, Pos. 75]

In the end, personalised and targetable consumer profiles of individuals – data doubles – were perceived by the participants as an ideal tool for their organisations, not only to provide them with a highly personalised picture of their existing as well as potential consumers, but also to enable the possibility of addressing the consumer on a personal level. The desired effect of the personalised consumer conceptualisations in the form of the data double is that the individual consumer has the impression of being personally addressed and personally favoured by the organisation. Generally, as many participants mentioned, the personalisation of consumer conceptualisations goes hand-in-hand with the targeting of individualised consumers. The participants, as well as others in their organisations, expected that through big data analytics and digital marketing, the targeting of consumers would become more feasible. In this, participants stated that the goal was to specifically address individual consumers with personalised messages and offers, based on detailed and datafied consumer-oriented knowledge.

As such, many participants mentioned how the personalisation would be a benefit for both their organisations and for the individual consumer. Their organisation would be able to sell their products and services more easily, for a higher price due to being able to better assess what prices individuals would pay, and how to retain their consumers more easily. In this process, the individual consumer was perceived by the participants as benefiting from the personalisation thanks to receiving individualised information, which

allowed the organisations to send out advertising that was relevant to the individual instead of them receiving general advertisements and offers. Further benefits for the consumers were identified by the participants as customers no longer having to make certain consumption choices since their organisation was able to limit the choices to what was relevant, having the necessary information to do this.

The personalisation was further perceived as a way of building a kind of personal relationship with consumers. Some of the participants specifically addressed their focus to customer experience and customer engagement in their organisations. Enhancing customer experience was often not only identified as a direct outcome of the personalisation of consumer conceptualisations but also as a good way of binding the individual consumer to the organisation. Establishing consumer loyalty by trying to create a personal relationship with the individual consumers was perceived as being increasingly possible by having access to the individualised data double conceptualisation and through personalised targeting.

While the benefit for the individual consumers was mentioned from time to time, in the end, the participants saw the main benefit of having data double consumer conceptualisations for their organisations, primarily in an economic sense. In general, the goal was to target high-value consumers in the hope that this would lead to a conversion, thus requiring less effort to achieve a sale. The personalisation of consumer conceptualisations aims at producing offers that are more likely to raise the interest of the targeted individual consumers, and thus nudges them into buying the products or services in question. At times, this can be on a relatively specific and personalised one-to-one level, as the following quote by a digital marketing and behavioural change consultant, referring to one of her potential clients, shows:

> *"So we can we can, if we wanted to, we could see that John from (company) went on our website at 9:30 a.m. Yeah, we could pick up the phone to him and have a conversation with him and he'd obviously be in the hot state of being engaging with us as a company and it means that the types of leads we go after, it's quite a long sales cycle. So, it's very much, quality of leads over quantity, and that means that we have to try and nurture them quite closely, rather than it just being a numbers game. So, it gives us this really joined-up view of our leads and we can*

see when who's more interested and engaged and then use that to know when to approach them." [Key5_consulting, Pos. 32]

5.3.1. The validity of the data double

In order for the data double to establish individualised profiles that aim at targeting consumers and anticipating their behaviour, marketers need to be certain about its validity, as well as about their marketing measures working. Regularly, respondents mentioned how "*data will help us to plan and prove the effectiveness of our marketing endeavours and our marketing investment*" [F2_beverage, Pos. 38]. Datafied knowledge is perceived as delivering the necessary insight and helping in assessing marketing measures. Importantly, however, datafied knowledge is seen as helping to provide individuals in their organisations with the certainty that their consumer conceptualisations, their data doubles, are, indeed, exact representations of their real counterparts. As the following quote shows, it is not acceptable anymore to approximate consumers for marketing measures. Instead, it is as important to know who the consumers are with certainty and precision before marketing measures can be taken:

> "*But you know so, but I know that and so I've actually got to say: okay, well, I can't guarantee you they're all blokes, so I can't do something different. So, whereas NOW, that's not okay. We would only send something that was designed for somebody in the latter years of their life to somebody we KNOW is in the last years of their life. Not somebody who is, possibly, characterized in a group that might be old, you know.*" [Key1_entertainment, Pos. 52 – highlights are stresses expressed by the interviewee]

The data double as an ideal type of consumer conceptualisation in marketing is perceived as becoming possible through the use of big data analytics. Datafied knowledge, with its attributed wealth of detail and accuracy, provides the necessary consumer insight for the data double to work as a one-to-one representation of the individual consumer. However, looking at the actual implementation of big data analytics for market and consumer research, it becomes obvious that the data double is not as precise as the participants expected it to be and it would appear to be a slippery concept. For instance, in practice, there appear to be substantial differences regarding the actual possibilities and capabilities of big data analytics within the organisations that were part of the research. Only some of

the organisations had big data analytics capabilities in place that were technologically advanced, while others were still rather limited concerning their options. This difference could, for example, be observed in the types of data analytics, and to what degree of sophistication, the participating organisations were able to perform, which affected the resulting knowledge about consumers. Some already had big data technologies in place that allowed them to do automated analysis on individual consumers. A participant from an organisation in the insurance industry explained, for example, how they could calculate propensity scores, such as the probability of consumers converting to a specific product or offer. These scores were based on the consumer's demographics and behavioural profiles and these allowed algorithmic models to adapt and improve constantly, further increasing the accuracy of the individual's propensity scores.

Other organisations that took part in the research seemed not to have such technological capabilities in place. Some participants reported about recent – i.e. within the last three years – implementation of big data technologies in their organisations, and most of them mentioned the extensive plans their organisations had regarding the further implementation and development of big data analytics. As such, some of the participants and their organisations were not yet in the position of actually performing the detailed and high-level analysis and personalisation they were interested in – and, by extension, did not have the ability to establish a data double consumer conceptualisation. Nonetheless, a commonality between most of the participants was a tendency towards a common end-goal – the vision and expectation of creating exact data doubles of consumers – as the following quotes from a business analyst and a revenue manager in the travel industry show:

> L4: "*So that's the benefit from L1's team becoming bigger. It's like, well we have stored all this data but now it's like what do we do with it. We need to use this. […] Hopefully, the customer experience side of it as well, where sort of the digital side will come in as well and feed into each other, where we can get people buying patterns and all that kind of stuff, how it feeds together. And as soon as you got all that data integrated together it makes... Well, hopefully, it'll allow us to target people more smartly, I guess. Giving them exactly what they want, or what we think they want, or what we want them to have.*" *[…]*

> L4.1: "*...very specific to that single passenger. I know you booked three weeks before travel and you're always going to (region). So that's when I need to give you the optimal price. Which I think is the vision. […] I think it's basically being able to find this granular... I think for the revenue growth something that gives us this so granular but then also can throw the bulk on to them. So, you can offer them food and drink discounts, if we know what you like. Or you can arrange taxis for them. All the added-on stuff, that you know that individual people use. As well as their (transportation) ticket. And when they need to, and when you can upgrade people. I think it's the personalized approach (unclear). Finding out exactly what one person wants and making them think they're getting this... So, then the offer is just for them. Where every single person is getting a tailored offer.*" [L4 & L4.1_travel, Pos. 139-157]

In the way the participants described their capabilities of producing and using data double consumer conceptualisations, it became clear that not every participating organisation was producing individualised consumer profiles. Instead, the data double consumer conceptualisations ranged from specifically individual one-to-one profiles to broader segmentations that contained individual personalisations. In some cases, these differences stemmed from the level of progress of the implementation of big data analytics in the organisation, as discussed above. However, in other cases, the participants described how they willingly adhered to overarching segmentations. Instead of building individual profiles, big data analytics and datafied consumer-oriented knowledge was used to create tighter, individualised segmentations – sometimes termed "*super segmentations*" in the industry literature [Music:ally, Sandbox Issue 225, p. 2] – within these overarching segments. This deliberate choice can depend on respective marketing strategies. In one example from the beverage industry, the marketing strategies tended to be designed according to these larger segmentations, each marketing strategy getting personalised advertising based on the tighter and individualised preferences within each larger segment:

> "*So, we kind of helped build a, let's call it, a tighter definition within that DMP* [Data Management Platform] *database, which then helps both if we're doing that from a global level, but then also we provided guidelines for our markets.* […] *And then depending on what your campaign or activity is, you need to be looking*

at the [segments] interests of things like film, gaming, you know, sport, whatever it is. And then that allows us to kind of build little profiles or segments within the database to then actually go and actually actively target with our media money." [F1_beverage, Pos. 55-56]

Despite the differences between the degree of the segmentations that were created based on datafied knowledge, the participants, nonetheless, perceived all of them as data double consumer conceptualisations. The use of datafied knowledge, with its increased level of detail regarding consumer insight, contributed to this perception. Datafied consumer segmentations are – despite their broadness – often still perceived as more actionable, more precise and, most importantly, actually representative of the real consumers. Since the imagined consumer conceptualisation is largely considered to be inaccurate and too broad, big data analytics, with its more precise and granular knowledge, was embraced by the participants and changed how they saw and built consumer conceptualisations. The consumer conceptualisation as the data double was perceived as being on a very individual, personalised level, with the prospect of having every single consumer in a computer system and the ability to target them on an individual basis.

The particularity of the data double in practice is that this concept is often considered to be a factual representation of the individual. However, with the way consumer conceptualisations are built, the emphasis is mainly only on a selected number of consumer traits and information points deemed relevant to the organisation and related to the marketed product or service – as we will see in more detail in the second chapter of the findings. The selection of these specific traits and pieces of information is mostly based on their economic potential and value as the aim of the organisations in question is to increase their sales or long-term economic value. As a participant of the travel industry put it, ultimately "*what we are trying to do is figure out a demand curve of every consumer that comes through our site. Basically, do a profit maximisation on it. So, it's an economics operations research kind of thing, that's how I look at it*" [J3_travel, Pos. 8]. The flipside of this focus on relevant economic traits is that the real consumer will only be represented partially since other traits and points of information are left out.

Finally, as I will also elaborate in more detail in the following chapter, the way datafied knowledge is being created also shows that the information used to build the data double

consumer conceptualisation is rarely as exact and objective as the participants expected it to be. Not only is the process of doing the analysis complicated, influenced by many different aspects such as the different teams in the organisations, the technology itself or the underlying norms, but big data analytics also has the potential to produce erroneous results, involves varying degrees of accuracy or it can be biased. All of these things should be taken into account when creating data doubles but they are not always considered. Consequently, the data double risks becoming a slippery concept, one that is mainly based on the expectation that it is an exact copy of the real consumer – in a sense, an ideal type consumer conceptualisation. Although the ideal type might still be more detailed, finely grained, and even more accurate than the imagined consumer, the data double, in practice, will only ever approximate the real consumer.

5.4. Conceptualising consumers – the postulated consumer

The ideal type of consumer conceptualisations, as with the data double consumers, were, in practice, used by the participants as real representations of the consumers. The participants either assumed that the consumer was represented in a factual and accurate way or, at least, perceived the consumer in the way they wanted the consumer to become. In other words, the ideal type consumer conceptualisation of the data double became a self-evident postulation of the consumer. The conceptualisations are expected to be a representation of each consumer, an exact, one-to-one copy of their real counterpart, even though there is no clear evidence that she or he actually is. Instead of being a data double, the postulated consumer conceptualisation risks being nothing more than an assemblage of data in which the actual consumer is only partially represented through the datafied knowledge (c.f. Zwick and Denegri Knott 2009).

The emergence of postulated consumer conceptualisations was exemplified on several occasions by the participants, in particular when their organisations aimed to try to foresee and anticipate consumer behaviour in their marketing strategies. In these processes, the conceptualisations were perceived as factual copies of the real consumers. In some situations, as the following quote from the head of digital marketing in an organisation in the travel industry illustrates, this is all still very speculative, and rather represents her expectations and visions of big data analytics. The interviewee specifically noted how, ideally, the necessary consumer insight was provided by the system, i.e. the data analytics

technology. Big data analytics is seen as trustworthy in delivering an accurate and relevant representation of the consumers and enabling personalised targeting with offers and the anticipation of consumer behaviour:

> "*We are showing personalised… I mean this is still a couple years down the line, but* ***ideally*** *(!) if an exporter from [country] is on the site, looking at suppliers, some content from our brand will come on and say: here are some of our supplier tours, or you know, some of our premium tours that you can book on. So, coming to us to check it out before you export the product. That's ideal, that's the end-goal. So, to have all the systems talking to each other and really serving personalised content, that's relevant to people*." [Key3_travel, Pos. 43 – emphasis added to highlight the stress used by the interviewee]

This similarly can be observed in more technologically advanced organisations. In the following example from the insurance industry, we can even better see how the postulated consumer conceptualisation potentially emerges. Since consumer conceptualisations here rely almost solely on datafied knowledge, the end-users of this knowledge have no other choice than to trust the correctness and accuracy of that information. The consumer insight is the result of automated machine learning software, where an analysis reveals the relevant and significant information. This is a selection from a lengthy description of the process of this software, in which at no point the accuracy of the results was addressed. Nonetheless, the results – in the form of the data double – are handled as a postulated conceptualisation:

> K4: "*Yeah, so what we do with, with the machine learning, […] we've got lots of information here, we've got dozens or maybe hundreds of bits of information. Let's let the machines tell us what the significant things are, and what we find as significant are things like proximity to the effective date of the policy. In other words, if you do go and get your quote two months before it is due to expire, you know, you are in a different position to a day before. So that was kind of known, but to understand it a bit better, I think, in combination with some other factors was really important. The type of vehicle. And then you can kind of look up against some other information*" […]
>
> I: "*Aaaah, so you are using the Experian profiles then to add to that.*"

K4: "*Yeah that kind of thing,* […]. *But there are dozens of attributes or features. But there are also some in there, where you get a story that you uncover: that's, that is something new! Where you look at it and you think: wow that's interesting, I didn't know that, that those people, you know that those people converted more than those people!* […] *[W]e then extended that to a project where we were predicting what other insurers would price this quote at.* […] *So, what we do is, we learn from that data and we then predict on it. So, we learn effectively how the rest of the market calculates their price. And then, at point of quote, we'll say: so we already know things like, we predict that this person is this likely to convert, this likely to commit fraud, this likely to renew, and also we know that the rest of the market is likely to price it at this. And then in the kind of pricing engine you can then say, well actually we want this person, so what would happen if we reduced our premium, if we are currently the average top five price is this, and we are way above it, let's reduce our price to get into, you know towards the top of the screen. And then you still are both profitable and competitive, and that is really the thing that sets aside (our company) from what other insurers are doing currently*." [Key4_insurance, Pos. 80-82]

What both these examples, as with others, have in common is that datafied knowledge is in use to conceptualise the consumer on a very granular and personalised level and that the conceptualisation is envisioned as being a data double. And in both examples, the participants acted on the data double as if it were an actual one-to-one representation of the consumer. The datafied knowledge seemed to be the most accurate information the organisations had at hand for their operations, and there were seldom doubts about its accuracy. This was even the case where the participants realised that the consumer conceptualisation was only a partial representation of the real consumer, as the following example shows. Despite perceiving that the postulated consumer conceptualisation was only a simplified version of the actual consumer, they still acted on it as if it was a data double:

"*Insight now is directive. It's actionable. It's operationalized. So, not only do I produce for the CEO a view, a simplified view of, you know, who our customers are. That simplified view is actually a simplified illustration of what is an individual, once one personalized. An understanding of what an individual's […]*

propensity to perform a certain action is, at an individual level. And not only do we have it on the individual level, we have it in a way that we can use. When we talk to that person. When that person contacts us and it's that expectation that, it's not just a kind of an additional kind of illustration of the sort of guidance that you might want to take. What this is, is directing actions. Directly directing action rather than indirectly influencing the decision-makers that then instruct the action." [Key1_entertainment, Pos.45]

The data double as an ideal type of consumer conceptualisation thus is thus a slippery and elusive concept. The data double is often described as an improved and more detailed version of the imagined consumer conceptualisation. However, as it is used as if it were a virtual double of reality, the conceptualisation turns into a postulation. The following quote from an organisation in the insurance industry illustrates this slipperiness and inaccuracy well and also gives us an idea of the potential impact of these inaccuracies. Talking about different forms of fraud – at the time of the quote and the time of a claim – the participant stated that there might be a correlation for people committing both forms. However, considering the wording of the interviewee, there were a lot of uncertainties revolving around this correlation – irrespective of potential uncertainties that could have emerged before, such as in the data or while assessing the propensity of fraudulent behaviour. Nonetheless, if the data analysis provided a narrative for fraudulent behaviour by the data double, this was trusted. It also improved the insurance fraud detection statistics:

"[…] *if somebody is manipulating their quote, is there a, is there a way that you could say: Well, we're ok with that. If they're just changing their postcode, is it because they are looking up moving house, or are they actually saying: I pretend I live at my grandmother's house. You know, what are they trying to do?* […] *But actually, what happened was, that the quality of the book of business was so much improved, that cancellations went down. And actually, not just the detection of quote manipulation fraud but there is a, a correlation. It's not, I mean, I wouldn't say it's a kind of very strong correlation. But there is a correlation between people who commit fraud at point of quote and people who commit claims fraud. So actually, the claims fraud went down slightly as well.*" [Key4_insurance, Pos. 65]

Participants rarely reflected critically on the creation of the ideal type consumer conceptualisation of the data double, nor on the actual practical use of the postulated consumer conceptualisation. In some instances, they noted how these types of conceptualisations might be limiting the creativity of marketing, or – depending on the products the organisation provided – the creativity of the product design and development because they "*start briefing creative agencies with spreadsheets, rather than consumer insight*" [F2_beverage, Pos. 30]. Other critical points addressed by participants were seen in the potential of producing too tightly personalised conceptualisations, creating filter bubbles and removing a certain level of randomness that consumers might be expected to have. But the general concept of the data double and its use as a postulated consumer was not reflected on.

5.4.1. Postulations performing consumers

The general uncritical use of the postulated consumer can, however, have actual ramifications for the real consumers. The participants mentioned, on several occasions, how consumer conceptualisations were put into use and in what instances they anticipated the consumers to be represented in the data double and, subsequently, targeted with particular marketing measures. The following sections will exemplify these situations and highlight how they potentially act on the real consumers, mainly by influencing and limiting, to a large extent, their consumption choices, often without the consumer knowing or consenting to it.

In many examples that were provided by the participants, the use of the postulated consumer conceptualisation – in the expectance of it being a data double – often served to pre-select what services and products consumers could access. Participants described how, in their organisation, consumers were significantly differentiated, which translated into the offering of products and services, as well as advertisements and offers, to the consumers. This practice was generally vindicated by the participants in that through the use of these detailed and personalised consumer profiles, consumers would receive more accurate information and offers from their organisations. These were now specifically tailored to the individual consumers and their behaviour. Which, however, also meant that their organisations were specifically selecting and deciding who had access to their offers, their products and services, as well as information.

Several participants from different industries, including insurance, travel and entertainment, said that they were, at least partially, doing data analysis with the aim of anticipating the propensity of consumers to act in certain ways. Some, as in the example below from the insurance industry, were specifically addressing the use of propensity scores upon which certain actions were performed, such as deciding what offers consumers were presented. Despite potentially limiting the selection of offers consumers could choose from or, in the worst case, barring them from having access to offers at all, the potential inaccuracy of targeting the wrong real consumer through the use of postulated consumer conceptualisations was not addressed:

> "*So, the leads to quote was about 12%. And they had no way of prioritising them. So, they might get a lead and think: Well, who do we phone next. You know, we've got these people, they didn't, quote on the website, but we've got their phone number. Who do we call next? And when they just called them on a random basis, they were converting around 12%. What we did, was we gave them a propensity score and said: not only do you have their phone number, but we'll tell you how likely they are to become a quote. And then anyone who is, above 70%, call them. And then, then call the 60%, then 50ies and then the 40ies. And just don't bother calling the others, because it's…* [Key4_insurance, Pos. 78]

The use of propensity or churn scores in combination with postulated consumer conceptualisations had the potential of impacting the pricing of the product or service. Similar to the access to offers, depending on the score attributed to the postulated consumers, the price for the product or service could be and would be adapted: "*we are doing premium sweet spot identification*" [Key4_insurance, Pos. 86]. While the identification of the ideal price consumers were willing to pay could be beneficiary for certain consumers, others had the potential of being negatively impacted through these scores. Here again, participant [Key4] elaborated how, depending on the scores a potential consumer might have received – in terms of propensity of converting, propensity to commit fraud and premium sweet spot identification – their organisation might have decided to increase the price accordingly: "[…] *We'll cost it accordingly.* […] *Just make it worth it, you know. Double the price and if they do it, then you say: ok, yeah, we'll do it*" [Key4_insurance, Pos. 77].

Another way postulated consumer conceptualisations could act upon the real consumers was through the potential of nudging and manipulating. While manipulation through marketing is not a new development and has been discussed since the 1950s (cf. Wertenbroch (2015a)), the possibilities for manipulating consumers have progressed significantly with the implementation of big data analytics, as some examples show. Limiting the choices for consumers to select from generally has the potential of nudging them in the direction that is desired by the organisation. As an example from an organisation in the travel sector shows, this seems to be done without deliberating too much on the potential of nudging or manipulating consumers. With the vindication of consumers benefiting by getting a different – and potentially better – rate for a hotel room, the goal is rather to nudge consumers in the direction of filling out rooms that would stay empty otherwise:

> "*In essence what we are looking to do is, we are trying to take some of the unused capacity of a hotel, and to help effectively sell that, eh those rooms that would otherwise remain empty, at different rates in order to steer consumers towards helping that hotel, for to fill out its capacity over time.*" [J1_travel, Pos. 16]

While in the example above, the nudging tends to be only marginal, in other examples this has the tendency of going further. Sometimes, the aim is not only to influence consumer behaviour but rather to change consumer behaviour and, as such, the consumer as well. As the quote below from an interview with a consultant in digital marketing and behavioural change shows, the objective is to turn the postulated consumer more into what the organisations want the consumer to be instead of what the consumer momentarily is:

> "*I've got the kind of academic grounding in behavioural change psychology that, that the consultant and the company have me also work on our projects for clients and sometimes that will be helping global companies with behaviour change, maybe of their customers on their website. And so for them, those are people who are in, so we're helping people who are in marketing functions, who fit us up, like to design, for example, their user experience to change their customers' behaviour and the way they know whether that's working is through data and yeah, that's quite a powerful application of that.*" [Key5_consultant, Pos. 18-19]

In all these examples of how consumer conceptualisations were used in everyday marketing practices, participants rarely addressed the potential of relying on inaccurate data doubles, misrepresenting the actual consumer and impacting them in a potentially significant way. Some of the participants were partially aware of the potential impact conceptualisation and marketing practices can have on consumers. Although concerns rarely arose and, mainly, the positive impact on the consumers was highlighted, in some cases participants acknowledged that not everyone would profit from the developments in marketing and big data analytics. However, most participants who did raise concerns did not address the potential disconnect between the use of a data double conceptualisation and the postulated consumer. Instead, the concerns were more of a general nature related to the negative impact on consumers or society. As an example, the pricing sweet spot prediction was a case that, for some participants, had the potential to cause a serious negative impact for consumers. With the pricing sweet spot prediction, their organisations were trying to find the highest price for them to make a profit, which had to coincide with the highest cost the consumers were willing to pay without putting them off. The participants considered that because of the implementations and uses of big data analytics in marketing, some individuals would pay higher prices for products and services in comparison with others and that this would amount to discrimination.

Other concerns that were raised revolved around the potential of losing privacy and of consumers being followed around – metaphorically – by organisations. In these cases, participants often addressed these issues from their view as a consumer instead of in their role as a data analyst or marketer involved in these practices. These surveillance practices were often discussed on a broader level and involved the issues of data sharing and data selling through data brokers. While the consensus appeared to be that the responsibility of acting in an ethical and sensible way regarding personal data lay in the hands of individual organisations, almost all the participants considered their organisation to be acting in a responsible way. However, some participants – in their views as consumers – clearly did not seem to trust other organisations in doing so. The following quote exemplifies these two different roles and perspectives. Here the interviewee talks on the one hand as a product manager in marketing and big data analytics in the travel industry and on the other hand as a consumer:

"*So basically, yes, in (company), we are trying to be as responsible with our data as possible, of course. But then at the same time, we do have the paid growth campaigns and we do know that a lot of users find them very useful.*

However, I personally am not one of them. I hate those campaigns. I hate those ads following me ever after. So, it has become a routine for me to just do everything in incognito mode now on the internet. Basically, I only go to a bunch of trusted websites in normal load. Everything else, if I want to check something, okay, incognito tab by default." [I3_travel, Pos. 119-121]

5.5. Conclusion

The analysis presented in this chapter has shown that on multiple occasions different consumer conceptualisations can emerge in organisations. These consumer conceptualisations not only differ between before and after the implementation of big data analytics – and, thus, between traditional knowledge about consumers and datafied consumer-oriented knowledge – but there also appears to be a disconnect between the idealised data double and the practical use of the data double, the postulated consumer conceptualisation. This disconnect seems to stem from differences between the expectation that a data double can be produced which accurately represents the individual consumer and the practical implementation of this expectation. And while, in some instances, certain participants were aware of the potential impact of marketing procedures, such as personalisation and the targeting of individual consumers, there seemed to be almost no reflection on the accuracy of the data double and the impact of the use of a postulated consumer conceptualisation.

In the following chapter, I will further elaborate on this disconnect between the imagined consumer, the data double, and the postulated consumer. Addressing the second part of my research question, the next chapter will focus on why participants conceptualise consumers differently, depending on whether they rely on traditional market and consumer methods or big data analytics for consumer-oriented knowledge. As the descriptions by participants revolved around the knowledge that had been used for consumer conceptualisations, I will focus on how this knowledge was established. This will give a better understanding of the reasons why consumers were conceptualised differently when the participants and their organisations were using big data analytics,

and how the ideal type of data double seemed to persist so much – turning it into a postulated consumer conceptualisation.

6. Co-Producing Knowledge in datafied organisations

6.1. Introduction

To understand why different consumer conceptualisations were established in the participating organisations, I will focus in this chapter on the epistemology of the conceptualisation practices. The idiom of the co-production of knowledge established by Jasanoff (2006b, 2006c) can be a useful framework to analyse the epistemology of these practices and provide an understanding of how the construction of knowledge can lead to these different conceptualisations. As I have laid out in my theoretical framework in Chapter 3.2, while the idea of the co-production of knowledge initially was aimed at studying the process of sense-making in science and technology, it also helps in understanding sense-making in other knowledge production contexts, such as in market and consumer research. Linking the cognitive, material, social and normative procedures and settings of knowledge-making can help in understanding how consumer-oriented knowledge is being created. These different procedures not only have an impact on traditional knowledge about consumers, resulting from qualitative research methods such as interviews, focus groups and experiments, or quantitative research methods such as surveys, they can also be seen to be coming together in big data analytics and the co-production of datafied consumer-oriented knowledge. The way participants have described how data analytics is being done in their organisation reveals how the knowledge is not simply extracted out of the data. Instead, data is manipulated, handled, processed and acted on by a range of different actors, to result in a kind of datafied knowledge that is further communicated and, through this, is further processed.

The focus on the epistemology behind the conceptualisations is done for two reasons. First, it can help us to better understand why participants consider the imagined consumer conceptualisation an insufficient tool for marketing and why big data analytics is perceived as a better source for knowledge about consumers. Second, the way knowledge is being established can also account for much of the slipperiness that is observed in the unachievable data double, producing the postulated consumer conceptualisation. Consequently, the findings of this chapter address the changing epistemologies within the participating organisations, which seem to have occurred through the increased use of digital data and big data analytics. The epistemology behind the operations can be largely

divided into two sections. The first focuses on the actual process of co-producing knowledge, i.e. how participants describe the procedure of big data analytics and the potential difficulties and uncertainties that can emerge. The second focuses more on the resulting datafied knowledge and how its epistemic authority is established, contributing to the uncertainties that emerge in its co-production to be largely omitted.

The first section addresses the co-production of datafied knowledge. As discussed in Chapter 3.3, to date, previous studies that have focused on the practices of marketing and, specifically, on the segmentation or conceptualisation of consumers have done so by looking at the epistemology behind these practices (Grandclément and Gaglio, 2011). With the focus on the co-production of datafied knowledge, I will discuss here the epistemological practices used in establishing it as described by the participants. These are practices that are intended to be made more scientific through the use of big data analytics: witness how the participants often referred to the more objective technologies and the more accurate procedures as representing the 'real' consumer. Simultaneously, they describe a complicated process of creating datafied knowledge which involves many different procedures and influences, and which was described by participants as a slippery and uncertain process. Data analytics is not done in a sealed, neutral environment, but is affected by many different influences, such as the multiple teams and individuals involved, the technology, or the normative boundaries in which their organisations are embedded. Furthermore, the participants described how datafied knowledge could be inaccurate or erroneous, emphasising that the data double could be inaccurate or erroneous and, therefore, could not be a data double of the consumer.

In the second section, I will focus more on why the conceptualisation of the data double as an accurate representation of the consumer persists, despite the difficulties and uncertainties in co-producing knowledge. The role of the epistemic authority and the credibility of datafied knowledge in this is relatively important. I will discuss how the epistemic authority of datafied knowledge is established. On the one hand, this is through the simplification and translation process the knowledge is exposed to throughout its communication and dissemination, in which uncertainties tend to be dismissed. On the other hand, there is the dataist perception of datafied knowledge (van Dijck, 2014) being bigger, better, and faster, and its contribution to big data analytics being perceived, by the participants, as a superior method for gaining knowledge. The superiority of datafied

knowledge emerges, in particular, when it is considered in relation to the other kinds of knowledge. This all creates a kind of epistemic authority of datafied knowledge that is not only perceived to be superior to traditional kinds of knowledge but also appears to lead to an unquestioned use of the data. As such, it can be used as an explanation for the observed disparity in the representations of the real, and the multiple ontologies that emerge as a process. The co-production of knowledge, thus, serves as an explanation for the disparity in the expectations and meanings of datafied knowledge in the data double, traditional knowledge in the imagined consumer, and the practical use of datafied knowledge in the postulated consumer.

6.2. The co-production of datafied consumer-oriented knowledge

The disparities between the imagined consumer conceptualisation, the data double conceptualisation and its implementation in practice in the form of the postulated consumer are an outcome of the co-production of knowledge. The participants discussed the differences between the traditional types of knowledge, which was perceived as subjective and inaccurate, and the datafied knowledge that was expected to be an exact representation of reality, being precise and more complete, while also discussing how the latter contains uncertainties and errors. Looking at the practice of co-producing datafied knowledge as described by the participants provides an understanding of how these different perceptions and expectations of knowledge come to be.

The process of co-producing datafied knowledge entails several steps, as the participants have described. Normally, it starts with a decision-making process in relation to what is considered important and what should be achieved through big data analytics. This involves the practices of hypothesising and operationalising, as well as defining Key Performance Indicators (KPIs). But the analysis of the data in itself is a practice that is influenced by different procedures, such as technical limitations. Furthermore, the participants also mentioned the importance of interpreting in addition to communicating datafied knowledge in the process of making sense of data. In general, the whole process of co-producing knowledge is not just a linear process but is described as iterative and cyclical, in which the end-point – the datafied knowledge – can be elusive. All in all, the detailed accounts the participants provided regarding the co-production of knowledge through big data analytics helps to see how this slippery process is entangled with social

and organisational settings, such as the different teams and the expectations concerning the knowledge; the technological limitations, such as the algorithms; the cognitive procedures providing such limited understanding of the technological functions; and with the norms, such as the predominant capitalist functioning of profit-making in organisations. These will be elaborated upon in the following sections.

6.2.1. The sense-making of data – cognitive processes of co-producing datafied knowledge

> *"Big data […] is actually not valuable unless you know what's meaningful."* [Key1_entertainment, Pos. 59]

This first quote is a description of how the participants saw (big) data, and of the importance of the cognitive processes involved in the sense-making of data. These cognitive processes include the many mental operations necessary to turn data into knowledge, such as problem-solving, reasoning and intelligence. The participants were aware that just the possession of big data in their organisation was not enough for having datafied knowledge The data had to be made meaningful; there was a sense-making process involved in achieving the value of big data. The way of making data meaningful was described by many participants as being a specific process in its own right, that of big data analytics being a practice of knowledge co-production. Looking in detail at these cognitive processes provides a first understanding of how knowledge can lead both to the expectation of a data double consumer conceptualisation, but one that is, in practice, not at all a data double but rather a postulation.

Participants often described the co-production of knowledge in their organisations extensively and provided significant insight into what they considered to be meaningful data. There are, for example, some prerequisites that can assist in the process of co-producing knowledge, such as the necessity of understanding the wider business and the market to be able to create a kind of knowledge that is useful and impactful for an organisation. This step is described as specifically important for data analytics as it is crucial to know what the individual teams are doing, what is considered to be important information, and what should be addressed during the analysis. In other words, it is a process of understanding what variables are important, discovering what connections

within the variables could be significant and finding out what results the teams are looking for.

The practice of making data meaningful, i.e. that of datafied knowledge co-production, follows a pattern that is common to any traditional research design process. The first step in knowledge co-production was considered by the participants to be hypothesising. What questions to ask and what should the answers tell us was seen as a significant phase. In the ideal case of a big data analytics project, the participants pointed out that, in this process, enough time needed to be committed to deliberate with all the people involved over the questions that had to be addressed through the analysis. Yet, despite its apparent importance, this process was seldom as straightforward as planned but, on the contrary, could be slippery. Instead of it being a standardised structured process, participants described how the establishment of hypotheses happened along the way, with new ones being brought in by others in their organisations or emerging from the data and being the result of a constant ongoing and iterative process. Consequently, the participants were, at times, ambivalent about the soundness of the hypothesising in their organisations and mentioned that they would have preferred the development of hypotheses to be more of a standardised process, with fixed routines – an approach considered to be ideal for use in data analytics.

The next stage within the practice of co-producing datafied knowledge was described as the process of operationalising. Operationalisation tends to be aligned with the hypothesising and is frequently portrayed as requiring a thorough consideration of all the parts that relate to the problem or the issue that is being tackled. This is also described as how to answer the questions that are emerging as hypotheses. Here the previously described pre-requisites, such as knowing how the business operates, what individual teams require, etc. are also depicted as an important aspect for successful operationalisations. In the ideal case, as with hypothesising, operationalising is also done within a broad team of individuals that are directly involved in the co-production of knowledge, as well as by the outcome of the datafied knowledge. The following quote from a digital insights manager in an entertainment organisation explains the importance of the ability to fully define the problem. In particular, with big data analytics, without these processes of hypothesising and operationalising, the results are regarded as having no value. It is worth noting that she draws a comparison with traditional market and

consumer research and classical data analysis, where she considers the procedure of doing a thorough hypothesising and operationalising as not being as important:

> "*I guess the other challenge... Is probably knowing what to do and knowing what not to do, I think. So, when you are working with more traditional data, it's easier to let the data do the talking and you should do, you know, it's important. Therefore, it's quite easy for business areas to be lazy about really thinking through the problem.*
>
> *And, I think in the world that we're living in today, where pretty much anything is measurable, you can't do that. You can't approach a challenge, a business question, without properly defining the problem. Big data does not help you unless you find the problem. I've always jokingly but partly seriously said that I can get the data to tell you anything. Really can. So, tell me what!*" [Key1_part 2_entertainment, Pos. 23-24]

While thorough hypothesising and operationalising are described as the ideal case, the participants felt that, especially with data analytics, these cognitive processes risked being done by data analysts without direct consultation with other relevant individuals or teams. They were aware that this had the potential of resulting in situations where the data analysts failed to consider certain aspects that might have been relevant for some problems. These could have included missing out on relevant variables for the analysis, or not focusing on the right variables. However, the issue of not comprehensively collaborating during this process was often described as more of a problem at the beginning of big data analytics projects, when the analytical processes were still rather new and the hypothesising and operationalising had not yet been entirely adapted. As soon as these operations became a common practice in their organisations, the participants no longer considered this close collaboration as a necessity. This might also be why they mentioned that, ideally, these processes should be automated as soon as the hypothesising and operationalising of certain recurring issues had stabilised within their organisations.

With data analytics, a main aspect of the operationalisation is the establishing of certain metrics or defining KPIs. KPIs and metrics were seen by the participants as providing information about specific aspects of their organisation, such as sales or the impact and effects of marketing campaigns, in a numerical format. But these metrics are also

increasingly used to define and measure the consumers of their organisations. KPIs and metrics were portrayed by the participants as an easy way to categorise, compare and rank aspects of the consumers, and relate them to the products and services that were part of the market. KPIs were described as central to data analytics, as they allowed for the integration of consumer characteristics into the analysis and, thus, the assessment of the consumer on a numerical basis. Furthermore, they provided the possibility of evaluating the progress of all kinds of measures that were being implemented in their organisations, such as the success of marketing campaigns.

Consequently, as can be seen, the process of making data meaningful is a lengthy one, even before any type of data analysis is done. Particular attention is paid to these initial steps as they are considered crucial for the end-result of acquiring datafied knowledge. The participants suggested that these initial procedures should be done thoroughly, as this would influence the outcome and was seen as a necessity for obtaining "quality insight". The following quote, which is a continuation of the previous quote, further exemplifies the importance of hypothesising and operationalising in relation to big data analytics:

> *"If you don't think through the problem properly, in a big data world you get lost! You know, you just spend years and years and months and like in a quagmire and you don't really know, because it's not direct. It's not... It is so big and so enormous that you can go round and round in circles. So, you've... It forces you and as with everything, with data you learn this through trial and error, you learn it by getting it wrong, but it forces you to fully define the problem and ultimately you get much better quality insight as a result."* [Key1_entertainment, Part 2, Pos. 25]

Another important step that was described by the participants in the cognitive process of knowledge co-production is the actual step of researching and doing the analysis. In terms of big data analytics, the participants predominantly referred to the modelling and the use of variables for the models, as well as doing the statistical analysis in itself, which also included the constant refinement and adaptation of the existing models. A particularly important part in the analysis of big data was the automation of the processes. As most participants described it, ideally, the initial phase of data analytics aims to set everything in place, using algorithmic or machine learning models – depending on the analytical

capabilities – and decide upon what variables are correlating or providing the most accurate results. As soon as those steps are refined and in place, an automation process enables a constant flow of updated datafied knowledge for the end-users to use or provides them with simplified models so they are more flexible in selecting what variables to include or exclude, and are able to do part of the analysis themselves.

Lastly, the interpretation of the results was described by the participants as an important aspect of the knowledge co-production process as data and the results from the analysis were not seen as being self-explanatory. The participants specifically referred to the importance of people being able to interpret and understand the results of the analysis. The steps were similar to the operationalisations prior to the analysis, such as making the connections between the data or the results of the analysis and the problems to be solved. Again, this required a good understanding of the general operations of the business and its problems: knowing what kind of insight the end-users and teams required and how they would operate with this insight. Also here, cyclicity was a recurring component of the cognitive procedures of knowledge co-production. The participants had mentioned how, in ideal cases, the analysis and the interpretation of the results would lead to new questions, that again would lead to further research and analysis. As the next example shows, although general interpretations of the results seemed to happen, a systematic and more holistic reflection and interpretation of the data analysis and its results was not done in a regular way. Which, again, had the potential of leading to a situation in which either the interpretation, or the follow-up hypothesising and next research cycle, could be flawed:

> "*I mean, I know as a major label, we are lucky to have a data-insights team, a market research team, as well as a business intelligence team. But still… I mean (.) so in an ideal world, we would reflect – all together – after every case, everyone would have to prepare something, and we would study and interpret the analytics thoroughly. A bit in the sense of "where did we go wrong?" Just so we can improve on it next time*." [M2_entertainment, Pos. 60 – translated from German]

This quote highlights the cyclicality of the cognitive process of co-producing knowledge. The interviewee specifically explained how the improvement of the analytical process could benefit from critical reflection and interpretation. This would potentially lead to

more questions that would be addressed in a new cycle of hypothesising, operationalising, and analysing.

6.2.2. The sense-making of data – social processes in the co-production of datafied knowledge

As already partially addressed in the previous sections, big data analytics is often considered a process that involves multiple teams and the contribution of more than just data analysts. This shows that the co-production of datafied knowledge is not an isolated process but happens within an organisational setting, in which social processes take place. The participants across all the organisations noted the importance of a collaborative effort to make the data meaningful. Particularly in the early stages of hypothesising and operationalising of problems for the analysis, a significant team effort was believed to be required. These efforts often consisted of different teams and meetings coming together to decide on how they were going to operationalise and what they wanted to measure. Defining key metrics, setting goals and timeframes, as well as considering feedback measures were often seen by the participants as a collaborative process involving everybody who could be affected by the datafied knowledge. This process of collaboration included mainly senior-level decision-makers from different departments and, as a result, the organisational and hierarchical structures of companies tended to be reflected in these processes. A lot of decision-making around the co-production of knowledge was predefined by senior members of their organisations. They were in the position of deciding which aspects were currently deemed important knowledge, and which were not.

Regarding the collaborations between different teams and individuals within their organisations for the co-production of datafied knowledge, the participants indicated that these were increasingly necessary. A recurring theme that was often raised was the so-called organisational silos – insular, cut-off teams within organisations with whom communication did not take place and whose decisions impacted the data infrastructures. While the participants did not reveal whether the existence of these silos created difficulties in their pre-data analytics organisations, they were regularly mentioned in relation to big data analytics. Based on this, the process of data analysis was described as being inhibited and made increasingly more complicated through the presence of these silos in their organisations, which subsequently risked to slow down work processes,

decreasing the efficiency of operations and impacting the soundness of the datafied knowledge. At the same time, the participants also seemed to be hopeful that through the (further) implementation of big data analytics technologies, these silos would be broken up. This, ideally, should make the co-production of datafied knowledge easier and increase the flow of datafied knowledge throughout their organisations.

The potential for "breaking up" these organisational silos through the implementation of big data analytics relates to this necessity of having to collaborate with different teams to make data meaningful. The relationship between the data analytics teams with other teams was mostly described as being required to occur at the early stages of the co-production process, as seen above. In some situations, the increased collaboration and subsequent removal of silos even brought forth new roles within organisations, such as (digital) insight managers. They mainly dealt with the aspects of knowing what was happening in the wider business world and needed to be able to focus on what was seen as important in the knowledge co-production process. This role is described by the following quote from a data scientist in the travel industry:

> "*So, I have spent a lot of time listening to my colleagues, who, maybe, don't have a mathematical background, trying to understand, what were their real needs, what did they need to work more with data? Like, is this a case of do we need to teach about different tools that we can use, is it that we need to talk about the mindset you are trying to apply? Really there has been a lot of listening, coaching, bringing people along with us on that process*." [I1_travel, Pos. 62]

All these examples of the co-production of datafied knowledge and the involvement of cognitive, social, and organisational settings have already given us a first understanding of the intricacy of the different processes. Seeing how important they are shows that in all these steps that have been described, errors can occur, and negotiations have to happen for datafied knowledge to be co-produced.

6.2.3. The sense-making of data – technological processes in the co-production of datafied knowledge

While the social and organisational structures, as well as the cognitive procedures, play an important role in the co-production of datafied knowledge, so does the technology behind big data analytics. The role of technology is significant for big data analytics and

there are several aspects that emerge which influence the co-production of the knowledge. As described by the participants, data analytical technologies were relatively complex and consisted of many different parts and pieces, including the hardware, the software and the data itself. Each of the individual parts of big data technology could influence the way the participants co-produced datafied knowledge. As an example, the data was complex and participants had to address situations in which it, being a part of the big data infrastructure, had created issues. These related, for instance, to the multitude of data sources, external as well as internal, that were used for the analysis. Some participants also described how pre-existing data infrastructures still remained in a none-structured format or were archived in analogue repositories. Or they talked about the difficulties they had encountered accessing and using all the different external data sources that existed. All these different issues influenced the analysis of big data and the accuracy and reliability of the corresponding datafied knowledge.

Another part of the big data infrastructure is the software and the models used for data analytics, which affect the co-production of datafied knowledge. As seen above, the use of algorithms is crucial for the successful analysis of big data due to its automated nature and, at times, the participants referred to the use of machine learning algorithms to further improve their analytical capabilities. However, they also reported how the use of these algorithmic models impacted the way the analysis was done and the accuracy of the datafied knowledge. As the quote from one interview with a business analyst in the travel sector shows, outdated algorithms were believed to be responsible for not providing the required predictions or, in other words, the expected results. But, as the process was perceived as a black box, it was impossible to know. Interestingly, in the second part of the quote, he expected that, with the planned introduction of artificial intelligence and machine learning, this problem would be eradicated, allowing for better predictions:

> "*Yeah, the, the system currently is meant to be able to predict what's going to be happening, but it's not, the algorithms are so outdated and likely it's a bit of a black box we don't actually know what's going on in the middle. Where the new ones, sort of the second phase of it is going to be more. It's about artificial intelligence and machine learning, all the stuff that they're developing now. But with the ability to see how it's doing it. So, you can say actually that, that forecast*

isn't working I'll change it to a different model. Tweak that. So, I think that will do better predictions." [L4 & L4.1_travel, Pos. 79-80]

The software or the models are, however, not only responsible for doing the analytical calculations and operations, but they were portrayed by the participants as the main tools to ensure that the datafied knowledge could be disseminated effectively. For example, the software Tableau was described as providing a significant improvement in the communication of datafied knowledge, particularly between data analysts and end-users. As the following example from a commercial manager in the travel industry shows, their recent implementation of big data analytics technologies, and specifically the software Tableau, improved the way knowledge could be made accessible. Although it was not an organisational silo in the strict sense of the words, they still had many people who were not working on-site, rendering them insular, cut-off teams within the organisation with which the communication of knowledge had been difficult before, but had now improved through the use of the new software tools:

"*Now what Tableau has now enabled us to do is very quickly and in a very visual way for frontline people to understand and interpret. You share that information with them, personalize it to them or to their team leader or to their region* […]*"* [L2_travel, Pos. 73].

Frequently, participants referred to how these software tools helped to visualise complex data in such a way that the results were understandable for most. This process involved a simplification of the data, the analysis or the results and the participants referred to this simplification as a translation of the procedures and technical terms into layman's terms. This was portrayed as necessary since the technology used for big data analytics could be complicated, as well as the analysis itself which required a certain amount of statistical knowledge. This simplification of knowledge in the process of its communication and stabilisation was a common occurrence in the co-production of knowledge, as I have already described in the theoretical framework chapter.

While these examples of technological influence on the co-production of knowledge referred more to individualised pieces of the big data infrastructure, there were other examples in which the implementation of the technology had created difficulties, potentially impacting the co-production of datafied knowledge and its use. The

participants described how these difficulties often emerged amongst employees in their organisations who either had been working with a certain technology, like Excel, for a long time and now had to change, or amongst employees who had relied on entirely other means for co-producing knowledge. In these examples, the implementation of big data technologies was perceived as problematic. The increased use of data as a form of co-producing knowledge and making decisions can involve procedures which are new and challenging, sometimes requiring the learning of new skills and adaptation of routines. Big data analytics also led to a perception that participants had to justify why things have been done in the way they have over recent years. The implementation of big data analytics requires negotiation and persuasion if different individuals or teams are to accept it and use the corresponding datafied knowledge. A good example to illustrate the necessity of persuading others stems from an interview with a digital insight manager in the entertainment sector. She explains how her work and her role for two years had been to carry out specific pieces of data analytics in the form of recommendations that were mainly intended to slowly introduce some new big data technology into a specific team that was resistant to its implementation, and to illustrate to them the potential advantages that would come with its use:

> "*So as an example, if I delivered a recommendation, so I've done a piece of analysis. I delivered a recommendation that meant, that said: you know what? You normally send that piece of below the line marketing, or whatever, to a million people. I now know, I can tell you that 250,000 of those will not respond. They are not going to open it. So only send three-quarters of what you were sending. We would agree with Finance the monetary saving of over $250,000 not spent. It was a saving in principle because, obviously, that money was still in the budget and everything else. But what it was, was a visible line on the P&L* [profit and loss] *for the contribution made by the insight delivered. And it was purely a PR tool and it was an internal PR tool. That was my entire job for two years and what that did was it went slowly and it took a long time. It felt like a long time. But it slowly developed a common understanding and just expectation that insight delivers financial value. And it worked and then we stopped doing it because we didn't need to do it*." [Key1_part 2_entertainment, Pos. 17-18]

6.2.4. The normative procedures of datafied knowledge co-production

In terms of the normative procedures, there were two specific aspects mentioned by participants which could have a significant impact on the co-production of datafied knowledge. The first related more to the technical side of big data analytics, particularly the collection and processing of data. Here, laws and technical standards are the main contributors to how datafied knowledge can be co-produced. As an example, many participants explicitly mentioned the European General Data Protection Regulation (GDPR), that came into force in May 2018 and which altered the way their organisations could collect data and use it further on, mainly for their consumer profiling. The participants talked specifically about examples of how these practices were changing as a consequence. The following example stems from an interview with the head of digital experience and digital marketing in the travel industry and refers directly to the emergence of the GDPR. As result of the GDPR coming into force, her team had not been able to collect as much data as they had been doing before since these practices were no longer legal because of the new regulation. She believed that this had directly influenced the datafied knowledge, as well as the conceptualisation of their consumer:

> "*And I suppose it was a lot easier before GDPR. So, we had access to… Before launch, the strategy specifically in the States, because it is such a massive audience, was to really forensically target people based on their belief system and really align to those people, who are aligned with the brand traits of (the product). So those progressive, innovative, digital savvy, liberal, educated individuals. And that was, that could've been done through the usual first-party data, through Google. But we supplemented it at the time with third party data, such as Mastercard purchasing behaviour. So, we know that if people were more likely to donate to LGBT causes or social enterprises, they were right what we were aiming at. We don't have access to that data anymore. So, it is a lot harder to be able to forensically target these people based on their beliefs and their digital behaviours.*" [Key3_travel, Pos. 15]

Other normative influences which impact the co-production of datafied knowledge revolve around the perception and expectation of what data can achieve, relating to the different ontologies and worldviews that can be attributed to the different consumer and

market research methods. The participants often displayed a kind of techno-fundamentalist, digital positivist worldview of objective measurements produced by big data technologies when they were talking about how data analytics co-produces knowledge – and thus situated themselves in a realist ontology when it comes to how consumers can be conceptualised. As discussed in the literature review in Chapter 2.4, this relates to the ideology of 'dataism' as described by van Dijck (2014). The conception of digital positivism and dataism amongst participants is regularly connected with the intention of reducing the amount of decision-making in their organisations, which is based on a 'gut feeling' and subjective measurements. This was highlighted multiple times by a participant working in digital sales in the entertainment industry:

> "*I really enjoy this analytical approach, as it allows you to learn so much from it and it is not based as much on making gut-feeling decisions. It was different when I was doing my internship in the promo department, where a lot was about, ok, you need to know the people and gut-feeling: they could like this, or dislike that. Here it's more, you can look at the things and you know what you can do differently.*" [M1_entertainment, Pos. 26 – translated from German]

This conception also relates to the more managerial, economic principles that are persistent in marketing theory, which include positivist worldviews to argue for efficiency and measurement-based decision-making. The participants mentioned how they perceived decision-making to be reliable, mainly when they were being measured and reported quantitatively and economically. Objectives and achievements were portrayed in a way that meant they could not rely on being the result of simply gut feeling but had to be supported by something measurable, specific, achievable, and relevant. Similar approaches could also be seen regarding the initial investment in big data analytics technologies, where one of the key metrics for success (or lack of it) was its return on investment (ROI) as an economic metric.

However, regarding defining KPIs and deciding on what was measurable and how it could be measured, the participants talked about the difficulties involved in these practices. The following example stems from an interview with the head of digital in the beverage industry and gives us an understanding of how the ideology of dataism could impact the way datafied knowledge was being co-produced. The quote addresses the idea that

although the intention was to make as many things quantifiably measurable as possible, there were limitations as to what could be measured and how it could be measured. As we have seen above, the collaboration between different teams and individuals in the co-production of datafied knowledge was quite important to ensure that everybody knew what was possible and what was not, and how to properly hypothesise, operationalise and define the metrics. The interviewee mentions that otherwise, there was a risk of situations arising where the results were adapted in such a way to retrospectively "fit the story". This example furthermore shows that although big data analytics leads to a shift in paradigms towards (digital) positivism and dataism, it appears that depending on the situation, softer form of positivism emerges, as marketers and analysists are dealing with consumers and their behaviour which is difficult to quantify accurately:

> "*And, really, maybe then, after the event, when we go to do an evaluation, we realise in the path, well, we actually don't know how to evaluate that. It all sounded great on a PowerPoint slide [sound slip] how we quantifiably measure that. And I think bringing an analyst, a data person, into those discussions really early on helps to make concrete what you can and can't measure. And it's not to say that we propose that we measure absolutely everything that's part of a project or a campaign. But, you're clear about what we are going to measure and how, and we set ourselves… What are the KPIs and which are the relevant metrics to address each of those KPIs? Rather than doing it retrospectively and make the numbers fit whatever story we want, which is often the case with marketing people to retrospectively justify the investment.*" [F2_beverage, Pos. 74]

Thus, as these sections show, the co-production of datafied knowledge is quite a lengthy process that is intertwined in organisational procedures and structures, social and cognitive processes, with technological infrastructures as well as different norms. The participants have shown that the process of co-producing knowledge is not at all straightforward and it can be assumed that it requires many different iterations, depending on the organisation, the industry, or the maturity of the analytical operations. However, the slipperiness of the datafied knowledge becomes even more apparent when looking more specifically at the uncertainties, errors and the potential difficulties that can emerge through the analysis of big data. This topic often only came up after specifically being addressed by the interviewer or only tended to appear as side notes by the participants.

Also, at FCEs, the uncertainty of knowledge or difficulties in big data analytics were rarely specifically addressed and, if they were, they were rarely seen as part of the main discourse. However, if they were addressed by the participants, it became clear that in the process of the co-production of datafied knowledge, uncertainties of all kind started to emerge.

6.3. The uncertainty of datafied knowledge

Looking specifically at how the participants talked about uncertainties in knowledge and how they dealt with this in their organisations, several aspects emerged. Uncertainties were more acknowledged by the participants when they were talking about traditional market and consumer research methods or tacit forms of knowledge, such as experience. These were often evaluated as uncertain due to their subjective nature, which was perceived to be reducing the accuracy of the knowledge. Regarding datafied knowledge, however, participants rarely explicitly mentioned the potential uncertainties in its co-production, although they were present. When describing the process of co-producing datafied knowledge – as we have seen above – it becomes apparent that big data analytics involves many different aspects and steps, and the resulting knowledge potentially depends on a variety of actors, technologies, circumstances and ideologies. All of these have the possibility of (unwittingly) affecting the accuracy of the knowledge, bring in subjectivities, or even create false results and errors – and, thus, introduce uncertainties into the datafied knowledge.

The entire process of trying to make data meaningful and being able to turn data into actionable knowledge has several steps that have the potential of resulting in inaccurate, or even false, knowledge. During the process of establishing hypotheses, as well as during operationalisation, a certain number of subjective judgements are necessary and, as stated by the participants, also require the experience of data analysts. Making wrong assumptions, failing to consider all the necessary aspects, or focusing on irrelevant variables can create results that appear to be correct but are not suitable for the problem in question. I have already addressed the concerns raised by the participants regarding these processes being done by a data analyst on his or her own. But there are other, more subtle examples. The following is from an interviewee who works as a head of digital experience and marketing for an organisation in the travel industry. As she said, they

redeveloped their homepage, and a lot of work went into the analysis before, during and after its implementation in order to assess the impact and the effectiveness of the redesign, all of which resulted in findings they were expecting. However, upon launching the website, certain barriers emerged that were attributed to problems in the terminology which had not been considered before and could not be uncovered by data analytics:

> *"Well, often you would also get, conflicting insights. So, at the moment, we are looking to re-platform our website. We are in discovery stage at the moment, and we are just having a look at the current users, 'cause it's not a new product, it's already there, we are just looking to iterate and improve on it before we re-platform. So, we are having a review of all the user journeys and testing them. So, in the first instance, we looked at the Google analytics data. Are people doing what we want them to do? Are their user journeys working through all these different pillars? And Google analytics said: YES! [both laughing]. So really great. But then we took it to the live remote user testing, there were barriers, and there was, for instance, with the business journey, they identified that there was a problem with the naming of the website. People in the States didn't understand some of the terminology we were using on the website. So, that wasn't identified through analytics."* [Key3_travel, Pos. 30]

Other uncertainties that have been addressed by participants refer to the difficulty of quantifying or benchmarking certain features of a problem that needed to be solved. As we have already seen, the participants or data analysts were aware of the difficulties of trying to quantify as much as possible and knew the limitations of the analysis. While in some situations, KPIs and metrics were perceived as a good tool to evaluate developments, they did not seem to be fit for every purpose. Looking at how certain participants described the process of establishing these metrics provides some interesting insight into the potential uncertainties of datafied knowledge. While, generally, operationalising and interpreting were described as standardised steps, in that they are necessary and done in every instance of a big data analytics project, the process itself was often far from standardised. Instead, the participants talked about how this involved intuition knowing what variables to pick from a large data set, or how it sometimes was simply a case of guessing and employing trial and error in order to find the right variables, as the following quote shows:

> *"So, pick a couple of key metrics and then, literally, to be honest, sometimes, it's a bit finger in the air. Okay, so if we increase that by 15%, what do those numbers look like? Do we feel that's achievable? And then, in as many cases as possible, challenge ourselves and stretch on top of that."* [F2_beverage, Pos. 102]

The involvement of intuition in picking variables, experimenting with different metrics and the use of gut-feeling in big data analytics is interesting as traditional types of knowledge are often deemed subjective and inaccurate exactly for those same reasons.

Another uncertainty in the co-production of datafied knowledge that has emerged from the interviews relates to the inability of the participants to know how datafied knowledge was co-produced. In particular, the use of algorithms had the potential to obfuscate the operations of the data analysis, thus turning the algorithm into a black box, as we saw in the previous section. It was, for the participants, at times quite difficult to know if the resulting datafied knowledge was accurate or erroneous. This is an issue that has attracted increasing attention from researchers in recent years (Pasquale, 2015).

In some situations, the participants were partially aware of the potential of uncertain or erroneous outcomes of big data analytics, depending on the type of analysis and on the interpretation. This view of data analysis is not prevalent but was articulated in some instances by the participants. In these cases, they particularly addressed the difficulties that can emerge through big data analysis, such as the difficulty of representing minorities because they do not fit the model or are underrepresented in the data. Others also addressed the 'politics' of data, meaning that, depending on the recipients of the datafied knowledge, it is worth adapting the analytical process and the interpretation of the results to 'fit the purpose' – i.e. highlight certain values and results in meetings to have the results look better than they are.

On a similar note, there seems to be an awareness that a belief in dataism, in the sense of being able to quantify and measure every aspect of the consumer, can impact the quality of data, the accuracy of data analytics and the resulting datafied knowledge. As an example, a music industry marketing journal focused on the potential of fake statistics in the digital music industry, arguing that the hype around big data and the "*obsession with numbers and statistics*" provided a feeding ground for such developments and increasingly brought uncertain knowledge into data analytics:

> "*Given the modern world's obsession with numbers and statistics, you can understand why people feel tempted to try such chicanery – and all the more so in an industry that likes to chart most aspects of an artist's performance, from most weekly streams to the number of concert tickets sold.*" [Music:ally Sandbox, Issue 221, p.1]

As these examples have shown, the co-production of datafied knowledge can include uncertainties and subjectivities and is not free from error. As the participants noted, the process of hypothesising and operationalising prior to the start of the big data analysis involved a lot of these uncertainties and subjectivities. Not only because it seemed to be a process that should be done in a collaborative way to know what is relevant, but also because it was not a straightforward standardised process and involved practices such as experimenting with different variables, making assumptions and using intuition, as well as experience. Furthermore, as we have seen, the implementation of the technologies behind the operations, as well as during the analysis, can create obstacles, providing false results and errors, or creating other difficulties, such as the access and sourcing of data. Nonetheless, in many situations, the participants preferred datafied knowledge over traditional knowledge. As we will see in the next sections, this was the result of a constant comparison between the different kinds of knowledge that were in use – in a similar way as we have seen in the previous chapter in the difference between the imagined and the data double consumer conceptualisations – and the corresponding epistemic authority.

6.4. Addressing uncertainties and the epistemic authority of datafied knowledge

There were many aspects the participants brought up when considering how to make data meaningful, and how big data analytics contributed to the knowledge co-production in their organisations. Within this process they considered the resulting datafied knowledge as a better, more accurate and more precise type of knowledge in comparison with more traditional kinds of knowledge. As a result, the perceived superiority of datafied knowledge and the dataist belief of precisely quantifying the consumer led to the expectation of conceptualising the consumer as a data double – despite the uncertainties in the knowledge. In these final sections of the findings chapters, I will explore why datafied knowledge is considered a superior kind of knowledge, resulting in the data double, and how the uncertainties in datafied knowledge are dealt with.

As discussed in my theoretical framework, and further drawing from Jasanoff's (2006b) concept of co-production, epistemic authority does play a significant role in the stabilisation and use of different kinds of knowledge. This revolves around the credibility of that knowledge and the things which influence its authority. The epistemic authority of datafied knowledge is largely based on the expectations of the superiority of that knowledge, which is comparable to José van Dijcks (2014, p. 198) description of the ideology of dataism. Dataism consists of a "*belief*" of individuals in the possibility to objectively quantifying consumers through (online) data, and a "*trust*" in the correct collection, analysis and interpretation of this data. This trust is centred on how datafied knowledge is being communicated, which includes a simplification and translation of that knowledge. In these practices of communicating knowledge, there are many ways in which uncertainties in it can be dealt with. They can be addressed and mitigated but they, likewise, can be disregarded and omitted from the translation and communication of datafied knowledge. As a result, end-users of datafied knowledge, such as marketers, are often not explicitly aware of the potential uncertainties in it.

Furthermore, the "belief" in the objective quantification of consumers stems largely from depicting big data analytics as faster, better, more accurate and more complete in the process of co-producing knowledge, in particular in comparison with more traditional market and consumer research methods. In a similar manner, as the participants compared consumer conceptualisations, between the imagined consumer and the data double, and assessed the advantages of the data double, they were also comparing the different kinds of knowledge that were available in their organisations. This all contributes to stabilising datafied knowledge, making it credible and endowing it with epistemic authority, which tends to obfuscate the uncertainties that emerge in its co-production.

In the next sections, I will start by looking at how uncertainties are dealt with in the first place, and which of these uncertainties are often simply omitted through the simplification and translation process. As a feature of big data analytics, the complicated technological and mathematical processes contribute to these uncertainties being further ignored. In the second part, I will then more specifically address how the expectation of datafied knowledge as a superior kind of knowledge emerges through a constant comparison with traditional kinds of knowledge, based on the dataist ideology. This further promotes the use of datafied knowledge in the participating organisations,

convincing end-users to use it. In the final section, I will focus on the few critical voices among the participants regarding the use of big data analytics in marketing, those who partially addressed the potential uncertainties in datafied knowledge and how this could affect the consumer.

6.4.1. Uncertainties that are lost in translation

The slipperiness of co-producing datafied knowledge that has been described by participants can lead to many uncertainties emerging in it. While the participants reported different attempts to mitigate the uncertainties that emerge in the co-production of knowledge, inaccurate results in datafied knowledge persisted and uncertainties, unreliability and difficulties with knowledge were ignored, in the long run. This can be seen when the participants talked about how datafied knowledge was communicated and disseminated within their organisations. As Jasanoff (2017) has discussed in a recent article, the circulation or communication of knowledge, including datafied knowledge, is based on a simplification of complexity which can be described as a translation process. This simplification and translation of datafied knowledge was explained by the participants as well. Both from the producers' side, such as the data analysts and insight managers, as well as from the end-users' side, such as the marketers, the participants often referred to the necessity of the translation processes to make datafied knowledge manageable. In particular, the role of the insight managers seemed to be key in this translation process as they often served as a liaison between the data analytics teams and the marketing or sales teams. Within this translation, the simplification of the data analytics process is a key instance where the uncertainties and difficulties that might occur during the analysis are omitted.

As an example, the following quotes by a head of digital experience & digital marketing in the travel industry shows how datafied knowledge travelled through their organisation. In this process of stabilising knowledge, at every transmission point an interpretation and translation service contributed to the datafied knowledge being simplified, with uncertainties becoming "lost in translation":

> "*I am not very technical, so it is very hard for me to get my head around these things. […] What does it actually mean, and they even try and draw it down for me. They are like, we are looking to do a massive data warehouse. But I'm like*

yes, but what does that mean, draw me on like a paper! What is that, like here is organization one, here is organization two, data warehouse is about here. That's the kind of level of simplicity that I need. […] So, again that's when the smart data guys tell us what to do and we listen to them because we have no idea." [Key3_travel, Pos. 41-44]

"*So, the technical guys have to dumb it down for me, and then I have to dumb it down even further when I give my advice, […] I suppose it is a constant education piece, on my part and on their part. So there it is that kind of lost in translation function*." [Key3_travel, Pos. 60]

Similarly, this translation and simplification of knowledge can also be observed in other parts of knowledge communication and stabilisation. The participants mentioned that visualisations were an important factor of datafied knowledge, particularly as a means of presenting certain key metrics and key findings in a simple and visually pleasing way. The visualisation functions of big data analytics software – such as Tableau – have the advantage of displaying datafied knowledge in a simplified manner, so it can be processed quickly and without confusion by the recipients. In all this, the simplification processes affect, on the one hand, the amount of knowledge that is being transmitted, as there often is only a focus on the main metrics that are considered important for that specific issue. On the other hand, it affects the uncertainties behind the operations by omitting them from the presentation of datafied knowledge. This process of simplifying datafied knowledge assists its epistemic authority and its stabilisation within organisations while disregarding the difficulties of co-production and uncertainties of the knowledge.

As an example, when I asked a data scientist at an FCE, who was presenting about potential uncertainties and bias in big data analytics, about how he would communicate these issues with the end-users of his datafied knowledge, he replied that he made this assessment based on the situation. At times, he would specifically refer to the statistical accuracy and validity measurements of data analytics to highlight the potential of erroneous results. However, in other instances, he might leave out the possibilities of incorrect findings to simplify the situation and the communication of the knowledge. In this example, the end-user's awareness of potential uncertainties is at the discretion of the analyst.

6.4.2. The epistemic opacity of big data analytics

Particularly in comparison to the traditional research methods for consumer and market insight, such as focus groups, the end-users sometimes fail to understand how datafied knowledge is being co-produced, as the quotes above show well. A lack of mathematical and technical knowledge of the inner workings of big data analytics can result in the operations being perceived as the typical black box, which often leaves no other possibility than trusting the people creating the technology and the insight. In these instances, data analysts have mentioned that explaining the black box is sometimes just too complicated, or that there are simply not enough resources to help elaborate on the workings of data analytics. In other situations, even the technical experts are not fully knowledgeable about the potential uncertainty of the knowledge, as some participants working in big data analytics noted, saying that they did not always understanding the models and algorithms in their entirety.

In some situations, it seems like epistemic opacity, a lack of technical understanding, as well as the translation process of datafied knowledge contribute to situations in which the end-users of such knowledge appear to be entirely dependent on trusting those responsible for creating it – the data analysts, insight managers, managers – when judging its accuracy. In this setting, end-users have no means of checking or knowing whether there are uncertainties in the type of knowledge that they have available. Particularly if the technology behind the co-production of knowledge becomes more complex, end-users are left without much of a choice and must trust its correctness. Again, when addressed, the participants chose to focus instead solely on the advantages of big data and the corresponding knowledge. Being able to access datafied knowledge faster, on a more detailed level, in a more visually appealing manner and more accurately than before is sufficient for participants to disregard any potential inaccuracies. Furthermore, the participants sometimes argued that, despite the potential uncertainty, the results were still more accurate than if people were relying on their intuition or on the subjectivity of traditional knowledge.

6.4.3. Addressing and mitigating uncertainties

It should be noted that uncertainties in datafied knowledge are not only excluded through the simplification and translation of knowledge. The participants also described how, in their organisations, strategies were put in place aimed at dealing with the uncertainties that arose through the co-production of knowledge. As we have seen in previous sections, the cyclicality of co-producing datafied knowledge, as well as the multiple steps of the co-production process, can act as an additional function to address, rectify or improve false or unexpected outcomes. Furthermore, as the participants mentioned, the constant development and improvement of existing big data infrastructures have the aim of further improving the outcomes of the analysis and increasing the accuracy and details of the datafied knowledge. The same approach was relayed in terms of the development of skills in their organisations. These were also aimed at increasing the quality of big data analytics and had the potential to address uncertainties in datafied knowledge. This can, on the one hand, be in the area of big data analytics, where analysts are trained to be able to adapt to the quickly-changing field of big data technologies. On the other hand, as the following quote shows, the participants felt the need for everyone else in their organisations to develop basic mathematical and statistical skills to better work with the datafied knowledge, including skills such as interpreting the results and critical thinking in order to be aware of potential misconceptions in the knowledge, such as it having correlation without causation:

> "*So then, when it comes to the way the results of these data analytics projects actually penetrated the organisation, here I think one of the definite skills that we are requiring is being, not necessarily proficient but being adept as the analytic solutions in general. So even just things like Google Analytics and showing your aptitude as let's say, like people in data analysis in general.*" [I3_travel, Pos. 62]

In other examples, the results of unsatisfactory or uncertain datafied knowledge caused complete shifts in the process of co-producing knowledge. This was, for example, reported by participants whose organisations had been cooperating with external data analytics organisations to co-produce datafied knowledge. The participants referred to ventures of their organisations with external analytical firms which had been difficult and lengthy processes, resulting in unsatisfactory outcomes, such as irrelevant knowledge or

an insufficient support in effectively using the datafied knowledge. Consequently, these organisations performed a shift from relying on external big data analytics firms to an in-house big data analytics process. Despite the difficulties that can emerge from this switch to internal big data analytics, the participants noted that the results were considered to be better, and the process of the co-production of datafied knowledge to be easier.

Overall, looking at how the participants referred to how they or others in their organisations dealt with the uncertainties in datafied knowledge, there appeared to be a tendency, while being addressed at their root with the aim of mitigation, for them not to be communicated to the end-users. As a result, they were no longer considered when the datafied knowledge was being used in marketing, which contributed to an epistemic authority and credibility of the knowledge being perceived as an accurate representation of reality. In the practice of conceptualising consumers, this contributes to the creation of a data double that is considered to be a virtual copy of the individual consumer.

6.5. Establishing epistemic authority of datafied knowledge

The stabilisation of datafied knowledge and its credibility and epistemic authority is not only established through the simplification and translation of knowledge. Looking further at how the participants described datafied knowledge in general, the high credibility of this also stems from a constant comparison with traditional knowledge, in which the former is portrayed as more reliable, more accurate and superior to the latter. Traditional knowledge about consumers, although having a longstanding history and presence in marketing, was portrayed by the participants as increasingly outdated, whereas datafied consumer-oriented knowledge was seen as an improved type of knowledge that would – ideally – lead to exact representations of consumers. The comparison and assessment by participants of the two types of knowledge were based on multiple criteria that largely revolved around the 'quality' of the knowledge, which ultimately contributed to its credibility. Mainly, they described datafied knowledge as faster and less resource-intensive in its production and dissemination, better and more accurate in its results, and more complete in what it was able to represent. Lastly, in particular at FCEs but also during the interviews, participants often referred to the potential increase of economic value through big data analytics, which further contributed to the perception of the superiority of datafied knowledge, and its epistemic authority.

6.5.1. The epistemic authority of datafied knowledge: faster, better, more accurate and more complete

The perception of datafied knowledge as being faster, more complete, and more accurate was an important and recurring topic at the FCEs which I attended for the research. At these conferences, big data analytics was almost exclusively talked about in a very positive way – mainly focusing on how organisations could profit from data – and as something that was highly reliable and imperative for any organisation. Big data analytics was portrayed as a technology in which organisations needed to invest and which would dominate all business-related operations in the present and certainly in the future. Data was seen as the key component for innovation, which also meant that everyone who does not rely on data is seen as losing out against the competition. The FCEs were promoting a hyped idea of data and big data analytics, which was supported by the many big technology companies that were represented and were often aimed at promoting and selling their own big data analytics products. This hyped idea of big data analytics is what can also be observed in the discourses of the interviewees when they are talking about its advantages.

Datafied knowledge is perceived as being faster regarding the operations for obtaining and disseminating the knowledge. While the participants portrayed traditional market and consumer research methods as requiring a lot of manpower, money and time in order to obtain the necessary knowledge, big data analytics was seen as more efficient in this regard. The participants thought datafied knowledge was produced more quickly and without too much effort However, they disregarded the lengthy and expensive processes that are involved in making big data analytics functional in the first place. But, the general perception amongst the participants was that, by obtaining knowledge quicker through big data analytics, other operations increased in speed as well. This included how quickly decisions could be made; communications, briefs or reports could be sent out; and how consumers could be conceptualised, or these conceptualisations adapted. They perceived datafied knowledge as increasing efficiency within their organisations, in which their pre-data analytics organisation was seen as slow and inefficient and which was being improved using big data analytics. For example, a data analyst in the travel industry explained how the implementation of big data analytics and, particularly, the use of the

data science and analytics software *Alteryx* had significantly improved the speed of certain operations, such as data processing:

> *"So, Alteryx has basically changed our world, and the amount of time it takes to process data has, has been cut. For example, there is a, we work on four-week periods. So, we don't work on months, we work, we have 13 4-week periods. And after every period we have to do a period end-processing. And there are few tasks that took three days, processing using Excel and Access. And using Alteryx that, we cut that down to probably two or three hours. [I: Ohhhh] And you think that's so much time that you just freed up [I: Yeah, yeah, sure!] Just overnight!"* [L1_travel, Pos. 11]

The other differentiation participants made between the two types of knowledge was the reflection on the accuracy or, in other words, on how well knowledge was able to represent reality. In this, the pre-data analytics organisation was considered to be reliant on unspecified knowledge about consumers in the sense that traditional knowledge was quite broad and based on traditional market and consumer research methods that largely lead to imprecise knowledge and subjective assumptions about the consumers. As a result, the participants considered that in marketing, they could only establish wider demographic and behavioural consumer categories on which decisions had to be made and consumer conceptualisations created. With the emergence of big data analytics, they believed that their organisations could shift away from these imprecise and assumption-based decisions and, instead, rely on data-driven decision-making, which was perceived as more precise and allowed for more fine-grained consumer insight. Since datafied knowledge was seen as being more accurate, this also meant that more precise consumer conceptualisations were expected to be created, which would be able to represent the real individualised consumer. The participants argued that instead of having to rely on subjective knowledge, the implementation of data analytics within their organisation enabled the possibility of relying on "exact facts".

This is also a common rationale presented by the participants for the use of KPIs and measurements. Numbers are perceived to be more reliable, accurate and retraceable than qualitative assessments about the consumers. Therefore, KPIs are considered to be a way of bringing more objectivity into consumer-oriented knowledge available in their

organisations. It, furthermore, is perceived as an opportunity to trace financial investment in organisational operations and its potential return more easily. Through big data analytics, knowledge in the form of numerical measurements can also serve as a better tool to argue for an investment, e.g. in marketing strategies, than through other forms of knowledge, such as experience or focus groups.

An example of the comparison between the accuracy of the two types of knowledge can be seen in the following quote. In this case, the interviewee specifically stated that asking consumers directly about their behaviour would not provide the desired results as, often, consumers do not know or realise how they behave. As such, using big data analytics to uncover consumer behaviour will provide more interesting, as well as more accurate, results:

> K5: *"One of the, one of the issues with using any method which asks people what they think is that people don't always think they behave the way that they actually do behave."*
>
> I: *"Yeah."*
>
> K5: *"Anyway that...* [short pause] *Anyway, that you can kind of* [short pause] *get a picture of people's behaviour, which isn't asking them in a direct way, is always going to be a bit more interesting than when you ask people explicitly*." [Key5_ consulting, Pos. 44-46]

The possibility or expectation of producing more exact and more accurate consumer-oriented knowledge through data analytics is thus seen as one of the big advantages for the process in marketing. As datafied knowledge about consumers is thought to enable better and more informed decisions, it is seen as being able to make more of an impact on consumers. The following quote from an ad-operations and analytics manager in the entertainment industry sums up this perceived difference between traditional and datafied knowledge, in which the first is seen as an indecisive and broad approach, whereas big data analytics brings in the necessary precision. This allows us to know who the target consumers are and allows for a more individualised approach when communicating with them:

> *"So, previously, we quite often had a shotgun approach, kind of we have this and that of a budget, let's do something, somehow. And now our goal is to exactly and specifically target our groups, which we know, we partially know really well and precisely, we have our analytics from Facebook and Google and our insights from the market research team."* [M3_entertainment, Pos. 28 – translated from German]

Another quality assessment participants tended to make when comparing different kinds of knowledge was based on the completeness of knowledge. Concerning knowledge about consumers, a lack of available information about the consumers was mentioned and was subsequently perceived as complicating their conceptualisations. For example, the participants expressed how the traditional consumer segmentation profiles, the imagined consumers, were not ideal for marketing strategies as they were an incomplete representation of the (potential) consumers of their organisations. The assumption was that, by relying on traditional knowledge, they would not have the ability to represent the consumer in its entirety as some information, some parts, would always be missing. Conversely, datafied knowledge was expected to be a means of completing the knowledge about their consumers and of providing a fuller picture of who they exactly were. Since their operations were increasingly moving into the digital world and their consumers were potentially leaving more and more data on their online journeys, the participants considered that it was easier for them to collect or access this data and be able to rely on a more complete knowledge, as the following quote shows:

> "*The other really important part is around the data that you gather. So, a real way of differentiating is by making sure that you are collecting a lot of data, and making sure that you are doing that in a responsible way, and in a way that gives you all of the information that you need to then go and lever it later. So, for example, at (company), one thing that we have got a lot of, we have got a lot of data around the trends in travel, and the behaviours people exhibit before they travel and after they travel and all that sort of stuff. So, we can use that information for [inaudible], because we have got years and years of very rich data sets.*" [I1_travel, Pos. 14-15]

6.5.2. The epistemic authority of datafied knowledge: The value of data

Another factor that was regularly addressed by the participants and which was also a topic at the FCEs is the economic rationale behind the knowledge co-production process, which further influences the epistemic authority and stabilisation of datafied knowledge. In the way the participants described the processes of knowledge co-production and the conceptualisation of consumers, economic rationales are likely to dominate other rationales and are often the prevailing discourse – as they are in the area of knowledge communication and usage. Regularly, as participants were comparing the different types of knowledge, the advantages of datafied knowledge were seen as creating or increasing economic value in their organisations. As the economic functioning is an important – if not the most important - feature of organisational operations, it has the potential of having a significant effect on the epistemic authority, the stabilisation and the credibility of datafied knowledge.

This can be observed at an early stage. When talking more generally about data, the participants perceived it as an economic commodity that could be bought and sold. In this context, the correct processing of data was necessary to lever additional financial gains from it, instead of leaving it in its "raw" state. Similar economic reasoning was made in relation to the improvement of and innovation in their organisations. The participants related how investment in data should have an ROI that can either be the result of an increase in sales – due to making more informed decisions – or due to a reduction of resources expended, such as time or manpower. At the FCEs, the value of data was also sometimes extrapolated to entire local and national economies with claims that these could benefit from a more effective use of data. As Scotland's First Minister Nicola Sturgeon explained in her address to the Data Summit 2019:

> "*We're doing all of that because we understand the obvious and huge economic opportunity that there is. Studies suggest that across the economy as a whole, businesses could soon be benefiting from £3.8 billion per year if they use data more effectively*." [First Minister Nicola Sturgeon, Data Summit 2019]

The economic value of data was also a dominant discourse in the interviews. As discussed, data, as in using improved information that is more accurate, more precise and more detailed, is seen to result in better knowing the consumer. As a consequence, the

participants regarded consumers to be better targeted, more specifically and personally, with the products and services they needed, offered prices they would be willing to pay, as well as receive advertising refined in such a way as to improve communication with them. Ideally, the consumer would subsequently spend more money on products or services of that organisation. Another potential means of gaining economic value through data is by the increase in efficiency it offers, which relates, for example, to adjusting resources in marketing strategies. If marketers know that a certain amount of people will not read or respond to e-mail marketing – uncovered through big data analysis – they can exclude them from the communication list and, in this way, save on financial resources. The potential for increasing economic value through data is largely attributed to gaining datafied insight and knowledge, which has the potential of being used in marketing, sales, human resources, and for internal, as well as external, insight.

6.5.3. Traditional knowledge vs datafied knowledge – finding the right mix

In the comparison between traditional and datafied knowledge, the participants tried to convey a more nuanced and reflective view in some instances. Mainly when they were talking about the application of the knowledge, they would try to highlight the advantages of either type. As an example, while in most instances datafied knowledge was portrayed as being the better kind of knowledge, in the case of testing or experimenting products or marketing strategies, the participants considered traditional consumer research approaches as an equally useful, if not better, approach. They often argued that, in the ideal case, their organisations relied on both traditional consumer research methods as well as big data analytics, as the combination of both provided a more holistic kind of knowledge. Certain organisations, such as those in the entertainment and the travel industry, used both traditional approaches and big data analytics in their consumer and market research. Furthermore, the participants suggested that just trying things out and relying on tacit knowledge could be equally important and helpful as not everything needed to be meticulously calculated and assessed before a decision had to be made. The following quote gives a good example of such a situation, where the interviewee describes how decisions did not always have to rely on big data analytics. Sometimes it was acceptable to rely merely on experience and expertise or by just trying things out:

> *"I'd say another significant one is, you know, I've spoken to the marketing director as well as the head of the wider department and we've agreed that you know, there is too much actual focus on data on a granular level. Like, so frequently a marketer would come to us and be like, oh we're interested in, you know, testing this kind of approach in the US. You know, what evidence do we have that that would work and we're not talking about big sums of money and you know. It's just kind of like well try it, you know, we don't need to, you don't need to have every last piece of evidence before you just go and experiment, like your you should use also your own experience and expertise in your field to sometimes just make some calculated judgments."* [Key2_publishing, Pos. 27]

Still, while most participants initially considered that each kind of knowledge had its advantages and disadvantages, datafied knowledge was regularly portrayed as the preferred and dominant type of knowledge in their organisations. This could be, on the one hand, attributed to the limited resources in their organisations. Often, the ideal case of a mixed-methods approach in consumer research is described as too expensive and time-costly, which is why organisations would rather rely on the – perceived – faster and more efficient big data analytics. On the other hand, the dataism of big data analytics producing a more precise and complete type of knowledge leads to datafied knowledge, in its totality, being perceived as more credible. This is then translated into the conceptualisation of consumers, where big data analytics and datafied knowledge are expected to produce an accurate and complete representation of the real consumer, while traditional knowledge is perceived as inapt in providing the necessary knowledge about who the consumers are and how they behave.

6.6. Persuasion as a mean of endowing epistemic authority

Finally, another key component in endowing datafied knowledge with credibility, increasing its epistemic authority and diminishing attention to uncertainties in it can be attributed to a process of convincing, which mainly happens during the implementation of big data technologies. Big data analytics and datafied knowledge are not always accepted by everyone in participating organisations. At times, their acceptance requires negotiation and the convincing of others within these organisations – in particular, some end-users of the datafied knowledge. The following quote shows how such a process of

convincing others can play out. The example stems from an organisation in the travel industry. Here, the initial project sponsor – in the form of a senior manager and characterised as a visionary by the interviewee – had to campaign for a year before other senior managers were convinced enough to invest in big data analytics:

> "*But, (..) I suppose it took a really big visionary to bring all the organisations together. […], who at the time, three years ago now, was the director of marketing and digital of [company]. And he saw a good potential of bringing all the agents… all the agencies together to tell a unified story. So, he campaigned for that to happen for about a year, before people finally listened. And as you can imagine, you know organisations are used to work in their own way, so to be able to bring people together, takes a lot of effort and willpower. So then, you know, slowly people, senior management from the different organisations bought into this idea. And agreed to put budget and resource against it, and that was the first step*." [Key3_travel, Pos. 40]

Strategies for convincing regularly rely on the positive aspects of big data analytics – as being faster, better and more accurate – and the dataist belief of the ability to objectively quantify and measure consumers, or, as in the earlier example, streamline procedures and enabling greater market efficiency. The acceptance of big data analytics frequently requires these convincing strategies be applied. The participants noted how certain procedures of convincing were necessary for big data technologies to be accepted and put into operation throughout their organisations. They have described the convincing process as a form of education, in which the positive aspects and advantages of big data analytics are demonstrated to other people and teams in their organisations. The aim of convincing others is generally that the end-users of datafied knowledge will realise that they can benefit from big data analytics due to business operations being improved. However, the rationale used for convincing others is mainly by arguing the perceived positive aspects of big data analytics, these being speed and the production of more measurable, accurate and complete knowledge. The following quote from the publishing industry further exemplifies this process of steadily and practically showing the advantages and possibilities of big data technologies. Here, the interviewee highlights the importance of emphasising the practical impact big data analytics will have on the respective teams and their daily operations:

> "*There's always going to be people who need IT and it's not like you have to convince people that they need IT help but you do need to, you do need to convince people that you need analytics and that you do need consulting services, for example. And so, I think that it makes sense that for a lot of organizations there is that… that first phase is really kind of proving your utility and when you're doing that there's a lot more of you know. This is sort of obsequiousness that you know is just kind of a fundamental factor where it's, if you're trying to prove your relevance, you can't say: oh, excuse me I actually don't think that project has as much impact is as I'd like. You're going, yes, let me help you with that! Please, let me help you! Let me show you what we can do and it's a little bit an education piece*." [Key2_publishing, Pos. 20]

The process of convincing people to use big data analytics appears to not only promote datafied knowledge but, at the same time, demote others forms of knowledge. Being a common theme of the findings in general, the advantages of big data analytics are often compared to traditional or previous forms of knowledge production. As a result, the latter lose importance due to being portrayed as slower in the process of creation, as well as communication as with knowledge produced through qualitative methods, or not as streamlined in their dissemination such as with knowledge resulting from experience, which ultimately has the potential of changing how knowledge is perceived and stabilised. In the convincing process, as well as regarding datafied knowledge in general, many of the participants not only portrayed data as the primary source of information but also relegated all other forms of knowledge to an inferior level – to be considered as a supplement, if necessary.

This development can be exemplified by the language that is used and which can give an insight into the dataist ideology of the participants, mainly when they were talking about being able to uncover hard facts, the absolute truth, or the real customer, etc. Re-using an example from above, the following quote describes the dataist ideology in relation to big data analytics. The interviewee talks about the main difference between big data analytics and qualitative research methods, in which she implies that when asking people directly about certain things – such as their behaviour – they will not provide an accurate assessment of it. In her view, being able to uncover consumer behaviour through other

indirect means, such as big data analytics, will provide more interesting, and better, results:

> "*One of the, one of the issues with using any method which asks people what they think is that people don't always think they behave the way that they actually do behave.* [I: Okay] *Anyway that... [long pause] Anyway, that you can kind of get a picture of people's behaviour, which isn't asking them in a direct way, is always going to be a bit more interesting than when you ask people explicitly*." [Key5_consulting, Pos. 44-46]

6.7. Critical reflections on datafied knowledge

As a final point, the participants addressed the potential of inaccurate and uncertain knowledge in some instances, often in relation to how it might impact the individual consumers at the receiving end of the marketing operations. They saw that organisations had a great responsibility to ensure that the information that was being used for acting on consumers in such a potentially impactful way was accurate and without bias. They were particularly referring to known cases of misguided or wrong decisions based on inaccurate or biased datafied knowledge, such as the algorithmic decision-making in the US court system concerning potential re-offenders, where bias in the data led to faulty automated decisions.[6] A few participants were aware of the potentially harmful impact of decisions that had been made relying solely on the postulation of consumer conceptualisations. As the following example shows, concerning the hype around big data analytics, the interviewee stated how organisations risked relying ever more on datafied knowledge, despite potentially not needing to and not even knowing how to handle it correctly:

> K8: "*So, they think big data is just enhancing their portfolio anyway because they use… Just use it, not how they use it and why they use it. And sometimes, I think, like, it could be problematic that, probably… Using data, sometimes, can lead to worse decisions than not using, if you're using it wrong. You know what that means? Like, it's…*"

[6] An excellent critical analysis of such systems of governing crime and policing can be found in Sanders and Sheptycki (2017).

I: "*You're making decisions on inaccurate…*"

K8: "*Yes, just using it, and using it with half knowledge, usually, is probably going to… Lead to even worse results than not using it. So, that's the more pessimistic side of it. Because it's a hype. Everyone knows it's big data, and we need to do it. We need to have machine learning and AI and all that kind of stuff. And they hire data scientists. And then they're sitting there, it's like, okay, what are we going to do? They don't have a clear business sense why they're using it. So, it's like, we just need to use it to get further, but they don't know why they use it. So, that's a really… That could lead to, I would say, devastating outcome.*" [Key8_consultant, Pos. 144-149]

However, in general, the awareness that such problems might arise was only rarely mentioned by the participants. Interestingly, looking at the topics at the FCEs, a change in perception may be taking place. Some of the more recent FCEs that were visited in 2019 covered the potential impact of the use of algorithmic decision-making on individuals and consumers, often with a strong focus on the risk of using potentially inaccurate and biased datafied knowledge. For example, since datafied knowledge can become outdated relatively quickly, great care needs to be taken if automated decisions are made based on such knowledge. This can become problematic in organisations where such knowledge is being used – be it in the recruiting process or marketing. Decisions can be made on false rationales, potentially discriminating individuals or groups and leading to an exaggeration of these effects due to self-learning algorithms:

"*In particular, the currency of data may change over time – rendering it inaccurate or irrelevant.* [The] u*se of AI systems can exaggerate these effects, where the algorithm continually learns from and reuses the data. […] Discrimination can result from inaccurate data that makes assumptions about trends and behaviours to group individuals. […] Issues can be unfairly applied to individuals where companies can choose their own variables and labels. For example, defining 'good employees' and 'productive workers.'*" [Intelligent Automation 2019, Pos. 131-139]

The FCEs, furthermore, have addressed the potential wider impact of big data analytics and automated decision-making on individuals and society. Here again, popular cases and

scandals, such as Cambridge Analytica's role in interfering with elections in the US and the Brexit referendum, have often been used to show how, on a large scale, major issues might emerge. Generally, the responsibility here is often shared between individual organisations and the regulators. The duties of organisations include them needing to handle the data of individuals in an ethical and sensible manner. This was the message that was directed at the attendees of the FCEs. They are seen as needing to provide a strict regulatory framework that will, on the one hand, limit the misuse of data while, on the other hand, still enable their organisations the freedom to use big data analytics and allow them to gain an economic advantage from it.

It seems clear that there can be a potential impact on individuals, groups and even entire societies through big data analytics and decision-making based on postulated consumers in organisations – of which some participants were at least partially aware. The question that remains is whether this perception will be further disseminated in the future, and whether this awareness will potentially change how datafied knowledge is used in marketing – and elsewhere. Because, as this final quote exemplifies, while the participants were mindful of the potential power lying in big data analytics, allowing the manipulation of consumer or individual behaviour, and thus impacting large groups and societies, this awareness might not change how this process is used, but rather make it even more opaque and hidden:

> "*Yeah. I think it's a little bit scary to be honest, actually. When you can do this kind of thing at scale. And it has such big… Like it doesn't have a big impact if you just do it on, to ten people but when you've got the scale, when you've got a company which has the scale to be doing it to billions of people. Even though they're retaining the freedom to choose what they want, if there is, if there is a deliberate choice architecture that's been set up. Yeah… Because we know that it nudges people to spend more money or something like that. Then it can be, I think it can be a little bit scary. And for that reason, I think you know for some of our clients aren't happy with us talking publicly about what we do with them.*" [Key5_consultant, Pos. 50-51]

6.8. Conclusion – the epistemic authority of the data double and the postulated consumer

To conclude, as I have elaborated upon in the two findings chapters, the participants provided lengthy accounts of how big data analytics influence the conceptualisations of consumers through the changes that occur in the knowledge co-production and the stabilisation and epistemic authority of that datafied knowledge. With the use of datafied knowledge, consumers are increasingly conceptualised through data doubles, creating the expectation of exact representations of their real counterparts. This expectation is the result of a dataist belief amongst the participants, that of trusting the ability to quantify the consumer in a more accurate, detailed and complete manner – thus reassembling the consumer as a virtual data double.

The dataist ideology and the epistemic authority and credibility of datafied knowledge seem to involve a constant comparison and assessment with other types of knowledge, such as more qualitatively gained knowledge or tacit forms of knowledge. A comparison in which the positive aspects of data analytics were highlighted by the participants. This hype surrounding big data analytics results in an increased push for big data technologies and the convincing of organisations to choose these technologies over others. A common rationale of the hype for data analytics revolves around the economic potential of data, which further increases the apparent necessity for big data technologies. All this bestows datafied knowledge with increased epistemic authority and further confers this knowledge with the credibility of being the exact truth.

While participants regularly referred to a mixed methods approach as the ideal case scenario, their main focus was on big data analytics. This may have partially been due to the limited resources of their organisations and the time constraints involved in co-producing knowledge, in general. Particularly regarding marketing and the necessity to profile and predict consumers, personalise offers and advertising and target consumers accordingly, datafied knowledge was perceived by the participants as being the inevitable choice of knowledge, vastly superior to traditional types of knowledge. As a result, the data double was perceived to be vastly superior to the imagined consumer conceptualisation. Being more precise, less reliant on subjective interpretations – of

marketers as well as consumers – and more actionable, the data double was perceived by the participants as enabling necessary and innovative marketing measures.

However, as I have explained – based on the descriptions of the participants – despite the expectation that datafied knowledge will be accurate and an exact representation of the consumers, the practical co-production of datafied knowledge has been shown to be complex, slippery and cyclical. It is not a straightforward, standardised process but, instead, involves negotiation amongst different teams and individuals, requiring a good understanding of the organisation, market, and consumers, as well as a range of technical and statistical skills. Big data technologies enable, but can also complicate, an analysis. And here, ideologies influence the process of creating datafied knowledge. Furthermore, as the participants stated, datafied knowledge has the potential of being imprecise, can include uncertainties, and can produce false results. While these are not always translated and communicated further to the end-users, these imprecisions and uncertainties still persist. As such, they have the potential of being included in the conceptualisations of the consumers.

Despite these problems, the emerging datafied consumer conceptualisations are often perceived as exact copies of their real counterpart, even though they are unable to represent the consumer as an exact data double. Decisions are made based on these conceptualisations as if the profiles were the real consumer. This is rarely questioned, or at least not to the extent that traditional consumer conceptualisations are. The ideal type of the exact consumer profile thus becomes a postulation, as it is assumed to be true, and it is used as a basis for action, without proof of its accuracy and rigour due to its epistemic authority. This results in the generalised credibility of the exactness of the knowledge. The postulated consumer conceptualisations then serve to anticipate consumer behaviour and target consumers, accordingly. This can have ramifications for the actual consumers as these conceptualisations are used to initiate behavioural change, steer consumers in certain directions, or decide on what offers consumers are able to access or what prices they should pay. The concept of the data double as it was defined by Haggerty and Ericson (2000) 20 years ago, separated from the human body and reassembled in the virtual, still does hold true, with the difference being that postulated consumer conceptualisations are generated by continuous amalgamations of data, derived from different sources, of different consumers that only tend to work in relation with each other and are stripped of

any real connection with actual individuals. Instead of being data doubles, the postulated consumer conceptualisations are approximations with uncertainties, inaccuracies and faults. Nonetheless, they are still seen as accurate data doubles by marketers and, because of this, have a real and substantial impact on individual consumers.

7. Discussion of results, limitations, future research, and conclusion

7.1. Introduction

In this chapter, I will discuss the findings presented in Chapters 5 and 6. First, I will consider how the research I have conducted has answered the research question, which is followed by a discussion on how my key findings relate to the current literature. Here, I will elaborate on how my findings provide empirical evidence in response to the issues that have been discussed in the Critical Marketing, Critical Data, and Surveillance Studies literature, as well as how my theoretical approach contributes to the research of consumer-oriented knowledge and marketing. This will be followed by practitioner recommendations. Finally, I will conclude the discussion chapter by reflecting on the limitations of the study, as well as present an agenda for future research.

7.2. Revisiting the research question

This thesis began by establishing that big data analytics has proliferated in every domain, including the governmental sector, in healthcare, as well as in education. In the private sector, big data analytics is used widely, such as to optimise supply chain management, in industrial and agricultural processes, to assist hiring practices in human resources, or, in marketing, to better profile consumers (Kitchin, 2014). Academic literature and empirical research on the latter have substantially increased over recent years. In the mainstream marketing and big data literature, the potential of big data analytics for consumer research and marketing has been discussed extensively, focusing on how big data analytics improves the knowledge organisations have about consumers and other market actors and, thus, marketing practices. One of the main arguments in this literature is that improved datafied consumer-oriented knowledge leads to better profiling of consumers – in the form of segments or individual profiles – as well as prediction of their future behaviour, the personalisation of offers and advertising and more precise targeting (Ascarza et al., 2018; Bello-Orgaz et al., 2016; Matz and Netzer, 2017; Trusov et al., 2016).

The mainstream marketing and big data literature, however, only addresses the potential of big data analytics and the expectations marketers have and deals less with the impact

of big data analytics on marketing in practice. Exceptions can be found in the field of operations research, where big data analytics is discussed in relation to its impact on organisational structures and the limitations of the technology, addressed mainly through a managerial lens (Brock and Khan, 2017; Cabrera-Sánchez and Villarejo-Ramos, 2019; Khan and Brock, 2017). However, little has been written concerning the practical use of big data analytics to produce consumer-oriented knowledge, as well as the practical use of that datafied knowledge to profile or conceptualise consumers.

A more critical approach towards big data analytics in marketing can be found in the literature of Surveillance Studies, Critical Data Studies, and Critical Marketing Studies. In these areas, a more thorough and critical analysis of the use of big data analytics for consumer conceptualisations is provided. Critical Data Studies, for example, criticises how big data analytics is often perceived through the ideology of dataism, which stands for the idealised and mythical perception that big data analytics creates factual knowledge, in general as well as in marketing (van Dijck, 2014). Applied to the conceptualisation of consumers, this leads to an expectation of marketers being able to produce an exact data double consumer conceptualisation through big data analytics (Haggerty and Ericson, 2000). However, in my thesis, I argue that when both big data analytics and marketing are seen as sociotechnical assemblages, the practice behind the production of datafied consumer-oriented knowledge and consumer profiles are neither neutral nor objective, but subjective and performative (Araujo et al., 2010; Callon, 2006; van Dijck, 2014). The conceptualisation of the consumer through datafied knowledge can impact and shape the real consumer through the everyday practices of data analytics, marketing and consumption. This has led to the question of *how* and *why* big data analytics and digital marketing leads to different consumer conceptualisations:

> How and why do marketers conceptualise consumers differently when using big data analytics in comparison with traditional market and consumer research methods?

Unpacking the practices of big data analytics and conceptualising consumers, and engaging with the normativities that are part of these practices, helps us to understand how consumers are enacted and performed through these conceptualisations (Jasanoff, 2017). Using the idiom of co-production (Jasanoff, 2006b, 2006c), I have specifically

examined the practice of analysing big data for marketing purposes. In this, social, material, normative and cognitive processes shape the co-production of datafied knowledge in the organisations and the conceptualisation of their consumers. Addressing my research question, the findings show that in the participating organisations, consumers are conceptualised in three different ways, depending on whether the marketers relied on traditional market and consumer research methods or on big data analytics:

1. First, there is a stark difference between the contribution of traditional consumer research methods and big data analytics towards consumer conceptualisations. With the first being perceived as slow, inefficient, inaccurate and subjective, the resulting consumer conceptualisations are described as not being actionable and representative of an imagined group of consumers. Hence, the outcomes are imagined consumer conceptualisations, in which only certain characteristics can be included in the conceptualisation and, even then, it is unknown whether the real consumers that are targeted based on these conceptualisations relate to these characteristics.
2. Big data analytics, on the other hand, is perceived as an improvement as it enables the creation of precise and accurate consumer conceptualisations. These conceptualisations are not only more detailed in terms of the characteristics that can be included, but closer to representing real and individual consumers. Marketers expect to rely on an exact knowledge about consumers, with the relating consumer conceptualisations that are created to be virtual data doubles. As many of the participants saw big data analytics in a dataist way, believing in the factuality and accuracy of the datafied knowledge, the expectation of producing a data double consumer conceptualisation was omnipresent.
3. However, since the produced datafied knowledge cannot be an exact knowledge, but is always only an approximation, which also entails uncertainties and errors, the same holds true for the consumer conceptualisation. The virtual data double, as a factual representation of the real consumer is not feasible. As the consumer conceptualisation is treated as if it were a data double, the conceptualisation becomes a postulated consumer conceptualisation, expected to be factual without any real evidence of its existence.

To answer the 'why' part of the research question, the emergence of these three different conceptualisations – the imagined consumer, the data double consumer and the postulated consumer – can be explained through how the respective consumer-oriented knowledge is being produced. Showing how datafied knowledge is co-produced and deeply entangled in social, normative, technical and cognitive procedures explains how datafied knowledge also entails uncertainties and inaccuracies, and emphasises the inability to produce exact knowledge about consumers. Furthermore, looking at the procedures of stabilising knowledge (Jasanoff, 2006b), it becomes apparent that the dataist ideology significantly contributes to establishing datafied knowledge as the main knowledge amongst the participants. Here, datafied knowledge is portrayed as faster in its creation and communication, more accurate in its representation of reality, more detailed in terms of the characteristics that are covered and, overall, judged to be a better kind of knowledge. More importantly, however, the uncertainties that can arise in this knowledge are omitted from these descriptions, and also largely ignored in the procedure of stabilisation.

Datafied knowledge is endowed with epistemic authority and made credible to the extent where end-users, such as marketers, have no reason to question either its credibility or its accuracy or factuality. The epistemic opacity of the knowledge further contributes to this status. As it is almost impossible for marketers to check how datafied knowledge is produced due to the technical and statistical characteristics of big data analytics, they have no other choice than accept and trust the knowledge they receive. As a result, the imagined consumer conceptualisation is negated as the contributing knowledge is perceived as inaccurate and the datafied knowledge has been made more credible and has more authority. However, as the uncertainties of this knowledge are rarely addressed, and not communicated further when it is stabilised, the postulated consumer conceptualisation is all that remains.

7.3. Implications

7.3.1. Theoretical implications

As set out in Chapter 3, the idiom of co-production has been used to research the production of knowledge about consumers in marketing. There are multiple advantages to using this framework. The first is that it helps to show how deeply entangled the co-production of consumer-oriented knowledge in marketing is – as it is in other settings as well. Similar to other studies in the cultural materialist strand of Critical Marketing which have researched how marketing is done in practice, my findings reveal the many different influence on these practices. As in Grandclément and Gaglio's (2011) research on the use of focus group in marketing, as well as Sunderland and Denny's (2011) study on the practices of segmenting consumers, my research shows how the production of knowledge about consumers is a process that involves sociality, normativity and technologies, irrespective of its methods of production.

Understanding that big data analytics involves a constant negotiation between those commissioning or requiring the knowledge, those responsible for the analysis, as well as other, more senior individuals in the organisation reveals how the resulting datafied consumer-oriented knowledge is not necessarily an exact representation of reality. Instead, it is the outcome of a compromise between what is needed, what is feasible – in terms of technological capabilities, as well as financial and temporal resources – and what has been done before. As the participants have revealed through their broad descriptions on how big data analytics is carried out in their organisations, the practice of analytics includes what Jasanoff (2017, 11f.) simply terms: the involvement of politics. This can be seen in the framing, in terms of defining the problem that is at the basis of the analysis. But this can also be seen through the constant selection of which variables to include in the analysis, which hypothesis will be prioritised, or even which big data analytics projects to initiate first. Finally, the politics of non-knowledge can be seen through the large number of uncertainties and errors the participants referred to, and how these are (not) handled. Certainly, in this regard, big data analysis runs the risk of over-relying on what is already known and not considering enough what is not (Jasanoff, 2017).

The idiom of co-production and its necessity to put the production of knowledge into its respective contexts shows that this processes is rarely a linear one (Jasanoff, 2006a), but

is cyclical, as I have been able to show. The endpoint of big data analytics is rarely reached but, instead, it raises additional questions, requiring follow-up analysis and – ideally – the creation of a constant process of improvement and new knowledge. Dubuisson-Quellier (2010) has discussed similar issues in relation to what she has called the qualification of products and consumers. This relates to the procedures in marketing and product development which conceptualises consumers and connects certain products to these consumers. Dubuisson-Quellier not only shows that these procedures are heavily dependent on normative, technological and social contexts, but also that these qualifications are not done in a one-off manner: there is no linearity in this practice (Jasanoff, 2006a). Instead, these conceptualisations "must be re-specified regularly" (Dubuisson-Quellier, 2010, p. 92) as market researchers and product-development experts are constantly producing new consumer-oriented knowledge and are constantly re-defining the consumers and the connected products.

The main advantage of using the idiom of co-production to research the use of big data analytics for consumer conceptualisations lies in its focus on the stabilisation of knowledge. This is a factor that is regularly overlooked in the existing research (Jasanoff, 2006b) but is a crucial aspect of the co-production of knowledge. Unpacking the stabilisation of knowledge reveals the powers, politics and practices that are involved in translating and communicating different kinds of knowledge. As shown in the previous chapters, during the practice of co-producing datafied knowledge about consumers, as well as the communication of that knowledge, the participants referred to the instances of simplification. In most cases, the simplification relates to the epistemology of datafied knowledge. The procedures that are at the core of co-producing datafied consumer-oriented knowledge are rarely disclosed to the end-users of that knowledge, such as the marketers or managerial staff. Instead, datafied knowledge is interpreted, specific findings are selected and then prepared for presentation and communication, very often through the new means of visualisation that go along with big data analytics (Kennedy et al., 2016; Kennedy and Hill, 2017). The participants, particularly those who were end-users of datafied knowledge, referred to the black box of big data analytics, in which the technology and necessary analytics skills were seen as complicated and as obfuscating the understanding of underlying analytical procedures. As a result, big data analytics was described as epistemically opaque by the participants.

As I have shown, this is a main contributing factor of big data analytics which leads to a kind of knowledge that is perceived as objective and factual. While datafied knowledge is stabilised, made credible and holds epistemic authority, the potential irregularities, uncertainties and subjectivities are lost through the simplification process, and not translated into the end-result. This is even more crucial as the new kind of datafied knowledge is in direct competition with more traditional types of consumer-oriented knowledge. While the first is increasingly stabilised, the latter is rendered unstable. As Jasanoff (2006a, p. 280) has mentioned, "co-productionist analysis is symmetrically concerned with both stability and instability", and both are an important part of the changing environment in which consumer-oriented knowledge is co-produced.

The strength of the co-productionist analysis is that it reveals the normativity and philosophical understanding of the different types of knowledge, which contributes to the stability and instability of these types of knowledge. As big data analytics has its strengths in highly quantified situations, a predominant realist ontology and (digital) positivist worldview is conveyed upon the use of such methods, as I also discussed in the beginning of my thesis (see also Fuchs, 2017). In its most radical manifestation, this worldview becomes ideological and thus becomes what van Dijck (2014) has termed dataism. The field of marketing was already accessible for a more positivist paradigm, with marketing-as-management largely relying on positivist principles, hence why the implementation and use of big data analytics for consumer and market research sees such an uptake. This has contributed to datafied knowledge being perceived as a credible source of consumer-oriented knowledge by my participants. If put into relation to the research of around 10 years ago by Grandclément and Gaglio (2011), where focus groups were perceived as an ideal method and credible source for knowledge about consumers, participants now referred to qualitative methods as slow, inefficient and, most notably, subjective and unreliable. Despite some mentioning the ideal case of a mixed-methods approach, datafied knowledge was portrayed as the better kind of knowledge. The idiom of co-production in my research has been able to show "how certain conceptual designs and cognitive formulations gain ground at the expense of others, and how, once adopted, these successful settlements come to be seen as natural, inevitable, or determined in advance" (Jasanoff, 2006a, p. 277).

So far, Critical Marketing Studies rarely have put their focus on the procedures behind the stability as well as the instability of knowledge. If at all, they rather have looked at the processes of stabilising knowledge, as Dubuisson-Quellier's (2010) example has shown. However, it is through both accounts of stability and instability, through the process of rendering certain kinds of knowledge credible at the expense of other kinds, that the normativities, worldviews, socialities, and technologies behind these practices are revealed. Through the epistemic opacity of the technology and the analytics operations, as well as fuelled by the epistemic authority through the paradigm of positivism and the ideology of dataism, datafied knowledge has become stabilised as the main kind of consumer-oriented knowledge. This leads to multiple ontologies of the real consumer emerging in marketing as marketers are enacting multiple representations. There is, in some instances, still the creation and enactment of an imagined conceptualisation of the consumer – at least in those organisations which still rely on traditional consumer research methods. There is the creation of data double consumer conceptualisations, which are perceived and enacted as a factual copy of a real consumer. But there is also the creation and enactment of the postulated consumer conceptualisation as the data double is a misconception that is not feasible. This shows the contribution of the idiom of co-production to the cultural materialist strand of Critical Marketing Studies as a framework for researching and understanding how knowledge about consumers is produced, used and leads to different enactments of consumer conceptualisations.

7.3.2. Empirical implications – Critical Marketing Studies

One of the main contributions of this study is that it has managed to unpack the sociotechnical assemblages that lie at the core of big data analytics and marketing. The findings have shown that big data analytics is deeply entrenched in normative and social processes. The paradigm of digital positivism and the ideology of dataism (boyd and Crawford, 2012; Crawford et al., 2014; Fuchs, 2017; van Dijck, 2014) emerged on a regular basis. Particularly when participants were talking about their expectations regarding big data analytics, how they perceived the resulting datafied knowledge, as well as how they further used that knowledge for creating consumer conceptualisations, the dataist ideology was omnipresent.

At the same time, the analysis has illustrated how the implementation of big data analytics, as well as the practice of doing analytics, are not neutral procedures that can be separated either from the organisational context, the markets they are operating in or from the wider society of which they are part. Big data analytics is a very collaborative, social practice, in which decisions are made on a subjective basis, often in terms of what works and not in terms of what is required, as Ariztía (2018) has discussed before. Furthermore, the participants elaborated on the technological limitations as well as the statistical limitations of big data analytics, through which datafied knowledge is created, that is only applicable within a certain variance (Agostinho et al., 2019; Kaplan et al., 2014; Symons and Alvarado, 2016).

An important finding contributing to the cultural materialist strand of Critical Marketing Literature is that participants showed how, in their organisations, datafied knowledge leads to consumers being defined largely based on their marketable qualities (Fourcade and Healy, 2017a; Krenn, 2017a). Datafied knowledge about the consumers is relevant if it relates to the product or service in question, and it contributes to better promotion or selling of this product or service. Big data analytics contributes to this focus on the marketable qualities of the consumer. As Pridmore (2012, p. 321) has alluded to, big data analytics for consumer conceptualisations relies on the "patterns and associations deemed to be 'of value'" and are "directed towards obtaining the maximum current and potential profitability from differing sets of consumers". Through the stabilisation of that datafied knowledge, particularly in the form of the consumer conceptualisations, the measures taken on behalf of the consumers, the decisions regarding what offers to display or what insurance premiums to pay, rely solely on the consumers' marketable qualities. The data double has a purpose, which is to increase profit at the expense of the consumer. This is despite many participants referring to the advantages for the consumers due to only receiving 'relevant' advertisements and offers.

Consequently, these measures have an impact on the real consumers. Through the lens of performativity of marketing practices, developed by ANT researchers in Critical Marketing, (Ariztía, 2014; Callon, 2006; MacKenzie, 2006; Roscoe, 2014), my findings show that the Actor-Network of marketing through its methods of co-producing consumer-oriented knowledge, actively shape how consumers ought to look and how they should behave. As some participants in this research stated, for them, the ultimate aim of

big data analytics in marketing is to have consumers behave in such a way as the marketers want them to behave. Thus, through the precise consumer conceptualisations and the intention of predicting how consumers will act and react in the near future, the expectation is to have a consumer that perfectly fits the representation that has been created about them in the first place: a representation in which the consumer's marketable qualities are predominant and are reinforced in real-life settings.

All these issues are occurring in the opaqueness of the sociotechnical assemblage of big data analytics, as well as in the organisations employing such digital marketing measures. This is why the consumers are, most of the time, unaware of these operations. And with marketing operations turning increasingly towards manipulating consumer consumption behaviour through these digital and datafied means (Wertenbroch, 2015a), consumers are left under the impression of having free and deliberate choices when, in fact, many choices have already been made on their behalf (Sunstein, 2015). Even some of the participants appeared to be unaware of – or had not reflected on – the negative implications that might occur through using big data analytics in marketing.

7.3.3. Empirical implications – Critical Data and Surveillance Studies

The expectation of producing a data double consumer conceptualisation through big data analytics, which is a postulated consumer conceptualisation, also contributes to the Critical Data and Surveillance Studies literature. The idea of the virtual data double was established by Haggerty and Ericson (2000) when describing the process behind the 'surveillant assemblage.' The surveillant assemblage, and thus the data double does not only relate to marketing and consumer surveillance. Still, Haggerty and Ericson (2000, p. 616) discussed the data double in relation to marketing settings, where its aim was to create surplus value through the creation of consumer profiles, to "refine service delivery and target specific markets." As with other contributions in the field of Surveillance Studies (see, for example, Matzner, 2016), my findings build and expand on the idea of the data double.

As I have discussed in my findings and in the previous section, the co-production of consumer-oriented knowledge is not a linear process but shows a significant level of cyclicality in many of the steps that are involved (Jasanoff, 2006a). This results in the consumer conceptualisations not being created as one-offs but, instead, requiring constant

re-qualification, as Dubuisson-Quellier (2010) has also shown. This means that different consumer conceptualisations exist simultaneously, not only due to the different methods used to produce knowledge but also because each conceptualisation, each data double, is constantly being re-worked and re-qualified. Matzner (2016) theorises on the existence of multiple data doubles of one individual, building on the concept of "dividuals", as coined by Deleuze (1992). As data doubles are only created for certain instances and based on particular characteristics of the individual, data doubles have to be multiple, and all versions act on the individual through the performativity of data (Matzner, 2016). Not only do my findings show that the data doubles are (re-)produced in a constant cyclical process, but they also further develop the concept of the data double by introducing the concept of the postulation. The postulation is the result of the false premise which emerges from the expectation that through big data analytics, factual and infallible knowledge can be produced.

It is important to discuss the idea of the postulated conceptualisation as it is the foundation of all the measures that are taken which directly impact targeted individuals – be that in the area of marketing or in any other areas where surveillant measures are profiling and targeting individuals. As shown, these conceptualisations were used by my participants and their organisations to define which advertisements and offers certain consumers would receive, or – in the case of the insurance company – what kind of insurance premiums individual consumers were offered and what price they would be paying for it. As the findings here suggest, many of the participants had a relatively deterministic view of their consumers, meaning they considered their consumption behaviour and consumer identity to be deeply ingrained in the consumers and of which the consumers were often not consciously aware of. Uncovering this exact knowledge about consumers was seen as means of helping consumers realise their preferences and needs by proposing products and services they might need before they had even thought about needing them (Araujo et al., 2010; Bauman and Lyon, 2016).

As with the performativity discussed in the previous section, this means that, from a consumer point of view, the shaping of who they are and how they can behave as a consumer lies entirely in the hands of the data analysts and marketers, who control the identity and behaviour of consumers (Zwick and Denegri Knott, 2009). It can be argued that the control and power over consumption and the consumer have always been in the

hands of the organisations and the marketers. Many studies in the field of Critical Marketing have shown this before (see for example Ariztía, 2014; Cochoy, 2010; Desroches and Marcoux, 2011; Garcia, 1986; Hardie and MacKenzie, 2007). Yet, as I show with my thesis, the difference is that before the implementation of big data analytics, in many domains of marketing, these conceptualisations were broader and less individualised – they were imagined. As conceptualisations are becoming individualised profiles, consumers are impacted more specifically and directly about how their consumption behaviour should take place and how they should look, irrespective of whether this represents how they see themselves (Ball et al., 2016).

This leads to the second main issue in relation to Surveillance and Critical Data Studies: discrimination through marketing practices. By considering datafied consumer conceptualisations as postulated consumers instead of data doubles, the discriminatory potential of digital and datafied marketing becomes even more obvious. As discussed in the literature review and as shown by the participants, the co-production of datafied consumer-oriented knowledge is not free from error and uncertainties (Agostinho et al., 2019; boyd and Crawford, 2012). And the use of algorithms and big data technologies does not turn operations of co-producing knowledge magically 'neutral' and 'objective' (Gillespie, 2014). Instead, algorithms are value-laden (Bilić, 2016; Mager, 2012), subjective and political (Allhutter et al., 2020), and have the effect of reproducing and reinforcing the biases and discriminations that are predominant in our society (Friedman and Nissenbaum, 1996; Mittelstadt et al., 2016).

Hence, the actions taken, and decisions made through datafied knowledge about consumers should be carefully considered. Denying applicants insurance policies or pricing them higher, depending on their risk to default or commit fraud, are obvious and economic choices to be made in insurance companies. Similar reasoning lies behind the personalised offers that are being sent out to (potential) travellers or personalised advertisements of the newest song, movie, or series that are being sent to entertainment consumers. As with any other form of knowledge that helps assist these decisions, datafied knowledge about consumers can be erroneous. As my findings show, participants rarely reflect critically on the use of datafied knowledge, nor do they rely on postulated consumer conceptualisations. If concerns arise, they attribute to their own role as

consumers. In the meantime, consumer-oriented knowledge is viewed as factual, uncertainties are disregarded and decisions that are taken are seen as justified.

Which finally leads to the receiving end of the marketing chain – the real consumers and the question of what they can do to counter these effects as well as (re-)gain control over their consumer conceptualisation. As discussed in the literature, performativity is reciprocal, which means that the co-production of datafied knowledge of consumers can only ever represent what the consumers themselves make available (Fourcade and Healy, 2017a). Marketing is only able to represent the consumers in such a way as the consumers present themselves. And while, through the datafication of many of the marketing measures, it becomes increasingly difficult for consumers to control their data, and thus also to control these representations and conceptualisations, there are tools available which help them to do so.

These are, for example, the privacy measures for consumers in the digital sphere, such as privacy-enhancing internet browsers or browser extensions, such as the Brave browser or the Ghostery extension. These restrict tracking while a user is browsing the web, making individualised data collection more difficult. The same tools are available for mobile devices, where more applications are available which restrict the tracking of the smartphone, such as Blockada. Also, through the implementation of the GDPR and in its aftermath, the tracking of consumer data has become increasingly complicated, while individual users have been provided with the possibility to further restrict the use of tracking cookies on their devices (Chester, 2012; Mellet and Beauvisage, 2020).

These are only some of the many options that individuals can take to counter the tracking and surveillance of their consumption behaviour. Marx (2003, 2009) has discussed further activities and responses individuals can take to counter the increasing data collection, such as “avoidance” of large retail chains that are known for their intrusive tracking, “piggybacking”, “switching” identity and “distorting” the data by sharing loyalty cards or shopping accounts to create a large data mess about individuals, or the “blocking” activities as discussed above, to name a few (Marx, 2003, 374f.). These are, however, all active measures individuals must take to prevent the tracking of their consumption behaviour and their socio-demographics and which enables the wide-ranging effect and practices I have disentangled here in my research. The question persists: should the

burden of alleviating these practices remain in the hands of the individualised consumers alone?

7.4. Practical recommendations

This finally leads to the question of what this means for the practitioners, the data analysts and marketers, working with datafied consumer-oriented knowledge. Here, there are two main points that should be addressed, and which are closely interlinked. The first is the necessity of a more thorough analysis and handling of errors and uncertainties. This leads to the second point concerning a more critical reflection regarding the co-production and use of datafied consumer-oriented knowledge in general in organisations. As I have mentioned in the introduction of my thesis, the aim of my research was not to criticise or defend big data analytics in consumer research and marketing, nor to judge how datafied knowledge about consumers is used. The goal was to open up the practices that lie at the core of these operations, and to engage with the normativities of these practices. This extends to a reflection on how conceptualisations of consumers are enacted and performed through big data analytics. This engagement with and reflection upon those practices are also necessary in organisations.

Both, for researchers as well as practitioners, there is a necessity to engage more with the errors and uncertainties that emerge through the co-production of knowledge, and particularly through the use of big data analytics (Ariztía, 2018; Jasanoff, 2017). As discussed in my literature review, particularly in social sciences or in marketing, the analysis of errors needs to be made an intrinsic part of data analysis, as is done in other disciplines already – certainly in the hard sciences. In the use of big data analytics, the analysis and general reflection on errors and uncertainties is a crucial task. As Busch (2014, p. 1736) states, "even a 95% confidence interval means that 5% of the reported results will be erroneous", and depending on how the data was sampled and how much "noise" the data includes, the confidence interval is unknown and the rate of error can be much higher.

Further reflections on other factors contributing to errors and uncertainties are also required, such as the potential for bias in the data and the analysis (Agostinho et al., 2019). As my findings show, these reflections are rarely present in the organisations that have taken part in my research. Instead, datafied knowledge is communicated and translated

without considering potential errors, distilling the findings to their core messages. The difficulties with reflecting more thoroughly about the errors and uncertainties of datafied knowledge can be manifold. Practitioners might lack the capabilities to do so on a wider basis as the analysis of errors might require them to recollect and reanalyse big data, as Busch (2014) mentions. Furthermore, as I have shown, the epistemic authority of datafied consumer-oriented knowledge often opposes the tendency to question the validity of that knowledge. As datafied knowledge is perceived as objective and infallible through the ideology of dataism, practitioners see no need to reflect on the uncertainties and analyse the errors.

There is an urgent need to introduce a more widespread perception and awareness about the issues of error and uncertainties, which should first and foremost start with the basis of the stabilisation and communication of datafied knowledge. As seen in the findings, particularly in this step, where the results of big data analytics are translated into datafied knowledge that is being widely disseminated, the errors and uncertainties are being dismissed. In a first step, acknowledging the potential of errors and uncertainties when communicating knowledge would be helpful in making end-users aware of the existence of these issues. This might also further contribute to demystifying big data analytics more generally and reduce the presence of the ideology of dataism amongst marketers.

Which leads to the second point of the practical recommendations. In general, there needs to be more critical reflection amongst practitioners about how datafied knowledge is being co-produced and enmeshed in cognitive, social, normative and technological procedures (Jasanoff, 2006b). The sociotechnical black box of big data analytics needs to be opened so the procedures behind the operations can be better understood. This relates, on the one hand, to what some of the participants expressed in my research, in that people working with datafied knowledge need to gain a better mathematical, statistical and technical understanding of how it works. While it is not necessary for everyone to become a data analyst or data scientist, understanding the core concepts of data analytics helps to better understand how datafied knowledge is being produced – and, thus, to understand the risks and uncertainties that can be included in the knowledge.

On the other hand, this also means becoming aware of the persistent ideology of dataism and, subsequently, steering away from it. While considering this, some participants

critically reflected on what they often called the hype around big data: big data analytics being perceived as the better and more efficient way of gaining knowledge in their organisations. Regarding this, it would be highly beneficial for the practical use of creating consumer conceptualisations to introduce a mixed-methods approach, something which some participants considered to be the ideal case scenario. This approach enables marketers to profit from the advantages of both methods, while also keeping in mind that both methods have their problems and downsides.

Steering away from the dataist approach to consumer-oriented knowledge and acknowledging the errors and uncertainties of big data analytics can help us to understand how big data analytics creates a more objective knowledge compared to other methods, but is not able to create exact knowledge about consumers. Ultimately, this should lead to an understanding that consumer conceptualisations cannot be an exact data double of the real consumer and, thus, their use should be considered carefully and their potential impact on the consumers reflected upon. This also means that consumer conceptualisations cannot turn into consumer postulations as marketers would not blindly trust the conceptualisations anymore.

7.5. Limitations and opportunities

In this thesis, I have looked at the different ways consumers are conceptualised against the backdrop of changing methods for consumer-oriented knowledge production, particularly through the introduction of big data analytics. Yet, as with all research, there are practical limitations that should be acknowledged as they have influenced the way this research has been conducted. This section will acknowledge and critically reflect on the limitations of my study. These are, on the one hand, limitations in the practical sense of my research methods, which are largely concerned with access to the site(s) of the research and the possibilities of my empirical study. These limitations have, however, brought forth new and unforeseen opportunities, such as the inclusion of industry conferences as a field of study. On the other hand, the limitations relate to my role and the researcher's role in general. As my research is also a practice of the co-production of knowledge, it is important to highlight the importance of contextualisation (Jasanoff, 2006a, p. 276) here, as well.

7.5.1. A changing research design – limitations and opportunities

As I have already explained in my methodology chapter, the difficulties encountered in accessing the research sites led to adapting my research design. From my first research concept of one large case study to the current design of smaller case studies in combination with key-informant interviews and FCEs, different iterations were developed, and modifications had to be made. Changing from an in-depth case study to multiple small case studies has limited the research in that thorough participant observations in the form of work shadowing, as initially planned, were simply not possible. Especially for the research of uncertainties in knowledge, the difficulties of co-producing datafied knowledge and the errors in big data analytics, these observations of the practices of co-producing knowledge would have allowed for a more external view on these issues. Talking with participants about these issues proved challenging at times as they rarely talked about these issues on their own. If these issues where addressed, they were often dismissed as non-crucial by the participants, despite several accounts showing that uncertainties were often present, could be persistent and were not rectified easily, as discussed in my findings. This is not to say that the participants did not provide sufficient insight into the presence of uncertainties in datafied knowledge, nor how these uncertainties were handled in the translation, simplification, and communication of the knowledge. Nonetheless, the observation of the entire process of co-producing knowledge in one organisation – including stabilising knowledge and endowing epistemic authority – could have provided a better understanding of how uncertainties in datafied knowledge are handled and translated in that organisation.

The changes in my research design have not only led to limitations within my research. A new opportunity in the form of FCE's emerged that I had not considered earlier. Initially, I had visited industry conferences on big data analytics as well as digital marketing to build a network through which I intended to gain access to research sites. However, after visiting the first two events in quick succession, I realised how rich in data these events were on the topic of dataism and the co-production of knowledge. As I was in need of alternative sources from which to gather data, I started to investigate methodological options for using conferences as a site of data collection. As explained in the methods chapter, the idea of researching conferences had been developed in the early 2000s conceptualised as Field Configuring Events. As Lampel and Meyer (2008b, p. 1025) have explained, FCEs are non-academic conferences which shape "the

emergence and developmental trajectories of technologies, markets, industries, and professions", and are sites where important decisions are made regarding the future development and uptake of technologies. The industry conferences I had visited had the characteristics of FCEs and seemed promising in terms of contributing to answering the *why* question of my research – why marketers conceptualise consumers differently when relying on big data analytics.

When looking at the events I consequently visited as a site of research, it emerged that these events had a very hyped and excessively positive attitude towards big data analytics, where critical topics were either missing entirely or dismissed quickly when addressed. This attitude was understandable, as the events were often intended as sale events for big data analytics technology. Still, considering that there was an impact of this hyped and highly marketised atmosphere on the end-users, such as data analysts and digital marketers, at these events – being potential consumers – the inclusion of FCEs as sites of research was fruitful.

The FCEs connected well with how normativities, the sociality, politics, discourse, identities, etc., impacted the co-production of datafied knowledge and played a role in the stabilisation of that knowledge. FCEs had become a site where different kinds of knowledge – traditional as well as datafied – competed in terms of credibility and epistemic authority, with big data analytics, again, being highly favoured. With some participants in my research being recruited at these FCEs, they were directly engaged and influenced by the discourses taking place there, by the normativity and the ideology of dataism. This, to a certain extent, contributed to the perception of the superiority of datafied knowledge and a belief in the potential of producing a data double consumer conceptualisation. Hence, the inclusion of the FCEs has provided additional insight into why marketers conceptualise consumers differently when relying on big data analytics.

7.5.2. Contextualising my research – co-producing a PhD thesis

Another limitation of my research, and one which applies to research in general, is that it requires a critical reflection on the context of its co-production (Jasanoff, 2006a). The knowledge I have used and created for my research and for writing my thesis is co-produced in the same sense as the datafied knowledge of which I talk at length is being co-produced. The context of my research – as a PhD thesis at the University of St

Andrews in Scotland – has limited it to a certain extent, in that my findings, although relevant, are not generalisable. Part of these limitations have been presented above, i.e. those which stemmed mainly from changing the research design, a move which may also be considered to have contributed to the context of my research. This has defined the circumstances of my study insofar as it has limited the depth of my empirical research. Future research could certainly benefit from aiming at investigating the co-production of datafied knowledge in depth in one or two organisations.

The research in my thesis involves a range of industries rather than focusing on one or two. This partially stems from the difficulties encountered in the access to data but is also a result of my snowball sampling strategy. The consequence of this is that it is difficult to consider whether industrial differences exist regarding the use of big data analytics or the conceptualisation of consumers. Some industries, such as insurance as part of the finance industry, have a lengthy history of using consumer data to make decisions based on this knowledge. In the US, this practice can be traced back to 1840 with the appearance of the first consumer credit reporting organisations, and with a stark improvement of operations in the 1960s through the introduction of database computing (Lauer, 2010). In other industries, the use of digital consumer data does not have such a history and, as a result, the shift towards using digital consumer data and big data analytics might be slower, less sophisticated, or may be considered more critically. This context could not be acknowledged in my research as there were too few participants from individual industries and too many different industries. Also here, future research – with the necessary access to the research sites – could research these potential differences arising, potentially, from the history of working with digital and non-digital consumer data.

Furthermore, the context of researching the use of big data analytics in marketing at a European university means that my research has a very Eurocentric and Anglo-American focus – in terms of the concept of my research as well as the location of my research participants, organisations and FCE's. As a result, cultural differences were not considered in my research, while being significant. Jasanoff (2006c, 2017) and Jasanoff and Kim (2009) have discussed, on multiple occasions, how there are differences in the co-production of knowledge and the conceptualisation or imagination of science and technology depending on the country or continent. As I have shown earlier, the co-production of knowledge cannot be considered without also considering social and

political actors and institutions. The same can be claimed for surveillant use of datafied knowledge, which also has a social and political context which can be included in the research. Some studies have specifically reflected on the cultural and national differences regarding surveillance and privacy, such as Luther and Radovic's (2013) research on the different understandings of surveillance and privacy in Japan.

Interestingly, this is something that one of my research participants reflected on regarding her work in marketing and behaviour change, as well as her academic background in psychology. As she noted, marketing research and behaviour change research generally express an undifferentiated, westernised view in their research, which often does not hold in a wider context. Research is almost exclusively focused on so-called "*weird* markets", the results of which cannot be simply transferred to other countries: for example, the Global South:

> *"The thing is, we're working with a global client that's trying to use data analytics to change their customers' behaviour. Yeah, there are interesting differences across different countries and cultures. I don't think anyone really fully understands yet because most of the psychological and behaviour change research has been done on, have you heard of weird market? (...) I think the W is Western. (...) The E is educated, the I is industrialized, the R is rich, and the D is democratic, I think? (...) Yes, so most of the research has been done on those groups. For example, like groups in, in Africa that there's one researcher who we've been in touch with at this place called the (research institute) and they're finding that lots of these things that we thought were universal principles in behaviour are actually not universal but are more kind of specific to these weird markets. So, as you try and roll these things out across different cultures like the same things that work in the UK, for example, just doesn't translate into Germany. So, I think that there's going to be a lot of work to be done on that and no one's really cracked that yet."* [Key5_Consultant, Pos. 74-81]

These considerations of cultural and nation-depending differences in the co-production of datafied knowledge, as well as in its use, would be necessary in this context of big data analytics and marketing. However, the inclusion of this angle of research was not possible in the current study due to time constraints and an even wider gap of access to potential

research sites. As with the previous limitations, this is certainly something that should be investigated further.

7.6. Agenda for Future research

In the previous section, I elaborated on the limitations of my research in which I discussed future research possibilities. Yet, there are more future research possibilities that need to be explored and which stem from my findings. As my methodology was deliberately exploratory in order to investigate the impact of big data analytics on the conceptualisation of consumers in the first place, further research can certainly profit from a wider methodological approach. This will be discussed in the first part of this section. In the second part, I will discuss other, more exploratory, possibilities that have emerged.

As noted in the previous section, as my research involved an exploratory study, the results are not generalisable, despite being novel and contributing to the wider research community. Future research would certainly benefit from a wider approach, one which covered more generalisable questions. As my findings have shown, the uncertainty of datafied consumer-oriented knowledge is an issue that impacts the conceptualisation of consumers and, through this, also impacts and performs the real consumer through the corresponding marketing practices. However, this uncertainty in knowledge is rarely acknowledged and, even less so, addressed or mitigated. Instead, as my participants have stated, uncertainties are ignored, not communicated further and – reflecting the heart of my research – not considered in the postulation of the consumer. As a result, future research on a wider scale, potentially quantitatively or using a mixed-methods approach, could investigate how uncertainties in datafied knowledge are handled more broadly. The intention of such research could be to investigate measures that are already in place in some organisations, and how to better communicate the uncertainties and errors in datafied knowledge they might identify. The aim would be to have marketers be more critical and reflective when conceptualising consumers and when using these conceptualisations to target consumers with offers and advertising. As Roscoe (2016, p. 132) rightly states, addressing the performativity of economics or marketing, as is the case here, is not so much about stating that marketing is performative, it is rather about asking "what kind of world we wish to see performed."

Another avenue for future research would be to take on a concept, the practice of which remained broadly unclear, which many participants addressed as the ideal case for consumer and market research – a mixed-methods approach in marketing. The participants often mentioned that, in the ideal situation and with enough resources, consumer and market knowledge should be the result of a mixed-methods approach, with big data analytics and qualitative research methods complementing each other. Also, in the academic and industry literature, some authors have started to call for a combination of big data and qualitative data: for example, by relying on "thick" data gained from ethnographic research (Wang, 2013). One possible approach would be to exploratively research how the combination of both research methods works in practice, how marketers aim at gaining the best result from both and how the issue of epistemic credibility and authority are handled, in this regard. The findings might help us better understand the conceptualisation of consumers and the way marketers target consumers based on their available knowledge.

Lastly, it is also important to further investigate the role and effects of algorithms in the co-production of datafied consumer-oriented knowledge. While, in my research, the topic of algorithms has been addressed in the literature review and the findings, I should stress that it has not been my primary focus. However, algorithms play an important role in the co-production of knowledge and contribute to the obfuscation of the uncertainties and errors in the datafied knowledge. Similar issues have been discussed in the literature, where the research of algorithms is increasingly becoming popular. Yet, as the field of application of algorithms is quite diverse, so is the research on their (social) impact. On the one hand, an entire academic discipline has emerged from within the Computer Sciences, which deals with "Fairness, Accountability and Transparency of Machine Learning" (FATML) (Binns, 2018; Dwork et al., 2011; Friedler et al., 2018). It is a worthy development that in addition to studies in technological academia and industry, societal concerns and issues with technology are emerging. However, a common criticism of FATML is that the proposed solutions are solely technologically-based. The wider issues behind algorithms and machine learning are rarely addressed, whether they are of a societal nature, such as bias and discrimination, or the technological and philosophical nature of epistemic opacity (Peña Gangadharan and Niklas, 2019).

On the other hand, as in other fields, critical research into algorithms and machine learning has become an important focus, and is applied to many topics, such as urban development (Coletta and Kitchin, 2017), crime and justice (Sanders and Sheptycki, 2017) or libraries (van Otterlo, 2016). There are already studies on the use of algorithms in marketing and advertisement. However, these have mainly been limited to the receiving ends of algorithmic profiling (Ruckenstein and Granroth, 2020) or on the issues of discrimination and bias through algorithmic-driven online advertising (Sweeney, 2013). An investigation into the practitioner side of the algorithm, looking at how algorithms contribute to the co-production of datafied consumer-oriented knowledge, might be worthwhile. With errors, uncertainties, bias, and discrimination being integral, although unwanted, parts of algorithms, the findings can help to further deconstruct the practice of co-producing consumer-oriented knowledge, and the subsequent postulating of the consumers.

7.7. Concluding thoughts

In this chapter, I have highlighted the main findings and contributions of my thesis in relation to the current debates in the literature, as well as to the practitioners who have been a vital source of insight for my findings. I have highlighted and addressed the limitations of my approach, as well as made suggestions on future research endeavours. The aim of the thesis was to explore and uncover how big data analytics contributes to the profiling or conceptualising of consumers in organisations. The academic literature has discussed how big data analytics is increasingly being put to use in marketing (Amado et al., 2018), as well as the wide range effects it has on consumers (Zwick and Denegri Knott, 2009), and its inherent surveillant nature (Ball, 2017). However, little research had been done on how big data analytics in practice contributes to the profiling, predicting, and targeting of the consumer.

Relying on the theoretical approaches from Critical Marketing Studies and the idiom of co-production of knowledge by Jasanoff (2006b), and analysing data collected through interviews, attending FCEs and gathering documents, has allowed me to investigate *how* and *why* big data analytics leads to different consumer conceptualisations. There are three key contributions resulting from this research:

1. I have shown that big data analytics indeed contributes to a different kind of consumer conceptualisation, but in a twofold way. First, the one that the participants expected to emerge, which is in the form of a data double consumer which is, however, non-existent in reality. And second, the one that emerged in practice in the participating organisations in the form of a postulated consumer.

2. As marketing practices are *performative* (Callon, 2006; Cochoy, 1998), these two different consumer conceptualisations can have a real impact on the consumers. The postulated consumer conceptualisation helps explain how bias, discriminations and errors, which stem from the big data analysis, ultimately affect the consumer. This is because they are not reflected upon.

3. My research is an empirical contribution which shows the usefulness of looking at knowledge and insight in marketing through the lens of co-production. Critical Marketing Studies largely rely on concepts of STS and consider marketing a set of sociotechnical assemblages (Callon and Muniesa, 2005). They have also considered the intricacy of producing knowledge in organisations (Grandclément and Gaglio, 2011; Sunderland and Denny, 2011). Yet, the stabilisation and credibility of knowledge and its co-production have, so far, been neglected.

The initial goal of my thesis was to open up the sociotechnical black box of big data analytics, to research and understand how big data analytics contributes to creating knowledge about consumers, and how marketers themselves try to make sense of this sociotechnical apparatus. Although the empirical research has not been as smooth as initially planned and expected, the originality and novelty of my research lies in the findings. They provide a rich account of how datafied knowledge is co-produced, uncertainties and errors in knowledge emerge and persist, and how this leads to multiple consumer conceptualisations being enacted. But my thesis has also highlighted the necessity to look further into how uncertainties in datafied knowledge are handled, particularly by focusing on the role of the algorithms.

The intention of my research has never been to judge the use of big data analytics in marketing. Instead, my thesis is meant as a reflective piece of research, one which has

unpacked the normativity and the practice of big data analytics in marketing and the conceptualisation of the consumer. As big data analytics, AI and ML will persist and, potentially, increase in its application, it will be important to remain vigilant about how it is used and how it might contribute to shaping the world we live in.

References

Aaker, J. L. (1999) 'The Malleable Self: The Role of Self-Expression in Persuasion', *Journal of Marketing Research*, vol. 36, no. 1, p. 45.

Agostinho, D., Ring, A., Veel, K., D'Ignazio, C. and Thylstrup, N. B. (2019) 'Uncertain Archives: Approaching the Unknowns, Errors, and Vulnerabilities of Big Data through Cultural Theories of the Archive', *Surveillance & Society*, vol. 17, 3/4, pp. 422–441.

Alharthi, A., Krotov, V. and Bowman, M. (2017) 'Addressing barriers to big data', *Business Horizons*, vol. 60, no. 3, pp. 285–292.

Allhutter, D., Cech, F., Fischer, F., Grill, G. and Mager, A. (2020) 'Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective', *Frontiers in Big Data*, vol. 3, p. 326.

Alvesson, M. and Ashcraft, K. L. (2012) 'Interviews', in Symon, G. and Cassell, C. (eds) *Qualitative organizational research: Core methods and current challenges*, Los Angeles, London, Sage, pp. 239–257.

Amado, A., Cortez, P., Rita, P. and Moro, S. (2018) 'Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis', *European Research on Management and Business Economics*, vol. 24, no. 1, pp. 1–7.

Anderson, B. (2006) *Imagined communities: Reflections on the origin and spread of nationalism*, London, MPublishing, University of Michigan Library.

Andrejevic, M., Hearn, A. and Kennedy, H. (2015) 'Cultural studies of data mining: Introduction', *European Journal of Cultural Studies*, vol. 18, 4-5, pp. 379–394.

Araujo, L. (2007) 'Markets, market-making and marketing', *Marketing Theory*, vol. 7, no. 3, pp. 211–226.

Araujo, L., Finch, J. and Kjellberg, H. (2010) 'Reconnecting marketing to markets: An introduction', in Araujo, L., Finch, J. and Kjellberg, H. (eds) *Reconnecting marketing to markets*, Oxford University Press.

Ariztía, T. (2014) 'Housing Markets Performing Class: Middle-Class Cultures and Market Professionals in Chile', *The Sociological Review*, vol. 62, no. 2, pp. 400–420.

Ariztía, T. (2015) 'Unpacking insight: How consumers are qualified by advertising agencies', *Journal of Consumer Culture*, vol. 15, no. 2, pp. 143–162.

Ariztía, T. (2018) 'Consumer databases as practical accomplishments: The making of digital objects in three movements', *Journal of Cultural Economy*, vol. 11, no. 3.

Arndt, J. (1985) 'On Making Marketing Science More Scientific: Role of Orientations, Paradigms, Metaphors, and Puzzle Solving', *Journal of Marketing*, vol. 49, no. 3, pp. 11–23 [Online]. Available at http://www.jstor.org/stable/1251612.

Arunachalam, D. and Kumar, N. (2018) 'Benefit-based consumer segmentation and performance evaluation of clustering approaches: An evidence of data-driven decision-making', *Expert Systems with Applications*, vol. 111, pp. 11–34.

Ascarza, E., Neslin, S. A., Netzer, O., Anderson, Z., Fader, P. S., Gupta, S., Hardie, B. G. S., Lemmens, A., Libai, B., Neal, D., Provost, F. and Schrift, R. (2018) 'In Pursuit of Enhanced Customer Retention Management: Review, Key Issues, and Future Directions', *Customer Needs and Solutions*, vol. 5, 1-2, pp. 65–81.

Azimont, F. and Araujo, L. (2010) 'Governing firms, shaping markets: The role of calculative devices', in Araujo, L., Finch, J. and Kjellberg, H. (eds) *Reconnecting marketing to markets*, Oxford University Press.

Ball, K. (2017) 'All consuming surveillance: Surveillance as marketplace icon', *Consumption Markets & Culture*, vol. 20, no. 2, pp. 95–100.

Ball, K., Di Domenico, M. L. and Nunan, D. (2016) 'Big Data Surveillance and the Body-subject', *Body & Society*, vol. 22, no. 2, pp. 58–81.

Bauman, Z. and Lyon, D. (2016) *Liquid surveillance: A conversation*, 2nd edn, Cambridge, Malden, Polity.

Beer, D. (2017) 'The data analytics industry and the promises of real-time knowing: Perpetuating and deploying a rationality of speed', *Journal of Cultural Economy*, vol. 10, no. 1, pp. 21–33.

Behera, R. K., Gunasekaran, A., Gupta, S., Kamboj, S. and Bala, P. K. (2020) 'Personalized digital marketing recommender engine', *Journal of Retailing and Consumer Services*, vol. 53, p. 101799.

Bello-Orgaz, G., Jung, J. J. and Camacho, D. (2016) 'Social big data: Recent achievements and new challenges', *Information Fusion*, vol. 28, pp. 45–59.

Bilić, P. (2016) 'Search algorithms, hidden labour and information control', *Big Data & Society*, vol. 3, no. 1, 1-9.

Binns, R. (2018) 'Fairness in Machine Learning: Lessons from political philosophy', *Proceedings of Machine Learning Research*, no. 81, pp. 1–11.

Bogner, A., Littig, B. and Menz, W. (2009) 'Introduction: Expert Interviews — An Introduction to a New Methodological Debate', in Bogner, A., Littig, B. and Menz, W. (eds) *Interviewing Experts*, London, Palgrave Macmillan UK, pp. 1–13.

Bogner, A. and Menz, W. (2009) 'The Theory-Generating Expert Interview: Epistemological Interest, Forms of Knowledge, Interaction', in Bogner, A., Littig, B. and Menz, W. (eds) *Interviewing Experts*, London, Palgrave Macmillan UK, pp. 43–80.

Bosch, V. (2016) 'Big Data in Market Research: Why More Data Does Not Automatically Mean Better Information', *GfK Research*, vol. 8, no. 2, pp. 57–63.

Bourdieu, P. (2005) *The Social Structures of the Economy*, Cambridge, UK, Polity.

Bowker, G. C. and Star, S. L. (2000) 'Invisible Mediators of Action: Classification and the Ubiquity of Standards', *Mind, Culture, and Activity*, vol. 7, 1-2, pp. 147–163.

Bowker, G. C. and Star, S. L. (2008) *Sorting things out: Classification and its consequences*, Cambridge, Mass., MIT Press.

boyd, d. and Crawford, K. (2012) 'CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society*, vol. 15, no. 5, pp. 662–679.

Brannan, M. J. and Oultram, T. (2012) 'Participant Observation', in Symon, G. and Cassell, C. (eds) *Qualitative organizational research: Core methods and current challenges*, Los Angeles, London, Sage, pp. 296–313.

Braun, V., Clarke, V., Hayfield, N. and Terry, G. (2019) 'Thematic Analysis', in Liamputtong, P. (ed) *Handbook of Research Methods in Health Social Sciences*, Singapore, Springer Singapore, pp. 843–860.

Brinkmann, S. (2018) 'The Interview', in Denzin, N. K. and Lincoln, Y. S. (eds) *The Sage handbook of qualitative research*, Los Angeles, Sage, pp. 576–599.

Brock, V. and Khan, H. U. (2017) 'Big data analytics: Does organizational factor matters impact technology acceptance?', *Journal of Big Data*, vol. 4, no. 1, p. 5.

Busch, L. (2014) 'A Dozen Ways to Get Lost in Translation: Inherent Challenges in Large-Scale Data Sets', *International Journal of Communication*, vol. 8, pp. 1727–1744.

Cabrera-Sánchez, J.-P. and Villarejo-Ramos, Á. F. (2019) 'Factors Affecting the Adoptation of Big Data Analytics in Companies', *Revista de Administração de Empresas (Journal of Business Management*, vol. 59, no. 6, pp. 415–429.

Çalışkan, K. and Callon, M. (2009) 'Economization, part 1: Shifting attention from the economy towards processes of economization', *Economy and Society*, vol. 38, no. 3, pp. 369–398.

Çalışkan, K. and Callon, M. (2010) 'Economization, part 2: A research programme for the study of markets', *Economy and Society*, vol. 39, no. 1, pp. 1–32.

Callon, M. (1998) 'Introduction: The Embeddedness of Economic Markets in Economics', *The Sociological Review*, vol. 46, no. 1, pp. 1–57.

Callon, M. (1999) 'Actor-Network Theory—The Market Test', *The Sociological Review*, vol. 47, no. 1, pp. 181–195.

Callon, M. (2006) *What does it mean to say that economics is performative?*, CSI Working Papers Series 005.

Callon, M. (2010) 'Marketing as an art and science of market framing: Commentary', in Araujo, L., Finch, J. and Kjellberg, H. (eds) *Reconnecting marketing to markets*, Oxford University Press.

Callon, M., Méadel, C. and Rabeharisoa, V. (2002) 'The economy of qualities', *Economy and Society*, vol. 31, no. 2, pp. 194–217.

Callon, M. and Muniesa, F. (2005) 'Peripheral Vision: Economic Markets as Claculative Collective Devices', *Organization Studies*, vol. 26, no. 8, pp. 1229–1250.

Canhoto, A. I. and Backhouse, J. (2007) 'Profiling under conditions of ambiguity—An application in the financial services industry', *Journal of Retailing and Consumer Services*, vol. 14, no. 6, pp. 408–419.

Chakrabarti, R. and Mason, K. J. (2015) 'Can the pragmatist logic of inquiry inform consumer led market design', *Advances in Consumer Research*, vol. 43, no. 6, pp. 212–217.

Charitsis, V., Zwick, D. and Bradshaw, A. (2018) 'Creating Worlds that Create Audiences: Theorising Personal Data Markets in the Age of Communicative Capitalism', *TripleC*, vol. 16, no. 2, pp. 820–834.

Chen, H., Chiang, R. H. L. and Storey, V. C. (2012) 'Business Intelligence and Analytics: From Big Data to Big Impact', *MIS Quarterly*, vol. 36, no. 4, pp. 1165–1188.

Chester, J. (2012) 'Cookie Wars: How New Data Profiling and Targeting Techniques Threaten Citizens and Consumers in the "Big Data" Era', in Gutwirth, S., Leenes, R., Hert, P. d. and Poullet, Y. (eds) *European data protection: In good health?*, Dordrecht, New York, Springer, pp. 53–77.

Christmann, G. B. (2009) 'Expert Interviews on the Telephone: A Difficult Undertaking', in Bogner, A., Littig, B. and Menz, W. (eds) *Interviewing Experts*, London, Palgrave Macmillan UK, pp. 157–183.

Cochoy, F. (1998) 'Another discipline for the market economy: marketing as a performative knowledge and know-how for capitalism', *The Sociological Review*, vol. 46, S1, pp. 194–221.

Cochoy, F. (2010) 'Reconnecting marketing to 'market-things': How grocery equipment drove modern consumption (Progressive Grocer, 1929–1959)', in Araujo, L., Finch, J. and Kjellberg, H. (eds) *Reconnecting marketing to markets*, Oxford University Press.

Cochoy, F., Trompette, P. and Araujo, L. (2016) 'From market agencements to market agencing: An introduction', *Consumption Markets & Culture*, vol. 19, no. 1, pp. 3–16.

Coletta, C. and Kitchin, R. (2017) ‘Algorhythmic governance: Regulating the ‘heartbeat’ of a city using the Internet of Things’, *Big Data & Society*, vol. 4, no. 2, 205395171774241.

Collins, F. S., Morgan, M. and Patrinos, A. (2003) ‘The human genome project: Lessons from Large-Scale Biology’, *Science*, vol. 300, no. 5617, pp. 286–290.

Côrte-Real, N., Oliveira, T. and Ruivo, P. (2017) ‘Assessing business value of Big Data Analytics in European firms’, *Journal of Business Research*, vol. 70, pp. 379–390.

Cox, M. and Ellsworth, D. (1997) ‘Application-controlled demand paging for out of core visualization’, *Proceedings. Visualization '97 (Cat. No. 97CB36155)*, pp. 235–244 [Online]. Available at http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=663888&isnumber=14360.

Crawford, K., Miltner, K. and Gray, M. L. (2014) ‘Critiquing Big Data: Politics, Ethics, Epistemology: Special Section Introduction’, *International Journal of Communication*, vol. 8, pp. 1663–1672.

Creswell, J. W. (2014) *Research design: Qualitative, quantitative, and mixed methods approaches*, 4th edn, Thousand Oaks [etc.], Sage.

Criado Perez, C. (2019) *Invisible women: Exposing data bias in a world designed for men*, London, Chatto & Windus.

Darke, P., Shanks, G. and Broadbent, M. (1998) ‘Successfully completing case study research: Combining rigour, relevance and pragmatism’, *Information Systems Journal*, vol. 8, no. 4, pp. 273–289.

Darmody, A. and Zwick, D. (2020) ‘Manipulate to empower: Hyper-relevance and the contradictions of marketing in the age of surveillance capitalism’, *Big Data & Society*, vol. 7, no. 1, 205395172090411.

Davies, I. (2015) ‘“Theory” into “Knowledge”: Operationalizing Consumption Research into Behavioral Change with Pragmatism’, *Advances in Consumer Research*, vol. 43, no. 6, pp. 212–217.

Degli Esposti, S. (2014) ‘When big data meets dataveillance: The hidden side of analytics’, *Surveillance & Society*, vol. 12, no. 2, pp. 209–225 [Online]. Available at http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/download/ analytics/analytic.

Dekimpe, M. G. (2019) ‘Retailing and retailing research in the age of big data analytics’, *International Journal of Research in Marketing*.

Deleuze, G. (1992) ‘Postscripts on the Societies of Control’, *October*, vol. 59, Winter, pp. 3–7.

Desjardins, J. (2019) *How much data is generated each day* [Online], World Economic Forum. Available at https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/.

Desroches, P. and Marcoux, J.-S. (2011) ‘The Making of the Sensuous Consumer’, in Zwick, D. and Cayla, J. (eds) *Inside marketing: Practices, ideologies, devices*, Oxford, Oxford University Press, pp. 162–184.

Dewey, J. (1929) *Experience and Nature*, New York, NY, Norton.

Di Salvo, P. and Milan, S. (2020) ‘The Four Invisible Enemies in the First Pandemic of the 'Datafied Society'’, *SSRN Electronic Journal*.

Dorstewitz, P. and Kremer, A. (2016) ‘Introduction: Pragmatist Perspectives on Science and Technology Studies and Contemporary Dewey Studies’, *Pragmatism Today*, vol. 7, no. 1, pp. 5–10.

Dourish, P. (2016) ‘Algorithms and their others: Algorithmic culture in context’, *Big Data & Society*, vol. 3, no. 2, pp. 1–11.

Dubuisson-Quellier, S. (2010) ‘Product tastes, consumer tastes: The plurality of qualifications in product development and marketing activities’, in Araujo, L., Finch, J. and Kjellberg, H. (eds) *Reconnecting marketing to markets*, Oxford University Press.

Ducange, P., Pecori, R. and Mezzina, P. (2018) ‘A glimpse on big data analytics in the framework of marketing strategies’, *Soft Computing*, vol. 22, no. 1, pp. 325–342.

Duhigg, C. (16.02.2012) ‘How Companies Learn Your Secrets’, *The New York Times Magazine*, 16 February [Online]. Available at http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html (Accessed 28 November 2017).

Dwork, C., Hardt, M., Pitassi, T., Reingold, O. and Zemel, R. (2011) *Fairness Through Awareness* [Online]. Available at http://arxiv.org/pdf/1104.3913v2.

Elia, G., Polimeno, G., Solazzo, G. and Passiante, G. (2019) ‘A multi-dimension framework for value creation through big data’, *Industrial Marketing Management*.

Erevelles, S., Fukawa, N. and Swayne, L. (2016) ‘Big Data consumer analytics and the transformation of marketing’, *Journal of Business Research*, vol. 69, no. 2, pp. 897–904.

Espeland, W. N. and Sauder, M. (2007) ‘Rankings and Reactivity: How Public Measures Recreate Social Worlds’, *American Journal of Sociology*, vol. 113, no. 1, pp. 1–40.

Espeland, W. N. and Stevens, M. L. (1998) ‘Commensuration as a Social Process’, *Annual Review of Sociology*, vol. 24, no. 1, pp. 313–343.

Eyben, R. (2005) *Ethical and Methodological Issues in Researching International Aid Organisations: Note of a Workshop held at the Institute of Development Studies, 8 December 2004* [Online]. Available at https://www.theasa.org/downloads/ethics/workshop.doc (Accessed 1 December 2017).

Ezrahi, Y. (2006) ‘Science and the political imagination in contemporary democracies’, in Jasanoff, S. (ed) *States of knowledge: The co-production of science and social order*, London, New York, Routledge, pp. 254–273.

Feldman, M., Friedler, S., Moeller, J., Scheidegger, C. and Venkatasubramanian, S. (2015) *Certifying and removing disparate impact* [Online]. Available at http://arxiv.org/pdf/1412.3756v3.

Floridi, L. (2012) ‘Big Data and Their Epistemological Challenge’, *Philosophy & Technology*, vol. 25, no. 4, pp. 435–437.

Fong, N. M., Fang, Z. and Luo, X. (2015) ‘Geo-Conquesting: Competitive Locational Targeting of Mobile Promotions’, *Journal of Marketing Research*, vol. 52, no. 5, pp. 726–735.

Fourcade, M. and Healy, K. (2017c) ‘Categories All the Way Down’, *Historical Social Research*, vol. 42, no. 1, pp. 286–296.

Fourcade, M. and Healy, K. (2017a) ‘Seeing like a market’, *Socio-Economic Review*, vol. 15, no. 1, 9-29.

Fourcade, M. and Healy, K. (2017b [2013]) ‘Classification Situations: Life-Chances in the Neoliberal Era’, *Historical Social Research*, vol. 42, no. 1, pp. 23–51 [Online]. DOI: 10.12759/hsr.42.2017.1.23-51.

Friedler, S. A., Scheidegger, C., Venkatasubramanian, S., Choudhary, S., Hamilton, E. P. and Roth, D. (2018) *A comparative study of fairness-enhancing interventions in machine learning* [Online]. Available at http://arxiv.org/pdf/1802.04422v1.

Friedman, B. and Nissenbaum, H. (1996) ‘Bias in computer systems’, *ACM Transactions on Information Systems*, vol. 14, no. 3, pp. 330–347.

Froschauer, U. and Lueger, M. (2003) *Das qualitative Interview: Zur Praxis interpretativer Analyse sozialer Systeme*, UTB GmbH.

Fuchs, C. (2011) ‘Web 2.0, Prosumption, and Surveillance’, *Surveillance & Society*, vol. 8, no. 3, pp. 288–309.

Fuchs, C. (2014) ‘Social Media and the Public Sphere’, *TripleC*.

Fuchs, C. (2017) ‘From digital positivism and administrative big data analytics towards critical digital and social media research!’, *European Journal of Communication*, vol. 32, no. 1, pp. 37–49.

Gandomi, A. and Haider, M. (2015) ‘Beyond the hype: Big data concepts, methods, and analytics’, *International Journal of Information Management*, vol. 35, no. 2, pp. 137–144.

Garcia, M.-F. (1986) ‘La construction sociale d'un marché parfait’, *Actes de la recherche en sciences sociales*, vol. 65, no. 1, pp. 2–13.

Garud, R. (2008) ‘Conferences as Venues for the Configuration of Emerging Organizational Fields: The Case of Cochlear Implants’, *Journal of Management Studies*, vol. 45, no. 6, pp. 1061–1088.

Gerring, J. (2004) ‘What Is a Case Study and What Is It Good for?’, *American Political Science Review*, vol. 98, no. 02, pp. 341–354.

Gieryn, T. F. (1983) ‘Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists’, *American Sociological Review*, vol. 48, no. 6, p. 781.

Gieryn, T. F. (1999) *Cultural boundaries of science: Credibility on the line*, Chicago (Ill.), University of Chicago Press.

Gillespie, T. (2014) ‘The Relevance of Algorithms’, in Gillespie, T., Boczkowski, P. J. and Foot, K. A. (eds) *Media Technologies: Essays on Communication, Materiality, and Society*, MIT Press Scholarship Online, pp. 167–194.

Gitelman, L., ed. (2013) *"Raw data" is an oxymoron*, Cambridge (Mass.), MIT Press.

Gläser, J. and Laudel, G. (2009) ‘On Interviewing “Good” and “Bad” Experts’, in Bogner, A., Littig, B. and Menz, W. (eds) *Interviewing Experts*, London, Palgrave Macmillan UK, pp. 117–137.

Gläser, J. and Laudel, G. (2012) *Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen*, Wiesbaden, VS, Verl. für Sozialwiss.

Grandclément, C. and Gaglio, G. (2011) ‘Convoking the Consumer in Person: The Focus Group Effect’, in Zwick, D. and Cayla, J. (eds) *Inside marketing: Practices, ideologies, devices*, Oxford, Oxford University Press, pp. 87–114.

Greene, J. C., Caracelli, V. J. and Graham, W. F. (1989) ‘Toward a Conceptual Framework for Mixed-Method Evaluation Designs’, *Educational Evaluation and Policy Analysis*, vol. 11, no. 3, p. 255.

Grosman, J. and Reigeluth, T. (2019) ‘Perspectives on algorithmic normativities: Engineers, objects, activities’, *Big Data & Society*, vol. 6, no. 2, 205395171985874.

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M. and Feldberg, F. (2017) ‘Debating big data: A literature review on realizing value from big data’, *The Journal of Strategic Information Systems*, vol. 26, no. 3, pp. 191–209.

Hacking, I. (1986) ‘Making up People’, in Heller, T. C., Sosna, M. and Wellbery, D. E. (eds) *Reconstructing individualism: Autonomy, individuality, and the self in Western thought,* Stanford, Calif, Stanford University Press, pp. 161–171.

Hagberg, J. (2010) ‘Exchanging agencies: The case of NetOnNet’, in Araujo, L., Finch, J. and Kjellberg, H. (eds) *Reconnecting marketing to markets,* Oxford University Press.

Haggerty, K. D. and Ericson, R. V. (2000) ‘The surveillant assemblage’, *British Journal of Sociology,* vol. 51, no. 4, pp. 605–622.

Hardie, I. and MacKenzie, D. A. (2007) ‘Assembling an economic actor: the *agencement* of a Hedge Fund’, *The Sociological Review,* vol. 55, no. 1, pp. 57–80.

Hardwig, J. (1991) ‘The Role of Trust in Knowledge’, *The Journal of Philosopy,* vol. 88, no. 12, pp. 693–708.

Hardy, C. and Maguire, S. (2010) ‘Discourse, Field-Configuring Events, and Change in Organizations and Insitituional Fields: Narratives of DDT and the Stockholm Convention’, *Academy of Management Journal,* vol. 53, no. 6, pp. 1365–1392.

Harrison, H., Birks, M., Franklin, R. and Mills, J. (2017) ‘Case Study Research: Foundations and Methodological Orientations’, [34 paragraphs], *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research,* vol. 18, no. 1 [Online]. Available at http://nbn-resolving.de/urn:nbn:de:0114-fqs1701195.

Hartley, J. (2004) ‘Case Study Research’, in Cassell, C. and Symon, G. (eds) *Essential Guide to Qualitative Methods in Organizational Research,* 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom, SAGE Publications Ltd, pp. 323–333.

Harvey, C. (2019) *Data Storage 2020: Trends in Data Storage* [Online], Enterprise Storage. Available at https://www.enterprisestorageforum.com/backup-recovery/data-storage-2020-trends-in-data-storage.html (Accessed 30 April 2020).

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A. and Ullah Khan, S. (2015) ‘The rise of “big data” on cloud computing: Review and open research issues’, *Information Systems,* vol. 47, pp. 98–115.

Hey, T., Tansley, S. and Tolle, K., eds. (2009) *The fourth paradigm: Data-intensive scientific discovery*, Redmond, Wash., Microsoft Research.

Hirsh, J. B., Kang, S. K. and Bodenhausen, G. V. (2012) 'Personalized persuasion: Tailoring persuasive appeals to recipients' personality traits', *Psychological science*, vol. 23, no. 6, pp. 578–581.

Hofacker, C. F., Malthouse, E. C. and Sultan, F. (2016) 'Big Data and consumer behavior: Imminent opportunities', *Journal of Consumer Marketing*, vol. 33, no. 2, pp. 89–97.

Hughes, J. and McDonagh, J. (2017) 'In defence of the case study methodology for research into strategy practice', *The Irish Journal of Management*, vol. 36, no. 2, p. 243.

ITU (2019) *Measuring digital development: Facts and figures*, International Telecommunication Union.

Jagadish, H. V. (2015) 'Big Data and Science: Myths and Reality', *Big Data Research*, vol. 2, no. 2, pp. 49–52.

James, W. (1897) *The Will to Believe, and Other Essays in Popular Philosophy*, London, UK, Longmans, Green & Co.

Jasanoff, S. (2006a) 'Afterword', in Jasanoff, S. (ed) *States of knowledge: The co-production of science and social order*, London, New York, Routledge, pp. 274–282.

Jasanoff, S. (2006b) 'Ordering Knowledge, Ordering Society', in Jasanoff, S. (ed) *States of knowledge: The co-production of science and social order*, London, New York, Routledge, pp. 13–46.

Jasanoff, S. (2006c) 'The idiom of co-production', in Jasanoff, S. (ed) *States of knowledge: The co-production of science and social order*, London, New York, Routledge, pp. 1–12.

Jasanoff, S. (2017) 'Virtual, visible, and actionable: Data assemblages and the sightlines of justice', *Big Data & Society*, vol. 4, no. 2, 1-15.

Jasanoff, S. and Kim, S.-H. (2009) 'Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea', *Minerva*, vol. 47, no. 2, pp. 119–146.

Johl, S. K. and Renganathan, S. (2010) ‘Strategies for Gaining Access in Doing Fieldwork: Reflection of two Researchers’, *The Electronic Journal of Business Research Methods*, vol. 8, no. 1, pp. 41–50.

Junqué de Fortuny, E., Martens, D. and Provost, F. (2013) ‘Predictive Modeling With Big Data: Is Bigger Really Better?’, *Big data*, vol. 1, no. 4, pp. 215–226.

Kambatla, K., Kollias, G., Kumar, V. and Grama, A. (2014) ‘Trends in big data analytics’, *Journal of Parallel and Distributed Computing*, vol. 74, no. 7, pp. 2561–2573.

Kaplan, R. M., Chambers, D. A. and Glasgow, R. E. (2014) ‘Big Data and Large Sample Size: A Cautionary Note on the Potential for Bias’, *Clinical and Translational Science*, vol. 7, no. 4, pp. 342–346 [Online]. DOI: 10.1111/cts.12178.

Kennedy, H. and Hill, R. L. (2017) ‘The Pleasure and Pain of Visualizing Data in Times of Data Power’, *Television & New Media*, vol. 18, no. 8, pp. 769–782.

Kennedy, H., Hill, R. L., Aiello, G. and Allen, W. (2016) ‘The work that visualisation conventions do’, *Information, Communication & Society*, vol. 19, no. 6, pp. 715–735.

Kennedy, H., Poell, T. and van Dijck, J. (2015) ‘Data and agency’, *Big Data & Society*, vol. 2, no. 2, 205395171562156.

Kerr, I. and Earle, J. (2013) *Prediction Preemption Presumption: How Big Data Threatens Big Picture Privacy*.

Khan, H. U. and Brock, V. F. (2017) ‘Are enterprises ready for big data analytics?: A survey-based approach’, *International Journal of Business Information Systems*, vol. 25, no. 2, p. 256.

King, N. (2004) ‘Using Interviews in Qualitative Research’, in Cassell, C. and Symon, G. (eds) *Essential Guide to Qualitative Methods in Organizational Research*, 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom, SAGE Publications Ltd, pp. 11–22.

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

Kitchin, R. (2015) ‘Big Data and Official Statistics: Opportunities, Challenges and Risks’, *Statistical Journal of the International Association of Official Statistics*, vol. 31, no. 3, pp. 471–481.

Kitchin, R. and Lauriault, T. P. (2018) ‘Toward Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work’, in Thatcher, J., Eckert, J. and Shears, A. (eds) *Thinking Big Data in Geography: New Regimes, New Research,* University of Nebraska Press, pp. 3–20.

Kitchin, R. and McArdle, G. (2016) ‘What makes Big Data, Big Data?: Exploring the ontological characteristics of 26 datasets’, *Big Data & Society*, vol. 3, no. 1, 205395171663113.

Kjellberg, H. and Helgesson, C.-F. (2016) ‘On the nature of markets and their practices’, *Marketing Theory*, vol. 7, no. 2, pp. 137–162.

Kosinski, M., Stillwell, D. and Graepel, T. (2013) ‘Private traits and attributes are predictable from digital records of human behavior’, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 15, pp. 5802–5805.

Krenn, K. (2017a) ‘Introduction: Markets and Classifications: Constructing Market Orders in the Digital Age. An Introduction’, *Historical Social Research*, vol. 42, no. 1, pp. 7–22.

Krenn, K. (2017b) ‘Segmented Intermediation: Advice Concepts in German Financial Services’, *Historical Social Research*, vol. 42, no. 1, pp. 123–151.

Kuckartz, U. (2014) *Mixed Methods: Methodologie, Forschungsdesigns und Analyseverfahren*, Wiesbaden, Springer Fachmedien Wiesbaden.

Lampel, J. and Meyer, A. D. (2008a) ‘Guest Editors' Introduction. Field-Configuring Events as Structuring Mechanisms: How Conferences, Ceremonies and Trade Shows Constitute New Technologies, Industries and Markets’, *Journal of Management Studies*, vol. 45, no. 6, pp. 1025–1035.

Lampel, J. and Meyer, A. D. (2008b) ‘Guest Editors’ Introduction: Field-Configuring Events as Structuring Mechanisms: How Conferences, Ceremonies, and Trade Shows

Constitute New Technologies, Industries, and Markets', *Journal of Management Studies*, vol. 45, no. 6, pp. 1025–1035.

Laney, D. (2001) *3D Data Management: Controlling Data Volume, Velocity, and Variety*, Meta Group, Application Delivery Strategies 949.

Lange, B., Power, D. and Suwala, L. (2014) 'Geographies of field-configuring events', *Zeitschrift für Wirtschaftsgeographie*, vol. 58, no. 4, pp. 187–201.

Latour, B. (1988) *Science in action: How to follow scientists and engineers through society*, Cambridge.

Latour, B. (1991) 'Technology is society made durable', in Law, J. (ed) *A sociology of monsters: Essays on power, technology and domination*, London, New York, Routledge, pp. 103–131.

Latour, B. (1993) *We have never been modern*, Cambridge, Mass., Harvard University Press.

Latour, B. (2000) *Pandora's hope: Essays on the reality of science studies*, 2nd edn, Cambridge, Mass. [u.a.], Harvard University Press.

Latour, B. (2005) *Reassembling the Social: An Introduction to Actor-Network-Theory*, New York, Oxford University Press.

Latour, B. (2016) 'Tarde's idea of quantification', in Candea, M. (ed) *The social after Gabriel Tarde: Debates and assessments*, 2nd edn, London, [England], New York, New York, Routledge, pp. 147–164.

Latour, B. and Woolgar, S. (1986) *Laboratory life: The construction of scientific facts*, 2nd edn, Princeton, N.J., Chichester, W. Sussex, Princeton University Press.

Lauer, J. (2010) 'The Good Consumer: Credit Reporting and the Invention of Financial Identity in the United States, 1840 - 1940', *Enterprise & Society*, vol. 11, no. 4, pp. 688–696.

Laughlin, P. (2014) 'Holistic customer insight as an engine of growth', *Journal of Direct, Data and Digital Marketing Practice*, vol. 16, no. 2, pp. 75–79.

Law, J. (2007) *Actor Network Theory and Material Semiotics* [Online]. Available at http://www.heterogeneities.net/publications/Law2007ANTandMaterialSemiotics.pdf.

Law, J. (2008) 'On Sociology and STS', *The Sociological Review*, vol. 56, no. 4, pp. 623–649.

Law, J. (2017) 'STS As Method', in Felt, U., Fouché, R., Miller, C. A. and Smith-Doerr, L. (eds) *Handbook of Science and Technology Studies*, 4th edn, Cambridge, Mass, London, UK, MIT Press, pp. 31–58.

Law, J. and Urry, J. (2004) 'Enacting the social', *Economy and Society*, vol. 33, no. 3, pp. 390–410.

Layton, R. (2016) 'Reframing marketing as a social science: A look back at the Special Session in Dublin', *Australasian Marketing Journal (AMJ)*, vol. 24, no. 3, pp. 241–243.

Lee, F., Bier, J., Christensen, J., Engelmann, L., Helgesson, C.-F. and Williams, R. (2019) 'Algorithms as folding: Reframing the analytical focus', *Big Data & Society*, vol. 6, no. 2, 205395171986381.

Lee, F. and Björklund Larsen, L. (2019) 'How should we theorize algorithms?: Five ideal types in analyzing algorithmic normativities', *Big Data & Society*, vol. 6, no. 2, 205395171986734.

Littig, B. (2009) 'Interviewing the Elite — Interviewing Experts: Is There a Difference?', in Bogner, A., Littig, B. and Menz, W. (eds) *Interviewing Experts*, London, Palgrave Macmillan UK, pp. 98–113.

Liu, S. Q. and Mattila, A. S. (2017) 'Airbnb: Online targeted advertising, sense of power, and consumer decisions', *International Journal of Hospitality Management*, vol. 60, pp. 33–41.

Luther, C. and Radovic, I. (2013) 'Perspectives on Privacy, Information Technology, and Company/Governmental Surveillance in Japan', *Surveillance & Society*, vol. 10, 3/4, pp. 263–275.

Lyon, D. (2014) 'Surveillance, Snowden, and Big Data: Capacities, consequences, critique', *Big Data & Society*, vol. 1, no. 2, 1-13.

MacKenzie, D. A. (2006) *An engine, not a camera: How financial models shape markets*, Cambridge, Mass, London., MIT.

Mager, A. (2012) 'Algorithm Ideology', *Information, Communication & Society*, vol. 15, no. 5, pp. 769–787.

Mager, A. (2017) 'Search engine imaginary: Visions and values in the co-production of search technology and Europe', *Social Studies of Science*, vol. 9, no. 2, 030631271667143.

Maguire, M. and Delahunt, B. (2017) 'Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars', *All Ireland Journal of Higher Education; Vol 9, No 3 (2017)*.

Malterud, K., Siersma, V. D. and Guassora, A. D. (2015) 'Sample Size in Qualitative Interview Studies: Guided by Information Power', *Qualitative health research*.

Mann, M. and Matzner, T. (2019) 'Challenging algorithmic profiling: The limits of data protection and anti-discrimination in responding to emergent discrimination', *Big Data & Society*, vol. 6, no. 2, 205395171989580.

Marshall, M. N. (1996) 'The key informant technique', *Family Practice*, vol. 13, no. 1, pp. 92–97.

Martens, D., Provost, F., Clark, J. and Fortuny, E. J. de (2016) 'Mining Massive Fine-Grained Behavior Data To Improve Predictive Analytics', *MIS Quarterly*, vol. 40, no. 4, pp. 869–888.

Marx, G. T. (2003) 'A Tack in the Shoe: Neutralizing and Resisting the New Surveillance', *Journal of Social Issues*, vol. 59, no. 2, pp. 369–390.

Marx, G. T. (2009) 'A Tack in the Shoe and Tacking off the Shoe: Neutralization and Counter-neutralization Dynamics', *Surveillance & Society*, vol. 6, no. 3, pp. 294–306.

Matz, S. C. and Netzer, O. (2017) 'Using Big Data as a window into consumers' psychology', *Current Opinion in Behavioral Sciences*, vol. 18, pp. 7–12.

Matzner, T. (2016) 'Beyond data as representation: The performativity of Big Data in surveillance', *Surveillance & Society*, vol. 14, no. 2, pp. 197–210 [Online]. Available at

http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/download/beyond_data/beyond.

Maxwell, J. (2009) 'Designing a Qualitative Study', in Bickman, L. and Rog, D. (eds) *The SAGE Handbook of Applied Social Research Methods*, 2455 Teller Road, Thousand Oaks California 91320 United States, SAGE Publications, Inc, pp. 214–253.

Mazzei, M. J. and Noble, D. (2017) 'Big data dreams: A framework for corporate strategy', *Business Horizons*, vol. 60, no. 3, pp. 405–414.

Mellet, K. and Beauvisage, T. (2020) 'Cookie monsters.: Anatomy of a digital market infrastructure', *Consumption Markets & Culture*, vol. 23, no. 2, pp. 110–129.

Mennicken, A. and Espeland, W. N. (2019) 'What's New with Numbers?: Sociological Approaches to the Study of Quantification', *Annual Review of Sociology*, vol. 45, no. 1, pp. 223–245.

Meuser, M. and Nagel, U. (2009) 'The Expert Interview and Changes in Knowledge Production', in Bogner, A., Littig, B. and Menz, W. (eds) *Interviewing Experts*, London, Palgrave Macmillan UK, pp. 17–42.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016) 'The ethics of algorithms: Mapping the debate', *Big Data & Society*, vol. 3, no. 2, pp. 1–21.

Mol, A. (1999) 'Ontological politics: A word and some questions', *The Sociological Review*, vol. 47, S1, pp. 74–89.

Mol, A. (2002) *The body multiple: Ontology in medical practice*, 2014th edn, Durham, London, Duke University Press.

Moon, K. and Blackman, D. (2014) 'A guide to understanding social science research for natural scientists', *Conservation biology : the journal of the Society for Conservation Biology*, vol. 28, no. 5, pp. 1167–1177.

Moran, J. (2014) *Marketing to the segment of one is closer than it seems* [Online], CMS Wire (Accessed 7 March 2020).

Morgan, D. L. (2007) ‘Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods’, *Journal of Mixed Methods Research*, vol. 1, no. 1, pp. 48–76.

Moro, S., Rita, P. and Vala, B. (2016) ‘Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach’, *Journal of Business Research*, vol. 69, no. 9, pp. 3341–3351.

Mosco, V. (2017) ‘Marx in the Cloud’, in Fuchs, C. and Mosco, V. (eds) *Marx in the age of digital capitalism*, Chicago, Haymarket Books, pp. 516–535.

Müller, H. and Hamm, U. (2014) ‘Stability of market segmentation with cluster analysis – A methodological approach’, *Food Quality and Preference*, vol. 34, pp. 70–78.

Netzer, O., Lemaire, A. and Herzenstein, M. (2019) ‘When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications’, *Journal of Marketing Research*, vol. 56, no. 6, pp. 960–980.

Neuman, W. L. (2014) *Social Research Methods: Qualitative and Quantitative Approaches*, 7th edn, Pearson.

Nowell, L. S., Norris, J. M., White, D. E. and Moules, N. J. (2017) ‘Thematic Analysis’, *International Journal of Qualitative Methods*, vol. 16, no. 1, 1-13.

Origgi, G. (2004) ‘Is Trust an Epistemological Notion?’, *Episteme*, vol. 1, no. 1, pp. 61–72.

Origgi, G. (2007) ‘What does it mean to trust in epistemic authority?’, in Pasquino, P. and Harris, P. (eds) *The Concept of Authority: A multidisciplinary approach: from epistemology to the social sciences*, Rome, Fondazione Adriano Olivetti, pp. 89–115.

Overdevest, C. (2011) ‘Towards a more pragmatic sociology of markets’, *Theory and Society*, vol. 40, no. 5, pp. 533–552.

Pasquale, F. (2015) *The black box society: The secret algorithms that control money and information*, Cambridge, Harvard University Press.

Peirce, C. S. (1878) ‘How to make our Ideas Clear’, *Popular Science Monthly*, vol. 12, pp. 286–302 [Online]. Available at http://www.peirce.org/writings/p119.html.

Peña Gangadharan, S. and Niklas, J. (2019) 'Decentering technology in discourse on discrimination', *Information, Communication & Society*, vol. 22, no. 7, pp. 882–899.

Pridmore, J. (2010) 'Reflexive marketing: The cultural circuit of loyalty programs', *Identity in the Information Society*, vol. 3, no. 3, pp. 565–581.

Pridmore, J. (2012) 'Consumer Surveillance: Context, perspectives and concerns in the personal information economy', in Ball, K., Haggerty, K. D. and Lyon, D. (eds) *Routledge handbook of surveillance studies*, London, Routledge, pp. 321–329.

Pridmore, J. and Hämäläinen, L. E. (2017) 'Market Segmentation in (In)Action: Marketing and 'Yet to Be Installed' Role of Big and Social Media Data', *Historical Social Research*, vol. 42, no. 1, pp. 103–122.

Pridmore, J. and Lyon, D. (2011) 'Marketing as Surveillance: Assembling Consumers as Brands', in Zwick, D. and Cayla, J. (eds) *Inside marketing: Practices, ideologies, devices*, Oxford, Oxford University Press, pp. 115–136.

Pridmore, J. and Zwick, D. (2011) 'Editorial - Marketing and the Rise of Commercial Consumer Surveillance', *Surveillance & Society*, vol. 8, no. 3, pp. 269–277 [Online]. Available at http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/download/4163/4165.

Quinn, L., Dibb, S., Simkin, L., Canhoto, A. and Analogbei, M. (2016) 'Troubled waters: The transformation of marketing in a digital world', *European Journal of Marketing*, vol. 50, no. 12, pp. 2103–2133.

Raghupathi, W. and Raghupathi, V. (2014) 'Big data analytics in healthcare: promise and potential', *Health Information Science and Systems*, vol. 2, no. 3, pp. 1–10.

Raguseo, E. (2018) 'Big data technologies: An empirical investigation on their adoption, benefits and risks for companies', *International Journal of Information Management*, vol. 38, no. 1, pp. 187–195.

Ritzer, G. (2015) 'Automating prosumption: The decline of the prosumer and the rise of the prosuming machines', *Journal of Consumer Culture*, vol. 15, no. 3, pp. 407–424.

Ritzer, G. and Jurgenson, N. (2010) ‘Production, Consumption, Prosumption: The nature of capitalism in the age of the digital 'prosumer'’, *Journal of Consumer Culture*, vol. 10, no. 1, pp. 13–36.

Roscoe, P. (2014) ‘‘Elephants can’t gallop’: Performativity, knowledge and power in the market for lay-investing’, *Journal of Marketing Management*, vol. 31, 1-2, pp. 193–218.

Roscoe, P. (2016) ‘Performativity Matters: Economic Description as a Moral Problem’, in Boldyrev, I. and Svetlova, E. (eds) *Enacting Dismal Science*, New York, Palgrave Macmillan US, pp. 131–150.

Roscoe, P. and Chillas, S. (2014) ‘The state of affairs: Critical performativity and the online dating industry’, *Organization*, vol. 21, no. 6, pp. 797–820.

Ruckenstein, M. and Granroth, J. (2020) ‘Algorithms, advertising and the intimacy of surveillance’, *Journal of Cultural Economy*, vol. 13, no. 1, pp. 12–24.

Saldana, J. (2013) *The Coding Manual for Qualitative Researchers*, 2nd edn, London, UK, SAGE Publications Ltd.

Sanders, C. and Condon, C. (2017) ‘Crime analysis and cognitive effects: The practice of policing through flows of data’, *Global Crime*, vol. 18, no. 3, pp. 237–255.

Sanders, C. B., Christensen, T. and Weston, C. (2015) ‘Constructing Crime in a Database: Big Data and the Mangle of Social Problems Work’, *Qualitative Sociology Review*, vol. 11, no. 2, pp. 180–195.

Sanders, C. B. and Sheptycki, J. (2017) ‘Policing, crime and ‘big data’; towards a critique of the moral economy of stochastic governance’, *Crime, Law and Social Change*, vol. 68, 1-2, pp. 1–15.

Saren, M. and Svensson, P. (2013) ‘Marketing’, in Alvesson, M., Bridgman, T. and Willmott, H. (eds) *The Oxford handbook of critical management studies*, Oxford, Oxford University Press, pp. 371–391.

Schneider, T. and Woolgar, S. (2012) ‘Technologies of ironic revelation: Enacting consumers in neuromarkets’, *Consumption Markets & Culture*, vol. 15, no. 2, pp. 169–189.

Schüßler, E., Grabher, G. and Müller-Seitz, G. (2015) 'Field-Configuring Events: Arenas for Innovation and Learning?', *Industry and Innovation*, vol. 22, no. 3, pp. 165–172.

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

Shareef, M. A., Dwivedi, Y. K. and Kumar, V. (2016) *Mobile Marketing Channel: Online Consumer Behavior*, Cham, Springer International Publishing.

Sheng, J., Amankwah-Amoah, J. and Wang, X. (2017) 'A multidisciplinary perspective of big data in management research', *International Journal of Production Economics*, vol. 191, pp. 97–112.

Sim, J., Saunders, B., Waterfield, J. and Kingstone, T. (2018) 'Can sample size in qualitative research be determined a priori?', *International Journal of Social Research Methodology*, vol. 21, no. 5, pp. 619–634.

Simakova, E. (2016; [2013]) *Marketing technologies: Corporate cultures and technological change*, London, New York, Routledge.

Singh, I. and Singh, S. (2017) 'Framework for Targeting High Value Customers and Potential Churn Customers in Telecom using Big Data Analytics', *International Journal of Education and Management Engineering*, vol. 7, no. 1, pp. 36–45.

Sivarajah, U., Kamal, M. M., Irani, Z. and Weerakkody, V. (2017) 'Critical analysis of Big Data challenges and analytical methods', *Journal of Business Research*, vol. 70, pp. 263–286.

Stake, R. E. (2005) *Multicase research methods: Step by step cross-case analysis*, New York, Guilford Press.

Sunderland, P. L. and Denny, R. M. (2011) 'Consumer Segmentation in Practice: An Ethnographic Account of Slippage', in Zwick, D. and Cayla, J. (eds) *Inside marketing: Practices, ideologies, devices*, Oxford, Oxford University Press, pp. 137–161.

Sunstein, C. R. (2015) 'Fifty Shades of Manipulation', *Journal of Marketing Behavior*, vol. 1, 3-4, pp. 213–244.

Sweeney, L. (2013) 'Discrimination in Online Ad Delivery', *SSRN Electronic Journal*.

Symons, J. and Alvarado, R. (2016) 'Can we trust Big Data?: Applying philosophy of science to software', *Big Data & Society*, vol. 3, no. 2, pp. 1–17.

Tadajewski, M. (2010a) 'Critical marketing studies: Logical empiricism, 'critical performativity' and marketing practice', *Marketing Theory*, vol. 10, no. 2, pp. 210–222.

Tadajewski, M. (2010b) 'Towards a history of critical marketing studies', *Journal of Marketing Management*, vol. 26, 9-10, pp. 773–824.

Tata, S. V., Prashar, S. and Gupta, S. (2020) 'An examination of the role of review valence and review source in varying consumption contexts on purchase decision', *Journal of Retailing and Consumer Services*, vol. 52.

Thomas, S. L., Nafus, D. and Sherman, J. (2018) 'Algorithms as fetish: Faith and possibility in algorithmic work', *Big Data & Society*, vol. 5, no. 1, 1-11.

Thompson, C. J. (2019) 'The 'big data' myth and the pitfalls of 'thick data' opportunism: On the need for a different ontology of markets and consumption', *Journal of Marketing Management*, vol. 35, 3-4, pp. 207–230.

Tien, J. M. (2013) 'Big Data: Unleashing information', *Journal of Systems Science and Systems Engineering*, vol. 22, no. 2, pp. 127–151.

Trusov, M., Ma, L. and Jamal, Z. (2016) 'Crumbs of the Cookie: User Profiling in Customer-Base Analysis and Behavioral Targeting', *Marketing Science*, vol. 35, no. 3, pp. 405–426.

Tzafilkou, K., Protogeros, N. and Yakinthos, C. (2014) 'Mouse Tracking for Web Marketing: Enhancing User Experience in Web Application Software by Measuring Self-Efficacy and Hesitation Levels', *International Journal of Strategic Innovative Marketing*, vol. 01.

Vaidhyanathan, S. (2012) *The googlization of everything: (and why we should worry)*, Berkeley, University of California Press.

van Dijck, J. (2014) 'Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology', *Surveillance & Society*, vol. 12, no. 2, pp. 197–208 [Online].

Available at http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/download/datafication/datafic.

van Otterlo, M. (2016) 'The Libraryness of Calculative Devices - Artificially intelligent librarians and their impact on information consumption', in Amoore, L. and Piotukh, V. (eds) *Algorithmic life: Calculative devices in the age of big data,* Abingdon, Oxon, New York, NY, Routledge, Taylor & Francis Group.

Venter, P., Wright, A. and Dibb, S. (2014) 'Performing market segmentation: A performative perspective', *Journal of Marketing Management*, vol. 31, 1-2, pp. 62–83.

Waddington, D. (2004) 'Participant Observation', in Cassell, C. and Symon, G. (eds) *Essential Guide to Qualitative Methods in Organizational Research,* 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom, SAGE Publications Ltd, pp. 154–164.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J.-f., Dubey, R. and Childe, S. J. (2017) 'Big data analytics and firm performance: Effects of dynamic capabilities', *Journal of Business Research*, vol. 70, pp. 356–365.

Wang, H., Xu, Z., Fujita, H. and Liu, S. (2016) 'Towards felicitous decision making: An overview on challenges and trends of Big Data', *Information Sciences*, 367-368, pp. 747–765.

Wang, T. (2013) 'Big data needs thick data', *Ethnography Matters*, vol. 13.

Wang, Y., Kung, L. and Byrd, T. A. (2018) 'Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations', *Technological Forecasting and Social Change*, vol. 126, pp. 3–13.

Wedel, M. and Kannan, P. K. (2016) 'Marketing Analytics for Data-Rich Environments', *Journal of Marketing*, vol. 80, no. 6, pp. 97–121.

Wertenbroch, K. (2015a) 'From the Editor: Manipulation and Marketing: The Elephant in the Room?', *Journal of Marketing Behavior*, vol. 1, 3-4, pp. 209–212.

Wertenbroch, K. (2015b) ‘From the Editor: Rational Choice as the Foundation of Behavioral Research in Marketing’, *Journal of Marketing Behavior*, vol. 1, no. 2, pp. 109–111.

Wheeler, S. C., Petty, R. E. and Bizer, G. Y. (2005) ‘Self-Schema Matching and Attitude Change: Situational and Dispositional Determinants of Message Elaboration’, *Journal of Consumer Research*, vol. 31, no. 4, pp. 787–797.

Wielki, J. (2013) ‘Implementation of the Big Data concept in organizations: Possibilities, impediments and challenges’, *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*, pp. 985–989.

Wilson, A. (2008) ‘Chapter 8 - Marketing research’, in Baker, M. J. and Hart, S. (eds) *The Marketing Book (Sixth Edition)*, Oxford, Butterworth-Heinemann, pp. 140–155.

Woolgar, S. (1991) ‘The Turn to Technology in Social Studies of Science’, *Science, Technology, & Human Values*, vol. 16, no. 1, pp. 20–50.

Woolgar, S. and Lezaun, J. (2013) ‘The wrong bin bag: A turn to ontology in science and technology studies?’, *Social Studies of Science*, vol. 43, no. 3, pp. 321–340.

Yao, Z., Sarlin, P., Eklund, T. and Back, B. (2014) ‘Combining visual customer segmentation and response modeling’, *Neural Computing and Applications*, vol. 25, no. 1, pp. 123–134.

Yin, R. K. (2014) *Case study research: Design and methods*, 5th edn, London, Sage Publication.

Zhang, H., Kong, F. and Wang, H. (2018) ‘Research on Power Big Data Processing Attribute Reduction Method Based on Cloud Computing Technology’, *IOP Conference Series: Materials Science and Engineering*, vol. 452, p. 42026.

Zuboff, S. (2015) ‘Big Other: Surveillance Capitalism and the Prospects of an Information Civilization’, *Journal of Information Technology*, no. 30, pp. 75–89.

Zwick, D. (2015) ‘Defending the Right Lines of Division: Ritzer's Prosumer Capitalism in the Age of Commercial Customer Surveillance and Big Data’, *The Sociological Quarterly*, vol. 56, no. 3, pp. 484–498.

Zwick, D. and Cayla, J. (2011) ‘Inside Marketing: Practices, Ideologies, Devices’, in Zwick, D. and Cayla, J. (eds) *Inside marketing: Practices, ideologies, devices,* Oxford, Oxford University Press, pp. 3–22.

Zwick, D. and Denegri Knott, J. (2009) ‘Manufacturing Customers: The database as new means of production’, *Journal of Consumer Culture*, vol. 9, no. 2, pp. 221–247.

Zwick, D. and Ozalp, Y. (2011) ‘Flipping the Neighborhood: Biopolitical Marketing as Value Creation for Condos and Lofts’, in Zwick, D. and Cayla, J. (eds) *Inside marketing: Practices, ideologies, devices,* Oxford, Oxford University Press, pp. 234–253.

Zwitter, A. (2014) ‘Big Data ethics’, *Big Data & Society*, vol. 1, no. 2, 1-6.

Appendix 1 – List of Industry Documents

Title	Issue	Date
Music:)ally Sandbox Music Marketing for the Digital Era	215	05.11.2018
Music:)ally Sandbox Music Marketing for the Digital Era	218	12.12.2018
Music:)ally Sandbox Music Marketing for the Digital Era	220	23.01.2019
Music:)ally Sandbox Music Marketing for the Digital Era	221	06.02.2019
Music:)ally Sandbox Music Marketing for the Digital Era	222	20.02.2019
Music:)ally Sandbox Music Marketing for the Digital Era	224	20.03.2019
Music:)ally Sandbox Music Marketing for the Digital Era	225	03.04.2019
Music:)ally Sandbox Music Marketing for the Digital Era	226	17.04.2019
Music:)ally Sandbox Music Marketing for the Digital Era	236	18.09.2019
Music:)ally The Report	419	19.12.2018
Music:)ally The Report	420	01.02.2019
Music:)ally The Report	421	27.02.2019
Music:)ally The Report	422	27.03.2019

Appendix 2 – Interview Guidelines

Hi, I am interested in your view and experience on data analytics for customer insight and engagement, as well as business insight. The questions will mainly revolve around the changes data analytics have brought in the area of marketing and customer engagement and how it effects the type of information and the flow of information, working routines and roles in these areas in organisations.

The interview will be entirely anonymised, there will be no identification of you, the interviewee but also the company and industry you are working in. Also, we don't aim to collect any strategic or competitive organisational information. You are at any time free to reject any question asked where you feel such information might come up.

In order to facilitate the analysis of the interviews, would it be ok for you if we record the interview digitally? The recording will be deleted after the transcription.

Introduction:

So, to start, could you tell me about your role in your organisation? What is that you do in your work, what current projects are you working on?

What is your vision for analytics in your organisation but also in a broader sense? Are these to do with the data themselves or are these more with the infrastructures?

Starting point (point of reference):

With your broader experience in market research, what have been the more traditional methods of gaining customer insight you have worked with, before starting on big data analytics projects? Please could you explain how that worked? How did the insight feed into the broader organisational strategy?

How did customer insight further feed into Customer Relationship Management processes, or marketing operations in general before working with big data analytics?

Changes in the organisation and operations:

How the practice of gaining customer insight in your experience changed upon the implementation of big data analytics?

How has this impacted different fields and areas in organisations: Have there been changes in the flow of information? Changes in team compositions and role changes? Has there been a demand for new skills? Have skills become redundant?

You have already shared your vision for analytics. To what extent has this met your expectations so far?

In what ways have operations changed? How has this impacted customer relationship management or marketing operations more broadly? Have there been changes in how customers are perceived and can be targeted?

Were there new or unforeseen challenges or difficulties that have emerged?

How are uncertainties – in the data or in the analysis – dealt with? How are they addressed and communicated with those working with the customer insight – i.e. the end-users?

Future trends/End:

What are the long-term plans in regard to data analytics at your organisation?

In the long term, where do you expect these changes will lead?

Thank you so much for your insight on data analytics and marketing, this was really interesting. Finally, do you have any questions or topics that you would like to discuss or address?

Appendix 3 – Ethical Approval Letter

University of St Andrews | FOUNDED 1413

University Teaching and Research Ethics Committee

05 March 2018

Dear Roger

Thank you for submitting your ethical application which was considered by the School of Management Ethics Committee meeting on 5th March 2018 when the following documents were reviewed:

1. Ethical Application Form
2. Participant Information Sheet
3. Consent Form
4. Debriefing Form

The School of Management Ethics Committee has been delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has granted this application ethical approval. The particulars relating to the approved project are as follows -

Approval Code:	MN13362	**Approved on:**	5th March 2018	**Approval Expiry:**	5th March 2023
Project Title:	Creating Postulated Consumers – Insight, Big Data and Targeting in Marketing				
Researcher(s):	Roger Ferdinand François von Laufenberg				
Supervisor(s):	Professor Kirstie Ball				

Approval is awarded for five years. Projects which have not commenced within two years of approval must be re-submitted for review by your School Ethics Committee. If you are unable to complete your research within the five year approval period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

If you make any changes to the project outlined in your approved ethical application form, you should inform your supervisor and seek advice on the ethical implications of those changes from the School Ethics Convener who may advise you to complete and submit an ethical amendment form for review.

Any adverse incident which occurs during the course of conducting your research must be reported immediately to the School Ethics Committee who will advise you on the appropriate action to be taken.

Approval is given on the understanding that you conduct your research as outlined in your application and in compliance with UTREC Guidelines and Policies (http://www.st-andrews.ac.uk/utrec/guidelinespolicies/). You are also advised to ensure that you procure and handle your research data within the provisions of the Data Provision Act 1998 and in accordance with any conditions of funding incumbent upon you.

Yours sincerely

Convener of the School Ethics Committee

cc Supervisor

School of Management Ethics Committee, The Gateway, North Haugh, St Andrews, Fife, KY16 9SS
management.ethics@st-andrews.ac.uk

The University of St Andrews is a charity registered in Scotland: No SC013532

Appendix 4 – Coding System

Code System	**Frequency**
Code System	4335
The role of data	0
Abstract Notion of Data	0
Big Data Hype / Buzzwords	67
Data and reputation	17
Data hoarding	10
Data is valuable	61
Data-driven	42
Improvement and innovation through data	34
What can data do?	16
Data as Information	0
Attributes of data	0
Sources of data	67
Data equals speed (or not)	56
Data is precise (or not)	34
Data permits easier visualisations	42
Data quality	20
The more the merrier	17
Data is valuable #2	57
Making data meaningful	50
What should data tell?	20
Why using data and what do we want to know?	77
Data for decision making	38
Data infrastructures	29

			Actual physical infrastructures					35
			Codes and Software as infrastructures					133
			Data as infrastructure					95
			Humans as infrastructures					75
			Infrastructuring					101
				Catalyst for infrastructuring				35
					Increase efficiency			28
					No more silos			26
	Producing and stabilising Knowledge							0
		What is (necessary for) good knowledge?						0
			Knowing the Business					21
			Knowledge creation process					0
				Operationalising problems and knowledge				93
					Defining KPIs			61
				Hypothesising				44
				Analysing / Researching				46
				Interpreting				38
			Prioritizing knowledge gaining processes					40
			Skill Sets					121
			The Technology					32
				Making the technology work				99
				Convincing Work				92
			Traditional insight vs data led insight					71
		Knowledge decision-makers						304
			Ideologies					81
			Levels of epistemic authority					66
			Learning to gain knowledge					28

		Communication		202
			Knowledge and insight	111
			Problems and issues	58
			Externally	33
		Kind of knowledge		101
			Exclusive knowledge	7
			Better Knowledge	52
			Having access to the 'right' knowledge	42
		Uncertainty of information		186
			Decisions of what to trust	100
			(Lack of) technical understanding	17
	Conceptualising / Segmenting Consumers			0
		Anticipating Consumers		205
			Personalising Consumers	35
			Targeting Consumers	45
			Impacting Consumers	47
			Consumer binding / loyalty	32
		Conceptualising / Segmenting consumers		171
			Imagined	51
			Postulated	84
		Consumer Insight		107
			Consumer Behaviour	63
			Consumer Demographics	16
	Difficulties and Uncertainties			0
		Data and Society		23
			Ethics	34
			Regulation	42

		Findings ways to make sense of the data		0
			Creating and using knowledge	22
			Under-reliance on data analytics	12
			Understanding how and where data will actually help	54
			Understanding the language	15
		Limitations and Inefficiencies		0
			Limitations & Inefficiencies of data analytics	45
			Limitations & Inefficiencies of traditional market and consumer methods	10
			Limitations & Inefficiencies of Marketing	4
		Uncertain infrastructures		0
			Lack of analytical tools	6
			Lack of communication between infrastructures	10
			Lack of skilled people / human infrastructures	9
			Organisational readiness	36
				Organisational structures	27
			Security	17
			The specificities of data infrastructure	61
	Practices			0
		The practices of data analytics / science		0
			Automation	31
			Answering data requests	4
			Supporting	3
			Modelling	22
			Data governance	40
			Analysing	7
			Communicating	6
			Consulting	12

			Experimenting		3
			Integrating		8
			Interpreting		2
			Measuring		79
				Segmenting	8
				Conceptualising	8
				Categorising	17
				Operationalisation	30
			Predicting		21
			Profiling		10
			Reporting		15
			Tracking		10
			Translating		12
			Visualising		7
		The practices of marketing			0
			Creativity		1
			Social Media Marketing		2
			Product visualisation / display		2
			Nudging		7
			Customer Engagement		11
			Experimenting		1
			Customer experience		15
			Personalising		29
			Targeting		35
			Segmenting		21
				Directive & actionable	6
				Illustrative	7

			Retaining		2
			Storytelling / PR		16
				Streamlining	5
			Content Creation		7
		The practices of insight management			0
			Sourcing Data or insight		19
			Disseminating		4
			Evaluating		6
			Pricing		8
			Aligning		15
			Operationalising		13
			Trial and error / creative testing / A-B Testing		23
			Researching		10
				Qualitative research methods	7
			Communicating		6
			Interpreting		13
			Consulting		34
				Translating	10
				Transforming	1