Università degli Studi di Cagliari

PHD DEGREE

Cycle XXXIII

WELCOME TO DIGITAL TRANSFORMATION ERA: FROM PROOF-OF-CONCEPT TO BIG DATA INSIGHTS CREATION

Scientific Disciplinary Sector(s)

SECS P-10: Business Organization

PhD Student: Maryia Zaitsava

Coordinator of the PhD Programme Prof. Vittorio Pelligra

Supervisor Prof. Maria Chiara Di Guardo

Final exam. Academic Year 2020 – 2021

Thesis defense: January-February 2020 Session

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, without her I would have not made it. Maria Chiara Di Guardo was guiding me at all stages of the research and sharing her tacit and tangible expertise throughout three years of Ph.D. starting from the research questions defining to pivoting the whole research without remorse. Not only her professional competencies are outstanding, but also Chiara is a role model for me as a woman in the research. She kept me focused on the research during my ups and downs.

I also would like to thank Elona Marku for her intellectual generosity. Elona was much more than a member of the research group I was involved in and co-author of the research projects. She was a mentor and a constructive critic that pushes you always to make a bit more effort than you think you can do. I am lucky to have benefitted from her precise thinking and encouraging mentoring and to develop my research thinking and methods further.

I would like to say a special thank you to the Liverpool John Moores University and specifically to David Bryde and Azar Shahgholian for their welcoming hosting me for the research period in Liverpool city. I appreciate very much the highest level of open and democratic research at this University and the highest knowledge level.

I am particularly grateful to the transportation authorities in the biggest UK cities (I cannot disclose them here) for allowing me to participate in designing and implementing their strategic changes. This became the ground for the whole thesis. I am excited to see how real business is ready to support the research.

Also, I would like to thank the CEO of Red Ninja technology studio and the CEO of iSensing technology company Lee Omar and Harvey Beilinsohn who were generously engaged in the research project in the UK and shared their knowledge, professional contacts, and insights. Thanks to them, I was able to accomplish the research that has great practical implications.

Last but not least, I would like to thank my family, mum and husband, who were supporting me emotionally throughout the three years of Ph.D. research. My mum even without having a clue of what is Digital Transformation was trying to listen and understand the contributions and importance of my research. My husband was cooking delicious dinners and lunches during a time of the submission deadlines and giving me all the necessary time for the research.

INTRODUCTION

Digital transformation (DT) is no longer an optional strategic priority, but the direction for managers of traditional firms that their success is built in the pre-digital era (Sebastian et al., 2017; Hess et al., 2016; Hanelt et al., 2020). Indeed, past success does not guarantee future success for traditional companies in this new era; businesses need to change to leverage digital era opportunities (Sebastian et al., 2017). With all hype around DT opportunities, it is rather a highly complex challenge that affects many or all segments of a firm (Hess et al., 2016) and more so at the early stages of DT.

Firms often struggle to understand the exact potential impact and commercial benefits of DT for their own firms due to its complexities and heavy technological nature (Parviainen et al., 2017). Thus, firms at the early stage of DT face the challenge of choosing among a big variety of existing and emerging technologies on the market (Kaplan and Tripsas, 2008), neglecting technological uncertainty (Kretschmer and Khashabi, 2020; Bhattacherjee and Premkumar, 2004; Ragatz et al., 2002), navigating through the technological solutions ocean and avoiding hype-driven decisions (Bohnsack et al., 2018), recognizing the commercial potentiality of Digital Transformation while being technology competence-less (Matt et al., 2015; Kaplan and Tripsas, 2008). With this respect, the phase preceding any adoption or rejection of a new DT initiative and aiming at the first meeting and proving feasibility and commercial opportunities becomes increasingly important (Parviainen et al., 2017).

Among challenges that the early stage of DT brings to firms, one is especially salient — the lack of tools and empirical-based conceptualization specialized on DT (Hanelt et al., 2020) that can help to overcome issues and technological uncertainty at the early stages of DT Therefore, in the thesis, I looked at the three different particular aspects of the earliest stage of

DT that, despite being well-known and established concepts, suffer from the lack of evidence base or one-dimensional view on the phenomenon.

Firstly, while proof-of-concept (PoC) is very often used to test a new concept at the early stages of innovation in a Digital Transformation (Fosso Wamba and Boeck, 2008; Papadopoulos et al., 2017; Kumar and Terpstra, 2004; Chemla and Tinn, 2020), surprisingly is not provided even with the definition in the technology innovation literature. An understanding of PoC is generalized to the test of an idea at the early stages of idea or technology development or adoption (Kumar and Terpstra, 2004; Fosso Wamba and Boeck, 2008; Papadopoulos et al., 2017; Chemla and Tinn, 2020). This explanation is not enough to fully capture the phenomenon of PoC that deals with all risks and opportunities of DT at the forefront (Kretschmer and Khashabi, 2020). Indeed, uncertainties, such as technology foresight, innovation, or adoption, accompany PoC projects. With this respect, what influences the dynamics of PoC and how knowledge on new technology is acquired and used to ultimately prove the concept is the black box for managers and researchers.

Secondly, although data-driven decision-making (DDD) is one of the most hyped concepts of DT (Herterich et al., 2016; Bohnsack et al., 2018; Ghasemaghaei et al., 2018) with the exception of the science fiction realm, a decision-making process that is completely driven by data is still very far from expectations (Gupta et al., 2018; Akter et al., 2019). Specifically, in the situation when DDD is focused on "supportive" functions of data algorithms, humans are still the main decision-makers (Raisch and Krakowski, 2020). This creates a particularly vulnerable to cognitive managerial biases context. Indeed, it is believed that human-related factors (intuitions, subjective opinions, and cognition biases) can influence data-based decisions (Kattel et al., 2019). Moreover, algorithms can be the subject of biases themselves. The biases can even leave out of consideration data-based insights (Mikalef et al., 2018; Surbakti et al., 2020; Hodgkinson & Healey, 2011). While in the existing research the two sides of the DDD are discussed separately, there is a lack of understanding of how the two agents (cognition and data) act together when first met or what are pitfalls when a firm firstly combine cognition and data.

Thirdly, it has been witnessed that the modern data-driven mode of decision-making is outperforming the classic intuitive school. However, the foregoing DDD step, namely, the creation of Big Data insights step, has been covered mainly through the heavy technological lens of Big Data Analytics (Sivarajah et al., 2017; Van Rijmenam et al., 2019; Dremel et al., 2020). Thus, not only Big Data insights creation is a black box for managers with no data analytics expertise (Court, 2015; Yadegaridehkordi et al., 2018; Zhang et al., 2019), but also purely Big Data analytics lens complicates the understanding of the usefulness of Big Data for managers (Court, 2015; Yadegaridehkordi et al., 2018; Zhang et al., 2019).

Following the overarching purpose of contributing to the building of empirical and conceptual evidence base on the early stage of DT phenomena, the thesis comprises three related objectives. Each objective is addressed by conducting independent research using comparative methods. The thesis consists of three essays and is structured as follows:

1. The first essay aims at opening the black box and disentangling technological uncertainty at the level of PoC. Specifically, I explore PoC for Digital Transformation when firms face the challenge of proving new technology potentiality under uncertainty and with low technology awareness. I use a qualitative case study methodology to capture granular details on how four cognitive biases, namely, herding, overconfidence, confirmation bias, and information avoidance, drive the perception of the technology potentiality. I explore the degrees of

technology awareness that developed as PoC is run by applying quantitative text mining method of the latent Dirichlet allocation.

2. The second essay aims at developing a better understanding of the cognitive and data biases and how they influence trust in data when DDD is first introduced. Based on an ethnographic methodology, I explore the case of a Commercialization team of the traditional transport authority in the North of the UK applying data to drive strategic decisions for the first time. Specifically, I investigate how managerial cognition and data interact when firstly met, focusing on how both cognition and data influence trust in data. Further, I couple qualitative data analysis methods with the new text mining quantitative analysis method of Sentiment analysis to derive the degree of trust in data and get the natural and real-time reaction of the firm on data and understand better the cognitive aspects.

3. The third essay investigates the core mechanisms of Big Data insights creation and disentangles it on the level of Big Data dimensions. Based on a participatory observation methodology I explore the pilot Active Travel Insights project launched in four European cities and specifically the role of Big Data dimensions and their characteristics in generating insights by conceptualizing Big Data dimensions in effects and building a matrix that explains the mechanism of different insights creation.

1. (DIS)PROOF-OF-CONCEPT: MANAGING TECHNOLOGY UNCERTAINTY IN THE DIGITAL TRANSFORMATION ERA

ABSTRACT

Proof-of-concept (PoC) is often used to prove a new concept at the early stages of innovation in a Digital Transformation. Despite the belief that PoC helps to overcome the most common challenges of Digital Transformation, innovation management research on PoC is still in its infancy. The present paper aims at disentangling technological uncertainty at the level of PoC. Using a qualitative case study, I explore PoC for Digital Transformation when firms face the challenge of proving new technology potentiality under uncertainty and with low technology awareness. The research allowed to open up the mechanism of how *perceived technology potentiality* changes during PoC due to the cognitive biases that affect it. Moreover, I found that *technology awareness* evolves step-wise from *borrowed* and *minimum acquired* to *enhanced technology awareness* as the firm accumulates knowledge on the technology. These two constructs directly influence the dynamics of PoC. This research makes several contributions to technology innovation literature. First, the work introduces the definition and understanding of PoC and its distinct constructs. Further, the research enlarges understanding of the existing cognitive biases and derived specific uncertainty situations during PoC when those biases may occur. The paper also introduced the notion of technology awareness degrees in the PoC phase. Finally, the research offers practical implications for innovation managers in the context of PoC for Digital Transformation.

Keywords: proof-of-concept, technology uncertainty, cognitive biases, technology awareness, digital transformation

1.1. INTRODUCTION

It is widely accepted that Digital Transformation is a critical component of strategic advantage in the long run (Cennamo et al., 2020), but how to achieve an effective Digital Transformation is not fully clarified and consensus on solutions is yet to be reached. In this context, proof-of-concept (PoC) is very often used to test a new concept at the early stages of innovation in a Digital Transformation (Fosso Wamba and Boeck, 2008; Papadopoulos et al., 2017; Kumar and Terpstra, 2004; Chemla and Tinn, 2020). At the first level, PoC is a rather simple approach that focuses on the two main aspects: assessing opportunities and learning practical aspects of new technology.

It is thus interesting to note that despite the growing role of PoC in implementing new ideas or new features in the existing system of most organizations involved in Digital Transformation, with the idea that PoC helps to overcome the most common challenges of this process, innovation management research on PoC is still at its infancy. The literature does not even provide a clear definition of the phenomenon and as PoC grows into a trend, new problems are emerging. Building a PoC needs a reasonable amount of time and effort for the development and it is subject to special uncertainty and risks, but for researchers and practitioners, it is still a black box.

The present paper aims to contribute to the technology innovation research and recent in Digital Transformation research by opening the black box and disentangling technological uncertainty at the level of PoC. Using a qualitative case study, the author explores PoC for Digital Transformation when firms face the challenge of proving new technology potentiality under uncertainty and with low technology awareness.

Specifically, the research examines the case of one of the leading transportation firms in North West England looking to employ external technology to improve its advertising revenue stream. In a typical Digital Transformation context, the firm runs a PoC specifically focused on exploiting sensors and Big Data technology to renew their marginalized advertising revenue channel. I chose the case as it is the representative situation for most traditional industries when a firm with low technology awareness decides to run PoC to prove the feasibility of new technology and assess its potentiality for the future digital shift.

The essay captured granular details on how four specific cognitive biases, namely: herding, overconfidence, confirmation bias, and information avoidance, driving what the author labeled as *perceived technology potentiality*. Moreover, it delineated three degrees of *technology awareness* that grow as a PoC is run, namely: *borrowed technology awareness*, *minimum acquired technology awareness*, and *enhanced technology awareness*.

The research allowed to open up the mechanism of how perceived technology potentiality changes during PoC due to the cognitive biases that affect it. Moreover, it revealed that technology awareness evolves step-wise from borrowed and minimum acquired technology awareness to enhanced technology awareness as the firm accumulates knowledge on the technology. These two constructs directly influence the dynamics of PoC.

This paper makes several contributions to existing literature. First, it suggests that PoC is especially vulnerable to cognitive biases occurrence. The author conceptualizes why this phase of technology innovation and testing is difficult, and how it was the subject of firstly over positive and later, negative attitude. Secondly, the research captured a specific relationship between cognition and technology awareness that designs the dynamic of PoC overall. In situations of especially inflated uncertainty and complexity, both cognition (Busenitz and

Barney, 1997) and technology awareness shape decision-making within PoC. Finally, the research offers practical implications for innovation managers in the context of PoC for digital transformation.

The paper is structured as follows. Section 1.2 provides a definition of PoC and discusses research on cognitive biases under technology uncertainty; section 1.3 describes the study's action research methodology. Section 1.4 summarizes the findings from the case study and the PoC dynamics. Section 1.5 discusses the contribution of the research literature as well as the implications for innovation management and provides concluding remarks.

1.2. THEORETICAL BACKGROUND

1.2.1. Proof-of-Concept

It is recommended to run PoC before technological advancements are attempted (Parviainen et al., 2017). Although PoC is a very intuitive term and a fair amount of case studies presenting PoC results are carried out (Athilingam et al., 2018; Kasurinen et al., 2018; Airehrour et al., 2019; Rosales-Morales et al., 2020), surprisingly, technology innovation literature does not provide its definition. An understanding of PoC is generalized to the test of an idea at the early stages of development or adoption (Kumar and Terpstra, 2004; Fosso Wamba and Boeck, 2008; Papadopoulos et al., 2017; Chemla and Tinn, 2020). However, this explanation is not enough to fully capture the phenomenon of PoC.

Therefore, the author borrowed an understanding of PoC from the philosophy of science field that suggests, PoC is the research instrument aimed at proving the feasibility of a new concept in practice and increasing practical knowledge on the new concept (Kendig, 2016).

In the Digital Transformation context, when firms are too positive and confident in estimating their necessity in digital shift, it can harm organizations and lead to misaligned expectations and fiasco in reaching planned results (Himmelstein et al., 2010; Andriole, 2017; Bohnsack et al., 2018). Not only face firms the challenge of choosing among a wide variety of existing and emerging technologies on the market and recognizing technology potentiality at the very early stage of its adoption but also firms are under pressure to better exploit technology being first to use the technology or being technology competence-less (Matt et al., 2015). Additionally, some firms might face the risk caused by the outside technology employed within the existing business structure (Stock and Tatikonda, 2008).

With this respect, applying the general understanding of PoC to the Digital Transformation context, the author defines PoC as *the managerial tool aimed at proving the feasibility of a new idea or technology in practice and developing technology awareness*. In the attempt to distinguish what PoC is not, the author highlights the main differentiating aspects between PoC and Minimum Viable Product (MVP) that is widely used in testing new technology and is often mixed with PoC (see Table 1).

	Ultimate goal	***Activity type***	***Knowledge types***	***Directionality of knowledge paths***	***Tactics***
PoC	Prove the feasibility of an idea or technology in practice	Complete	Developing technology awareness	Inside-in Outside-in	Prototype, Experiment, Simulation, Assessment Etc.
MVP	Test business model and customers' willingness to pay (Camuffo et al., 2020)	Iterative	Getting knowledge of customers and market needs (Ries, 2011; 2017)	Outside-in and Inside-out	

Table 1. Differences between PoC and MVP

Specifically, although PoC might use the same testing activities and tactics as MVP, such as experiment, prototyping, simulation, assessment, and others (Horton and Radcliffe, 1995; Fosso Wamba and Boeck, 2008; Teare et al., 2014; Jakkhupan et al., 2011; Papadopoulos et al., 2017; Chemla and Tinn, 2020), these two are not the same in terms of ultimate goals. Thus, PoC aims at proving an idea or technology in practice, while MVP has the main goal to test business model viability (Ries, 2011; 2017) and customers' willingness to pay (Camuffo et al., 2020). Moreover, while PoC is a complete type of activity as it has a clear beginning and the end (an idea is either proved or not), MVP aims at getting knowledge on the market iteratively presenting and getting feedback on a new changed business model (Ries, 2011; Eisenmann, Ries, and Dillard, 2012).

The second distinguishing aspect is technology awareness development. Firstly, the types of knowledge created differ. PoC is focused on developing technology awareness by moving from general to field-specific technology knowledge to use it later for Digital Transformation activity or similar cases (Kendig, 2016). Specifically, technology awareness qualitatively changes and moves from no technology knowledge level through acquiring basic to specific technology knowledge level (Bohn et al. 2005; Roca et al., 2017). While for MVP growing technology knowledge is not the priority, as it is focused on understanding the viability of the business model and market needs (Ries, 2011; Blank, 2013; Contigiani and Levinthal, 2019). Secondly, the directionality of knowledge paths is distinct too. Thus, PoC is concentrated on developing practical knowledge on technology within a firm, the knowledge path, therefore, is one-directional and aims at bringing from outside and developing knowledge inside. At the same time, the firm might accumulate knowledge differently. Thus, taking into consideration the Digital Transformation context, most often, firms deal with experimenting and proving an external technological innovation. Therefore, they firstly exploit technology knowledge developed by others (Pérez-Luño et al., 2011; Martínez-Román et al., 2020) and accumulate technology awareness by imitation at first (Damanpour and Wischnevsky, 2006). Exemplarily, technology awareness development might be supported by bringing in technology suppliers' or internal technology developers' knowledge (Ragatz et al., 2002; Fossas-Olalla et al., 2015). Alternatively, in the case of proving the potentiality of in-house developed technology, "learning by trying" is a common approach to accumulate technology awareness (Fleck, 1994). Thus, technology knowledge development happens by implementing small changes and modifications to make a new technology work (Roca et al., 2017). Comparing with MVP, that has a two-directional path – outside-in, from customers to a firm and inside-out, from a firm to customers (Reis, 2011; 2017).

1.2.2. Cognitive Biases Under Uncertainty

Technology uncertainty accompanies not only Digital Transformation (Kretschmer and Khashabi, 2020) but also PoC, as it deals with all risks and opportunities of Digital Transformation at the forefront. Specifically, technological uncertainty reflects the high degree of complexity, rapid change, or newness of a technology to test or adopt inside the existing business structure (Ragatz et al., 2002). Some of the PoC steps are characterized by uncertainty, including the technology choice, estimating, and proving feasibility while being technology competence-less. Translated into the context of technology innovation and new technology testing, these uncertainties are related to technology innovation, technology foresight, and technology adoption.

As the literature suggests, individuals tend to use subjective opinions or cognitive biases as the response to uncertainty (Kahneman and Tversky, 1977; Stubbart, 1989; Busenitz and

Barney, 1997; Haselton et al., 2005; Zhang and Cueto, 2017). Moreover, recent research suggests that people tend to rely on simplified strategies such as supporting general beliefs under the complexity and uncertainty of Digital Transformation (Solber et al., 2020). Kaplan and Tripsas (2008) emphasized the role of cognition for established firms in technology innovation and change context. Therefore, PoC that deals with technological uncertainty elements is especially vulnerable to cognitive biases too. Indeed, attitudes and perceptions drive technology adoption and use and trigger changes of beliefs, specifically at the early stages of technology use (Bhattacherjee and Premkumar, 2004). Further, technological forecasting or technological foresight focused on predicting the potentiality and characteristics of useful machines or techniques (Martino, 2003) or to formulate technology options or trends can be influenced by cognitive biases too (Bonaccorsi et al., 2020).

In particular, exiting studies showed that herd behavior is especially common for the technology adoption context Geroski, 2000; Duan and Whinston, 2009). Recently, Bonaccorsi et al. (2020) specified the role of overconfidence in the technology foresight context. Moreover, confirmation and information avoidance biases are well-known, especially while testing a new technology (Leventhal et al., 1994; Remencius et al., 2016) and mitigating uncertainty contexts (Lallement et al.,2020). Therefore, in this research, the author focus on four cognitive biases most often occurring in technology innovation uncertainty contexts, namely, herd behavior, overconfidence, confirmation bias, and information avoidance bias (Klayman and Ha, 1989; Banerjee, 1992; Geroski, 2000; Gaba and Terlaak, 2013; Gudmundsson and Lechner, 2013; Zhang and Cueto, 2017; Bacon-Gerasymenko and Eggers, 2019).

Herding or herd behavior generally refers to the phenomenon when individuals tend to mimic the actions of others (Banerjee, 1992; Ding and Li, 2019). The cognitive bias leads to a situation where a decision-maker tends to neglect their information and follow or imitate the actions of his/her fore-runner instead of making independent decisions (Banerjee, 1992; Sgroi, 2003). Decision-makers can be influenced by herding even when their prognosis suggests that the actions of others may be wrong choices, thus making inefficient choices (Scharfstein and Stein, 1990; Banerjee, 1992; Devenow and Welch, 1996; Barreto and Baden-Fuller, 2006). Furthermore, information cascade theory **(**Welch, 1992; Banerjee, 1992; Bikhchandani et al., 1992) suggests that the occurrence of herding requires two conditions, namely, uncertainty and the observed repeatable actions of other individuals (Banerjee, 1992; Ding and Li, 2019). When both conditions are satisfied, individuals observing one another typically follow the same action.

Herd behavior is recognized to have a significant influence on managers' behavior in the context of technology adoption when an alternative in adopting new technologies or business models exists (Bhattacherjee and Premkumar, 2004; Sun, 2013). Herding works as the underlying mechanism that leads later technology adopters to reduced willingness to invest in choosing between technologies (Geroski, 2000; Duan and Whinston, 2009). According to Geroski (2000), if one technology is proved by others to work better than the existing technology, firms have no motive to invest in testing other technology. Finally, firms can use herd behavior to learn from others in experiential learning (Gaba and Terlaak, 2013). Exemplarily, Lieberman and Asaba (2006) stated that when a new industry or commercial area appears, herd and experiential learning work together, and firms learn by concluding from the behavior of others. Research suggests that herding is not a simple reflexive copying but rather a mindful process that requires proper interpretation efforts (Strang and Still, 2006; Lounsbury, 2008; Gaba and Terlaak, 2013).

Overconfidence is a bias that results in perceiving a subjective certainty higher than the objective facts (Busenitz, 1999; Bernardo et al., 2001; Gudmundsson and Lechner, 2013; Zhang and Cueto, 2017). Among different tactics of overconfidence, Moore (2008) distinguished three facets that have been largely discussed as the ones the most frequently occurring when personal judgments need to be made. *Overestimation* deals with the bias of being too confident in estimating one's capabilities, performance, or chance of success. Overplacement reflects the overconfidence of being better than others; finally, overprecision affects one's beliefs in being accurate in estimates. Despite the variety of types, one common feature characterizes all overconfidence types - the bias affects the individuals' ability to objectively evaluate the degree of knowledge they have on the phenomenon their judgment is requested (Bonaccorsi et al., 2020). The recent research of Bonaccorsi et al. (2020) suggests that overconfidence occurs at different levels of technology foresight, such as assessing technology trends or technology options when personal estimates are the source of knowledge, but overall uncertainty is high. In this way, the situation when decision-makers being not experts in new technology are requested to foresight technology potentiality and give their judgments can be especially harmful (Kuusi et al., 2015).

Confirmation bias results from the effect of sticking to the goals or beliefs and it works more as a filtering process that makes individuals search for confirmation of prior beliefs rather than confronting information (Klayman and Ha, 1989; Lallement et al., 2020). Klayman and Ha (1987) discovered that individuals tend to test the hypotheses that have the best chance to verify existing convictions and not falsify them. Indeed, confirmation bias shows more substantial representation when individuals hold prior positive beliefs rather than negative ones. In practice, decision-makers tend to seek affirmation and support of a choice they have already made

(Klayman and Ha, 1989). Moreover, individuals are more likely to replicate and test hypotheses that are proved to be successful; the so-called “positive test strategy” does not necessarily have a negative effect (Klayman and Ha, 1987). It is proved that confirmation bias or positive test strategy could be a good heuristic for discovering the truth or errors of a hypothesis under realistic conditions (Klayman and Ha, 1987). However, it can lead to inefficiencies too. Confirmation bias can be a trap affecting software testing (Leventhal et al., 1994; Remencius et al., 2016). For instance, developers are more likely to test software to make their product work rather than challenge their code.

Information avoidance bias is characterized by non-acceptance of information, even when it could improve decision-making (Sweeny et al., 2010; Golman et al., 2017; Lallement et al., 2020) and stick to the beliefs that make individuals comfortable at least until they suffer the consequences of acting on these beliefs (Zhang and Cueto, 2017). The underlying mechanism of a deliberate choice to accept positive evidence and to avoid and neglect neutral or negative evidence is explained by selective information acceptance (Golman et al., 2017; Zhang and Cueto, 2017). Exemplarily, managers avoid recognizing arguments that contradict their prior beliefs and decisions, even when such arguments could help them avoid implementing measures that are insufficiently substantiated (Zaltman, 1983; Deshpande and Kohli, 1989; Schulz-Hardt et al., 2000). Among different tactics to avoid information, the most straightforward is *physical avoidance*, when individuals choose simply not to see information (Golman et al., 2017). However, there are other tactics, such as *inattention*, described as not focusing attention on information individuals already have (Golman et al., 2017). Thus, Sims (2003) suggests that attention as a resource can be allocated in the most efficient for a decision-maker way, therefore, presupposing rational inattention to some pieces of information (Sallee, 2014; Caplin and Dean, 2015). Moreover, individuals interpret evidence in the most comfortable way to support their beliefs and avoid results that challenge assumptions that they hold or would like to keep. In this way, managers may not draw obvious conclusions from data (Golman et al., 2017).

Concluding, PoC is especially of interest for firms willing to digitally transform, however, the technology innovation literature does not provide an answer to what exactly happens inside the black box of PoC when confirmative testing is run under technological uncertainty and with low technology awareness. The existing classical research on cognitive biases and technological awareness and analyzed above is a good ground to synthesize it with the very recent empirical experience presented below to develop a focused picture of how cognitive biases act and technology awareness change inside PoC.

1.3. METHODOLOGY

As the research background revealed the novelty of the PoC phenomenon is observed, therefore, the methodological solution is to explore the phenomenon in a real setting and its dynamics (Eisenhardt and Graebner, 2007). The present research employs a qualitative case study methodology as the overarching one to explore the PoC in its complexity in a real context and to investigate the *"how"* aspects of cognitive biases and technology awareness within the PoC context (Lee et al., 2007; Yin, 2009). Moreover, the inductive approach of the research was prompted by the pioneering and phenomenon-driven research topic (Eisenhardt and Graebner, 2007). Specifically, the author adopted a participatory observation as the primary method, as it is recognized effective to assess the antecedents and consequences of the behavior within a particular context (Carr et al., 2008; Bryman, 2012). The participatory observation that presupposes involvement in the phenomenon under investigation in comparison with just participant observation mitigates the drawbacks of the formalized communication and enables spotting alternative truths and to understand complex relationships (Clark et al., 2009) and spot the most natural reactions of PoC participants. The author complemented participatory observation by running a general inductive approach for qualitative data analysis (Thomas, 2006).

1.3.1. Case Description

In this paper, the author examines the case of PoC run by a firm willing to initiate the Digital Transformation path by employing external technology. Specifically, one of the leading transportation firms in North West England was focused on exploiting sensors and Big Data technology to renew their advertising revenue channel that was marginalized so far. Specifically, due to historical contracts with customers (advertising agencies and companies), advertising was losing more than one mln pounds per year. The identity of the firm is not revealed for privacy reasons. Henceforth, the author will refer to the transportation company as TC. Being an experienced firm in the transportation sector and advertising services, TC has not worked with Big Data. The firm was profit-oriented in its need to transform advertising services and wanted to capture the value of the digital shift as much as possible. The general objective of PoC was to understand if the specific sensor technology was able to produce reliable Big Data on pedestrian behavior to inform decisions and renew the ads revenue stream. Three more specific schemes to prove were: 1) possibility to support outdoor advertising agencies working with the firm to plan effective ads campaigns; 2) the ability of the technology to help to diversify services portfolio and build smarter advertising revenue streams; 3) ability to optimize and commercialize the company's spaces. The author chose the case as it is the representative situation for the majority of traditional industries when a firm with low technology awareness decides to run PoC to prove the feasibility of a new for the firm technology and assess its potentiality for the future digital shift.

There were two main parties involved in the PoC: the Commercialization division of the transportation company and the Big Data technology provider. The commercialization division was responsible for running PoC. The division was new (two years), and it was responsible for the commercialization of projects that lay apart from the main transport activities, such as outdoor advertising services delivery. Only two specialists were involved in the PoC project, the third specialist was presented at the first planning meeting only. The technology firm was a two years old startup using Wi-Fi sensors to generate Big Data about pedestrians. The main startup specialization was the transportation industry with a few projects in other spheres of cities' services.

As the Commercialization division almost did not have experience working with Big Data, the whole technical and strategic assistance was provided by the data technology supplier company. Thus, plans, strategy, KPIs, goals were developed by the technology provider. PoC sensor network was installed across an indicative study area; it was used to capture real-time data and examine historical data (within the project duration) on pedestrian movement and dwell-time, including origin/ destination, journey time statistics, footfall, duration. The technology solution was Wi-Fi sensors that search for Wi-Fi signals from mobile devices. The data generated was then visualized on the dashboard, making a clear picture of how citizens move around a city. Sensors used Machine Learning algorithms to catch a unique MAC number of a mobile device, track origin and destination, time spent in an area of a mobile device. The existing dashboard was used to visualize data; however, due to the specificity of PoC, the technology provider firm downloaded, analyzed, and visualized all data manually.

1.3.2. Data Collection and Analysis

The qualitative data collected with the participatory observation method was used as the first order empirically grounded facts that were strung together with inductively aroused points from other knowledge sources that are presented in Table 2. Observation methods enable revealing aspects that people do not will or are not able to discuss verbally, i.e., in interviews or surveys, and help to establish not hierarchical, such as researcher vs. research object bue rather non-hierarchical relationships (Clark et al., 2009) that is especially important for research on the sensitive topic, such as cognitive biases. Moreover, a longitudinal participatory observation is an excellent test-bed to explore technological awareness change over a period.

Data source	Description	Time/Pages
Participatory observations	Participatory observations were done during meetings, dials, and interviews with TC. Reactions, reasoning, changes, etc. were documented right after each event. Also, the author was observing how PoC was planned, technology was set up, etc. on the side of the technology provider.	92,2 hours of participatory observations, where, 12,2 hours – observations of TC and 50 hours technology setups, data analysis, and visualization. Plans – 30 hours 50 hours of documenting and reasoning the events afterward
Meetings	Meetings were dedicated to presenting and discussing the results of data generation. TC managers saw results for the first time only during meetings.	3 meetings, 4,5 hours in total
Dials	Calls were done at the initial stages of PoC to exchange expectations with the technology provider	2 calls; 1,5 hours in total
PoC planning	The researcher together with the CEO of the technology provider was planning and designing the PoC steps, KPIs, goals, etc.	20 hours of direct involvement
Technology set up, Data analysis and visualization	The researcher was involved in coordinating sensors set up outdoors, in choosing locations. The research conducted Big Data analysis, visualizations, and reporting.	50 hours
Face-to-face interviews	4 large in-depth semi-structured interviews were conducted as the follow-ups of the report meetings. The two TC managers were participating in the interviews. 1 large semi-structured interview with the technology provider representative	5,2 hours of interviews 1-hour interview with the technology provider representative 26,5 hours of transcript
TC: Analytics Suite for Outdoor Advertising	Final PoC plan, deadlines, KPIs, goals, descriptions, technical specifications, exact locations	12 pages

Raw Big Data	Generated Big Data, available in CSV files, data stored on servers every 15 mins: pedestrians counts, locations, time spent in the area, returned/new visitors.	3 locations in the city: 1. Train station 2. An area close to the train station 3. The main square 1 month of Big Data generation
9 Big Data reports	Every report was dedicated to the specific goal to test. They consisted of analyzed and visualized according to the developed specifically for PoC template.	44 pages
Emails	All emails from the PoC project from the side of TC. Emails from the technology provider were analyzed too to be in the context but were not included in LDA analysis.	50 emails from TC, were divided into 11 topics according to the flow of PoC: 1. Before the 1st call 2. Chase up after the 1st call 3. Pushing PoC start 4. Planning PoC 5. Before the 1st meeting 6. After the 1st meeting 7. Before the 2nd meeting 8. After the 2nd meeting 9. Clarifications 10. After the final report 11. Pricing, clarifications, projects

Table 2. Data Sources

The author was directly involved as the Digital Transformation manager working on the side of the technology provider firm and leading on the PoC project from March 2019 to August 2019. The author either directly participated or guided on all steps of PoC starting from the first initiating email until the last assessment meeting in the office, including PoC's time, finance, and strategic plans design, technology installs, Big Data analysis, and visualization (92,2 hours in total). Thus, following PoC in real-time, the author documented and interpreted all steps of the project, including planned milestones and unpredicted changes in the order of their occurrence, which took 50 hours. The summary of the steps of PoC is presented in the following Table 3. The participatory observation allowed getting the natural and real-time reactions of Managers of TC involved in PoC on results, technology awareness changes, and understand the cognitive aspects of changes within PoC. All documented actions and changes were analyzed by confronting with other data sources to balance the single participatory observer conclusions.

Main event	What happened	Comments
First contact	Technology supplier initiated the project itself.	This proactive step helped to push forward the project, as it might have taken too long to get to the start.
Proof-of-concept plan and goals designing	PoC plan – Analytics Suite - has been designed internally.	The Analytics Suite was based on the sent information about the TfL platform. TC was not initially involved in designing it.
First call	The call to discuss in short what are the needs and pain points of a public body. The Project Lead, on the side of the technology provider CEO of the technology provider Commercialisation Delivery Manager Assistant of Commercialization Delivery Manager	TC expressed readiness to start PoC, need for changes, discussed what are expectations and issues with data.
Chasing up emails from the technology supplier part pushing for the sigh-off by TC.	Waiting for approval.	5 follow up emails from the technology supplier company
Final decision to launch the PoC	The decision on which option to test was taken: all three packs and not be restricted and to get the most out of the trial.	**Three packs to test:** 1) estimate the engagement rate during the three day-time periods, find out the most engaging hour during the day: **Time-periods:** 7-9AM; 12-2 PM; 5-7 PM. **Location 1:** 2) a) Find out the most engaging time-periods for potential differentiated pricing. **Time-periods:** 7-9AM; 12-2 PM; 5-7 PM. **Location 2:** b) Find out which of the two locations billboards can be priced higher at the exact time-period (based on multi-level analysis of insights). **Time-period:** 12-2 PM. **Location 3:** 3) Estimate the level of attractiveness for a new digital billboard location. **Location:** Victoria, the location closest to the National Football Museum one of the stops on the Metrolink.
Pushing forward event	Transport for London publicized that they installed Wi-Fi sensors in all underground networks to provide a better passenger experience. TC reacted immediately and sent an email asking if it something iSensing do too.	The link on the headline: https://www.wired.co.uk/article/london-underground-wifi-tracking
Technology setups	All arrangements were done in 1 day	The technology provider was responsible for the whole technological process and set up.
Sensors switching on and templates to visualize data were designed	The technology provider was switching on sensors and checking data generation accuracy. Tailored templates to visualize data were	Using an existing sensors' network in the city, which was switched off. It was decided to use 10 meters range as the most precise for advertising purposes. The existing dashboard could provide more

	designed.	generic information, while TC needed more precise and specific insights, which was possible to get from raw data only.
Data gathering period	Data was permanently generated during this period, no disruptions observed	
Mid-term report	**Presented at the meeting:** 1. Project Lead 2. Commercialization Delivery Manager 3. Assistant of Commercialization Delivery Manager 4. Commercial Sales and Sponsorship Manager The first meeting delivering the first results of data	• The data gathered have been presented • Details discussed • Enthusiastic assumptions on how technology can be used in advertising **Additional questions on the:** Technical specification (radius of sensors, deviation rate, returned visitors) Price model for the future (licensing?) **Additional goals settings for TC:** Trying data tool for large-scale events
The data-gathering period continues		
Mid-term report	**Presented at the meeting:** 1. Project Lead 2. Commercialization Delivery Manager 3. Assistant of Commercialization Delivery Manager The results of the second round of data observation have been presented. Additionally, the dashboard visualizing the real-time data have been showcasing **The decision for the next meeting:** To present two versions of the analytics: the whole range (60 and 250 meters AND 60 m.) to see how it works different kinds of data for commercial purposes.	• The results were really surprising and unexpected for TC. TC expected that more people visit the area around sensors. • The whole meeting TC were discussing why previous intuition does not match new data, how they can use this information, how they can provide a bigger picture for advertising agencies. • It was required to change data visualization as the templates were not representative to show the whole picture
The request for raw Big Data	There was an extra request of raw Big Data as well as access to the dashboard	Managers were not confident in the data provided and wanted to double-check if the technology and results it was providing are correct. They sent raw Big Data to the Innovation team that has more experience of working with Big Data.
Final report	**Presented at the meeting:** 1. Project Lead 2. Commercialization Delivery Manager 3. Assistant of Commercialization Delivery Manager The final Big report was presented.	New visualization templates New data from a larger radius Discussing different options for the next steps: new focused pilots – where, when, price, etc.
Post PoC communication	Discussion of the focused pilots, pricing, details	The decision to continue with more focused pilots for the next fall.
Start of the new focused pilot.	The pilot has been paid	

Table 3. Summary of the steps of PoC

Further, three in-depth semi-structured interviews with the firm's three managers (5,2 hours in total) and one large interview with the technology supplier firm representative were conducted face-to-face (1 hour in total). The specificity of the interviews was that they were the supportive follow-up part of the official project meetings and participant observations. Therefore, although the interview guides with the open-end questions were prepared in advance to get answers on focused questions (Morris, 2015), interviews also included situational questions that arose after the meetings. The profiles of the participants are covered in Table 4. The dataset comprehends 26,5 hours of the transcript.

Position	Responsibilities in the PoC project	Working experience (field)	Working experience in TC (years)
Commercialization Delivery Manager at TC	Assessing PoC effectiveness, assessing the new data technology for better decision-making	13 years Advertising field and transportation field on the positions of identifying and pursuing new commercial opportunities within Commercial Media, to maximize revenue.	2 years
Commercialization Delivery Manager Assistant at TC		3 years Was working on the customer support side before.	1 year
Commercial Sales and Sponsorship Manager	Familiarizing with the data technology to be able to include it in the value proposition for clients.	15 years in sponsorship activation and business development	1,5 year
CEO of the technology provider	Technical setups	18 years of experience in Project and Product management of technological products	6 months partnership
Project Lead	Development and enabling the project milestones, KPI targets and goals meet from installs, Big Data visualization, and Business Intelligence, formal meetings and communication, dissemination of the results.	9 years Technology Marketing and Business Development.	6 months partnership

Table 4. PoC Participants Profiles

The researcher has got access to all emails sent between the technology provider and the firm before, during, and after PoC. The author analyzed all interviews and emails using the general inductive approach for qualitative data analysis (Thomas, 2006). The researcher firstly read all content of various internal documents, interviews, and emails several times to reveal emerging topics. Secondly, the author condensed all raw data into the summary format.

Thus, the researcher interpreted PoC documented steps and changes during participatory observations, confronting the results with emails, interviews, and all internal documents. Specifically, the author paid attention to the aspects driving the dynamics of PoC, such as motivation to start PoC, reasoning, and triggers to perform not planned changes, etc., as well as sources to acquire technology awareness and changes in technology awareness level.

1.4. FINDINGS

The research allowed me to explore the role of cognitive biases in tackling PoC uncertainties and driving attitudes towards technology potentiality. Moreover, it was found that technology awareness develops step-wise as the firm accumulates knowledge on the technology also benefiting from outside-in technology knowledge paths.

1.4.1. Cognition Biases on PoC Phase

The author observed the occurrence of cognitive biases at different phases of PoC characterized by specific to PoC technological uncertainty. Thus, herd behavior, overconfidence, confirmation, and information avoidance biases occurred when the firm had to choose the new technology to test, establish indicators to measure technology potentiality or prove technology potentiality while having low technology knowledge. Although sometimes it was not possible to distinguish biases from each other as some of them became a trigger for another, it became evident that some biases accompanied PoC from the very beginning till the very end, some biases occurred once. However, all four biases influenced the PoC dynamics by inflating expectations towards technology opportunity or defending prior beliefs.

In this case, herd behavior bias occurred as the response to the uncertainty raised by the need to choose a new technology to test among the big variety presented on the market. "*We have so many technologies nowadays and we really do not know which one to invest to. They can die tomorrow*" (Commercialization Delivery Manager Assistant, interview). Specifically, the Commercialization team having no knowledge of technologies that might support ads services, decided to test the technology that was already successfully proved by their senior counterparts – Transport for London (TfL). TfL is the leading London-based company in the transportation sector, and it sets the trends for the whole transportation sector in the UK and worldwide. "*There are no many organizations like Transport for London in terms of automation systems and technology. It is seen as a leader […]. And you see that because whatever city you go to - North America or South America, they mention TfL"* (CEO technology provider firm). Transport for London (TfL) implemented the same digital shift with the same Wi-Fi technology to improve the use of both transport and advertising assets. Thus, managers were referring to the TfL's technology solution as the "*holy grail*" and something extremely worthy to align with several times during PoC. Already in the first email discussing just the potential PoC project, Commercialization Delivery Manager specified that "*a system that was developed by TfL […] – it's the holy grail in terms of audience data, we'd love to have access to this level of data*." The Manager provided even screenshots of the platform to display the solution. Moreover, the author observed that the PoC project moved from the discussion to the real planning level just after the TfL case appeared in the leading news media. Right in the day when the Commercialization team spotted the headlines in news media, they wrote an email: "*I note that TfL Wi-Fi data use is making the headlines today, […]. The TfL pilot achieved some really interesting outputs […]"* (Commercialization Delivery Manager, email). While the time between the project initiation and the final decision to run PoC took one month all crucial details of the project were discussed and agreed upon right in the day when TfL made headlines and directly after the frim sent an email highlighting the success of the similar project. All technical setups to start PoC were done in one week after this event. Therefore, the researcher observed that herding supplied the positivism in perceiving the potentiality of the technology from the very beginning of PoC. This enthusiasm led to the inflated expectations towards the new technology initially. Moreover, the researcher found that herding accompanied PoC until the final decision on PoC success. This persistence in having herd behavior bias throughout PoC can be explained by the fact that once chosen even under bias, the technology remained the same. Thus, the final answer to the question from the technology provider "*Overall, is this level of data is enough for you to implement some strategic changes?",* Commercialization Delivery Manager replied, "*We need something like TfL example, yeah, something like that*" (interview, last assessment meeting).

Further, the PoC team having no experience in running PoCs and having low technology awareness did not establish any indicators or measurements that would signalize that technology proves to have a potentiality. This raised a special uncertainty moment when it came to the assessment of the first results. What the author observed is that managers started using their previously made intuitive estimations to compare with new technology results as potentiality proving indicators. With this respect, overconfidence bias arose as the confidence in the accuracy of prior personal estimations was higher than in the accuracy of the results produced by data technology. Specifically, although the use of prior estimates was not specified as an indicator in any document and was not directly mentioned in any interviews, the author observed that managers were reflexively referring to them as to the criteria to meet and were perceiving this match as the signal that the technology produces correct data. "*It looks correct, it is close to our* *estimates*" (Commercialization Delivery Manager). Paradoxically, an enthusiasm was expressed towards technology that met prior intuitive estimates and not to the reverse that is more logical in the PoC context. The managers perceived the fact of "matching" prior estimates and new technology results with enthusiasm saying that new Big Data insights "*have the potential to improve decision-making on the differentiated prices strategy*." (First assessment meeting, notes from observation). As a contrasting point, although the project had well-articulated KPIs in the PoC plan, all of them were connected either with the responsibilities of the technology provider, such as delivering "*Reports and Analytics twice per month* [...]" or with the technical part of the technology, such as "*sensors up time 99,9%"* (Analytics Suite for Outdoor Advertising). Going further into overconfidence bias, the author found that during the second assessment, when data showed a dramatically lower number of pedestrians than historical estimates, managers started expressing doubts about the overall technology potentiality, however, not about the accuracy of their prior assumptions. They stated that new technology is not able to add value to their services: "*It is an extremely low number of citizens on the station. It cannot be like this. This data cannot promote our services to our clients*" (2nd assessment, notes after the observation). Although managers were quite persistent in trying to get what exactly caused this mismatch between prior decisions and new results, they were focused more on the questioning technology performance rather than their historical assumptions. Thus, being latched onto the historical assumptions and to the wrong indicators, the Commercialization team requested access not only to the visualization dashboard but also to raw Big Data to double-check if new results were correct, although it was not presupposed and discussed initially by the PoC conditions. The team also wrote several catch-up emails (although during PoC, the communication between the technology provider and TC trams was at its minimum level) asking if "*Is it possible to get the* *raw data from the pilot in the interim?*". Further, the team shared raw data with the third side, the TC Innovation division, which had access to other sources of data to double-check new information.

Further, the author found that the desire to stick to prior beliefs and find confirming rather than confronting facts led to the confirmation bias. Specifically, the PoC team was proving the concept by confirming and looking to prove their expectations or slightly different from intuitive calculations facts. Exemplarily, the technology showed that all peak and off-peak times during working days are as estimated previously by the Commercialization division except the one peak-time, which was confronted by data (8-10 AM instead of 7-9 AM as expected by managers). Rather than questioning why it is so, managers perceived this fact as the confirmation of the good performance of technology, as it was confirming their expectations about the general peak-times trend line and did not influence the commercialization ability of assets. "*Interesting! Never thought this!*" (1st assessment, notes after the observation, Commercialization Delivery Manager Assistant). The author observed that a slight change in what was expected from the technology and actual results were perceived as confirming the overall effectiveness of technology under the test. Thus, at this level, confirmation bias supplied a positive perception of the technology potentiality.

Moreover, the author observed that the team also was proving by avoiding confronting results. Specifically, information avoidance was observed when managers avoided considering the results that were not consistent with their prior estimations until the data showed the same results over again later. Specifically, new data stated that the number of people in the exact train station was well less than expected, managers, although having recognized this fact ("*Interesting number*", Commercialization Delivery Manager, 1st assessment, notes from the observation) switched their attention to another “positive” facts and these very results were not discussed during the first assessment meeting. The information avoidance bias was possible to spot, as the fact of a lower number of pedestrians was critical for TC, it would influence the potentiality of the area to be priced higher for advertisers, and not paying attention to the key estimate was a rather salient fact. Moreover, one PoC pack was designed specifically to prove the ability of the technology to price higher in some areas, as they are more pedestrians there. Although according to the researcher’s observations, information avoidance occurred once during PoC, the author found inattention avoiding some pieces of information helped to remain positive about the overall technology potentiality, initially.

1.4.2. PoC Technology Awareness

In line with the theoretical background on PoC, the author introduces the technology awareness variable in this section. The researcher observed that while, firstly, the directionality of the technology knowledge path was outside-in, later it changed to the pure inside-in, where the Commercialization division was developing its own unique technology awareness. Participatory observation and qualitative content analysis allowed us to capture not only the fact of growth of the technology awareness during PoC but also specific stages of what the author labeled as 1) *borrowed technology awareness*, 2) *minimum acquired technology awareness*, and 3) *enhanced technology awareness.* Additionally, the author runs latent Dirichlet allocation[1] content analysis to witness the changes in technology awareness level at the initial and final stages of PoC.

Specifically, *borrowed technology awareness* is characterized by the low level of own knowledge about technology and, therefore, the usage of expertise of more experienced fellows’.

[1]LDA is a generative probabilistic model of a text corpus that uses an unsupervised Bayesian machine learning algorithm to discover context-specific hidden topics (Blei et al., 2002). The method has started gaining attention in strategic management related to cognitive sensitive topics to discover latent themes in collection of documents (i.e., Kaplan and Vakili, 2015).

An analysis of communication between the Commercialization team and technology provider revealed that the focus at the pre-launch phase of PoC (technology choice, PoC planning, and technology setups) was mainly on the discussion around commercializing assets need and the TfL case. Although, the purpose and capabilities of the technology were clear for the Commercialization division from the beginning ("*We also hold no data on pedestrian flows [...] to plan things like the position of digital advertising sites and customer information. What is the total audience of an advertising display on this stop? We hold no data on the total footfall surrounding our assets*", email after technology choose and before to launch PoC), there was no clear understanding how the technology works ("*is there much difference in the tech and data derived from their [TfL] system?",* email after technology choose and before to launch PoC). Additional questions on the technology performance were connected only with the technology functions (e.g., "*If we deploy on stop will there be a way to distinguish those that are tram passengers?",* email after technology choose and before to launch PoC). Moreover, there was no knowledge of how to plan PoC and set up the technology. Therefore, it not only borrowed the firm the experience of TfL but also of the experience of the technology provider and outsourced them the whole PoC roadmap and goals design as well as technology setups. The role of the Commercialization division in designing the PoC plan was passive as they only gave their "go-ahead" for PoC and accepted the initial proposed PoC plan with the three packs to test ("*I think we should proceed with all packs if it's not too much work? All the metrics will be useful to understand from a media perspective*", email after technology choose and before to launch PoC).

Minimum acquired technology awareness is characterized by the presence of more specific knowledge on technology features that might influence technology performance and the more active role of TC in running PoC. Specifically, once managers had to measure technology's first results and performance by themselves (at the 1st assessment), the firm started actively acquiring new knowledge on very specific aspects of the "*radius of sensors*", "*deviation rate of analytics*", including, mobile penetration rate or pedestrians with switched-off Wi-Fi search option in mobile devices (1st assessment and follow-up emails). New knowledge enabled the shift from the passive to a more active role of managers in running PoC. Thus, at the point where PoC started collapsing due to very low perceived technology opportunities, managers took over an initiative from a technology provider and started looking for solutions and suggesting to change the technology setups. Specifically, when managers recognized that the amount of pedestrians in a specific zone is extremely low compared to what they intuitively calculated, they proposed to enlarge the radius of sensors to track pedestrians from 10 to 60 and then to 150 meters to have a larger coverage range. From the beginning, sensors were operating in a radius of 10 meters as it was considered correct to provide more precise data on an engagement rate. While the majority of screens were big enough to see the advertising content from 60 meters or even 150 meters. Thus, the technology was adapted for the advertising specificity use and the proposed solution saved PoC from being closed.

Finally, *enhanced technology awareness* phase witnessed the shift from passive to the pro-active role of the Commercialization division in running PoC and from knowledge on technology functions to industry-specific technology awareness. Thus, at the final assessment and follow-up emails discussing future projects, managers expressed concrete vision and a plan, indicating where they want to install sensors, what exactly they need to get from data, and what they need to measure. They were not anymore outsourcing strategic decisions to the technology provider and minimized their involvement to the level of technical setups and data accuracy gathering. Further, managers discussed concrete technical aspects for future projects, such as "*footfalls*" counts, exact "*locations*", "*sensor networks*", quality of "*data*", time to launch (emails and last assessment meeting). Finally, the firm proposed ad-hoc and industry-specific technology applications that were never discussed during PoC focusing on the value of the technology (e.g., *"We don't need it [solution] constantly, we need the data to be able to quote what footfall is for one stop or for advertising quotes, and then in six-month time we might check to make sure we are selling on accurate data*". Or: "*There might be an opportunity to link two things up, where a sensor is married to a screen and there is a bank of content […]. So it's dynamic, programmatic, so the ad will play a screen that is relevant to the audience based upon some sort of cash of adverts that already stored on the CNS of the screen*", last assessment, Commercialization Delivery Manager Assistant).

Additionally, the change in the active vocabulary concerning technology awareness at the beginning and the end of PoC has been witnessed by quantitative content analysis. Specifically, the results of the performed LDA analysis presented in the following Table 5 allowed deriving four key topics at the initial phase and final phase of PoC.

The initial phase of PoC		**The final phase of PoC**	
Key Words	**Weight %**	**Key Words**	**Weight %**
Topic 1 "Future Vision" Vision and priorities of TfGM for today and the future, focusing on innovation,		**Topic 1 "Future focused Pilot"** Discussing new focused pilot to launch	
know	1.93	sensors	2.8
team	1.07	stop	1.34
manchester	1.07	peter	1.18
used	1.07	footfall	1.18
movement	1.07	accurate	1.18
analytics	1.07	useful	1.02
does	1.07	month	1.02
sensors	1.07	piccadilly	1.02
innovation	1.07		
tomorrow	1.07		
trends	1.07		
nowadays	1.07		
people	1.07		
really	1.07		
invest	1.07		
gather	1.07		
Wi-Fi	1.07		

die	1.07		
technologies	1.07		
city	1.07		
council	1.07		
Topic 2 “Call-to-action” Consists only of two terms; the firm does not discuss any details on how to set up the pilot at the very beginning at all		**Topic 2 “Usefulness of Big Data”** Concrete outcomes of PoC for the firm’s advertising division	
sensor	1.54	advertising	1.18
deploy	1.54	audience	1.18
		peak	1.18
		useful	1.18
		seen	1.18
		start	1.18
		sharing	1.18
		campaign	1.18
Topic 3 “Data for assets” Data to improve both transportation and digital assets use		**Topic 3 “Applicability”** Discusses a new idea of application that was born during PoC	
data	3.29	things	1.44
transport	1.41	think	1.44
digital	1.41	pilot	1.44
public	1.14	screen	1.1
use	1.14	advertising	1.1
lot	1.14	sales	1.1
assets	1.14	link	1.1
people	1.14	metrolink	1.1
		assets	1.1
		know	1.1
		people	1.1
		level	1.1
		direct	1.1
Topic 4 “Commercialization” Issues concerned historical advertising contracts and the need to do something new to better commercialize		**Topic 4 “Data add-value”** Data benefits for advertising companies (customers) the firm is working with	
assets	1.98	need	2.21
advertising	1.71	sensors	2.2
contracts	1.71	stop	1.85
team	1.44	network	1.85
make	1.44	way	1.5
commercial	1.44	locations	1.15
money	1.44	metrolink	1.15
based	1.16	companies	1.15
people	1.16	understand	1.15
think	1.16		
new	1.16		
grow	1.16		

Table 5. Topics and Key Words Distribution in LDA

Thus, the initial phase, on the one hand, is characterized by the presence of vision-driven orientation for tomorrow and an understanding of the need for changes for better business performance. On the other hand, low technology awareness, which can be proved by referring to wrong industry experience (transportation instead of solely advertising), absence of technology vocabulary, and by the passive role of TC company in launching PoC. While, the LDA analysis of the final phase proved the presence of acquired own technology experience, which can be proved by the active use of the specific technology vocabulary, the active role of TC in planning future digital transformation initiatives, finally, the orientation toward technology performance rather than vision. The analysis of the finishing phase of PoC indicated a clear intention to move further to future focused projects.

1.4.3. PoC Dynamics

Following the findings, the author observed that expectations and beliefs were driving the perception about technology under PoC, which the author labeled as *perceived technology potentiality*. The author discovered that while *perceived technology potentiality* was fluctuating in two directions – either up or down, *technology awareness* was always in the "accumulating" or growing direction. Moreover, technology awareness came into power driving PoC dynamics, specifically when it moved from *borrowed* (outside-in) to *minimum acquired technology awareness* (inside-in). To illustrate these dynamics within PoC, the author put the two constructs: *perceived technology potentiality* and *technology awareness* as the axis and draw the trend line of PoC dynamics (see Figure 1).

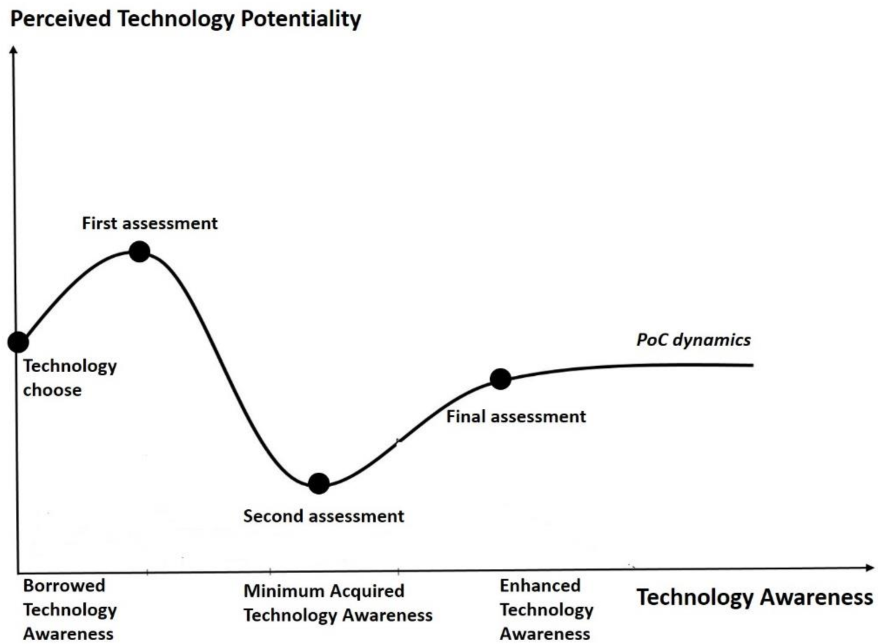

Figure 1. PoC Dynamics

In this vein, firstly, herd behavior caused positive expectations towards the technology and influenced the technology to choose a decision. Once PoC went to the active phase of testing the technology, other cognitive biases, specifically, overconfidence, confirmation, and information avoidance biases, appeared to respond to the PoC specific technological uncertainties. Mostly, biases were defending the *perceived technology potentiality* as the technology started showing confronting prior beliefs results. Therefore, this fact doubled with the *borrowed technology awareness* caused the highest peak (in this case, this occurred at the first assessment) of *perceived technology potentiality* during the whole PoC. However, as technology kept producing confronting results and the firm started acquiring *minimum technology awareness*, biases were not anymore able to defend *perceived technology potentiality*, and the curve declined. At the lowest point of PoC dynamics, that in this particular case, occurred during the second assessment, where the shrunk in trust towards technology happened, the team activated *minimum acquired technology awareness* as the second resource to keep running PoC. From the low point, the PoC dynamics started improving, and the curve did not show over-or under-optimism about *perceived technology potentiality*. Moreover, after the low point, the Commercialization team kept accumulating and enhancing technology awareness, which was used to make the final decision on the *perceived technology potentiality* and adoption.

Further, the author found that cognitive biases were occurring anytime when the Commercialization team had to act independently from the technology provider and take own decisions, such as choosing a technology to run PoC, running technology assessments, or deciding on adoption or rejection. The author observed that initial positive expectations caused by herding supported by overconfidence, information avoidance, and confirmation biases during PoC were mitigated by the acquired technology awareness; however, the herding bias remained till the end of PoC.

The author also found that in analogy with cognitive biases, *technology awareness* degree shift worked as the response to the new uncertainty posed by PoC confronting results. Thus, while overall *technology awareness* was growing throughout the whole PoC, the shift from one to another degree of *technology awareness* occurred at critical points where the firm had to take its own decisions. Specifically, *borrowed technology awareness* was activated when the firm did not have any knowledge of how to run PoC, thus, outsourcing the whole PoC plan design to the technology provider. Second, the shift from *borrowed* to *minimum acquired technology awareness* occurred at the lowest PoC dynamics point, when PoC had a risk to die. Finally, the shift from *minimum acquired* to *enhanced technology awareness* happened at the final decision on adoption and future projects after having a big drop in perceived technology potentiality before. That was again an uncertain situation.

1.5. DISCUSSION AND CONCLUSIONS

The present pioneering research explored PoC for Digital Transformation when firms face the challenge of proving new technology potentiality under uncertainty and with low technology awareness. The author used a qualitative case study approach to capture granular details on how four cognitive biases, namely, herding, overconfidence, confirmation bias, and information avoidance, drove what the author labeled as *perceived technology potentiality*. Moreover, the author delineated three degrees of technology awareness that developed as PoC is run, namely, *borrowed technology awareness, minimum acquired technology awareness,* and *enhanced technology awareness.* Finally, the author drew PoC dynamics explaining how the two constructs, *perceived technology potentiality,* and *technology awareness* strengthen or balance each other in the PoC milieu.

Specifically, the author found, PoC phase on the one side dealing with uncertainty and complexity of Digital Transformation (Hess et al., 2016; Bohnsack et al., 2018), on the other side with the PoC-related moments of uncertainty such as technology innovation, adoption, and foresight (Geroski, 2000; Bhattacherjee and Premkumar, 2004; Bonaccorsi et al., 2020), is especially vulnerable to cognitive biases occurrence. Following the literature stating that biases occur in uncertain situations (Busenitz and Barney, 1997; Haselton et al., 2005; Zhang and Cueto, 2017), the author enlarged the understanding of what are those situations within the PoC context, specifically, choosing the technology among the variety, proving performance and potentiality, or establishing clear indicators to estimate technology potentiality. Thus, the observed fluctuations of *perceived technology potentiality* were especially prominent in the points of raised uncertainty. Moreover, perceived *technology potentiality* was influenced by cognitive biases and, therefore, was the subject of firstly over positive and later, negative attitude.

The research deepened an understanding of the role of *technology awareness* at the PoC phase. The author argues that not only helps *technology awareness* to neglect uncertainty at the PoC phase (Bohn et al., 2005; Roca et al., 2017) but also the firm shifts step-wise from borrowed (outside-in) moving to minimum acquired (inside-in) to enhanced technology awareness (inside-in) in the points of specially raised uncertainty. Moreover, the author was able to capture the mechanism of the firm's different tactics to acquire *technology awareness*. While the firm used technology supplier knowledge source as the main at the beginning of PoC (Ragatz et al., 2002; Fossas-Olalla et al., 2015), later it started learning by trying (Fleck, 1994; Roca et al., 2017); this became the main tactic to acquire *technology awareness*. The author argues, while the firm was acquiring *technology awareness* using different tactics throughout the whole PoC, new technology knowledge was activated specifically in points of fluctuations in *perceived technology potentiality*.

Furthermore, the author captured a specific relationship between cognitive biases and *technology awareness* that designs the dynamic of PoC overall. In situations of especially inflated uncertainty and complexity, both cognitive biases and *technology awareness* shape attitudes towards technology within PoC. Moreover, they both become a part of experiential learning (Gaba and Terlaak, 2013) and enable learning from the firm's own experience (Shaver, Mitchell, and Yeung, 1997) in specific uncertainty points that constitute the curve of PoC dynamic. This made us conclude that the firm benefited from both an increase and decrease of *perceived technology potentiality* caused by biases by balancing those fluctuations with different degrees of *technology awareness*.

The present research bridges two streams of literature to initiate the discussion on the phenomenon of PoC that was surprisingly by-passed in the technology innovation literature. Firstly, the author contributed to the understanding of PoC by defining its main constructs, namely, goal, knowledge and activities types, and directionality of the knowledge paths. Further, the author contributed to the managerial cognition literature by focusing specifically on beliefs and attitudes in the technology innovation, foresight, and adoption context and combining it with fragmented research on technological uncertainties (Kahneman and Tversky, 1977; Banerjee, 1992; Busenitz and Barney, 1997; Ragatz et al., 2002; Zhang and Cueto, 2017). The author suggests that cognitive biases play a distinct role in the driving perception of the technology potentiality during the PoC phase. The author enlarged understanding of the existing cognitive biases (Banerjee, 1992; Klayman and Ha, 1989; Gudmundsson and Lechner, 2013; Golman et al., 2017) looking at them through the PoC lens and derived specific situations during PoC when those biases may occur. Further, the author introduced the notion of technology awareness degrees in the PoC phase and enlarged the existing research that has covered mainly the technology product innovation context so far (Fleck, 1994; Bohn et al., 2005; Roca et al., 2017). The author suggests that technology awareness should be included as an integral factor of PoC dynamics. Finally, the author introduced the new intriguing dynamic of PoC that explains through two constructs, namely, *perceived technology potentiality* and *technology awareness*, how cognitive biases and shift in technology awareness drive PoC.

2. IS DATA-DRIVEN DECISION-MAKING DRIVEN ONLY BY DATA? WHEN COGNITION MEETS DATA

ABSTRACT

Data-driven decision-making is becoming a pillar of strategic decision-making. With the exception of the science fiction realm, totally data-driven decision-making is still very far from expectations: managerial cognition can influence not only strategic decisions but also data and algorithms. This research aims at investigating what are and how cognitive and data biases influence trust in data when data-driven decision-making is firstly introduced. Using an ethnographic approach and exploring the case of a traditional transport authority in the North of the UK, I derive the distinct managerial cognition and data biases. I introduce several conceptual components characterizing what I called *cognition, data technology,* and *traps recognition zones*. I focus on how trust in data changes as managers fall into one of the three traps zones. The study sheds light on a better understanding of the mechanism underlying the interaction between cognitive and data biases that influence data trust and put the break into the parallel-competitive theory by suggesting the intriguing dynamics and synergy of cognition and data. In doing so, the study extends the literature on managerial decision-making contributing to the emerging stream on data-driven decision-making.

Keywords: managerial cognition, cognitive bias, data-driven, decision-making, sensitive analysis, ethnographic research

2.1. INTRODUCTION

"Opinions and intuitions play an important part even where the forecasts are obtained by a mathematical model or a simulation" (Kahneman & Tversky, 1977, p. 1).

Digital transformation is a strategic priority for all firms, especially traditional ones who built their success before pre-digital reality (Sebastian et al., 2017; Hess et al., 2016). Among all opportunities brought by ICT and the new technological enabler (*i.e.*, artificial intelligence, internet of things, big data), data-driven decision-making is becoming the holy grail for firms willing to improve the overall decision-making process (Ghasemaghaei et al., 2018; Bertsimas and Kallus, 2020; Camiña et al., 2020; Raisch and Krakowski, 2020). At the same time, with the exception of the science fiction realm, an entirely data-driven decision process is still very far from expectations (Akter et al., 2019; Gupta et al., 2018; Choudhury et al., 2020).

The common consensus among strategic management scholars on managerial decision-making is that there are two fundamental decision-making types: judgmental and logical (Hodgkinson and Sadler-Smith, 2018; Simon, 1987). The most recent view based on the parallel-competitive theory is that both types work in parallel, each contributing to the decision to take if necessary (Hodgkinson and Sadler-Smith, 2018). Therefore, a final decision is taken when intuitive analytical types have responded. When it comes to judgemental type, existing strategy research has highlighted that it is often the subject of cognitive biases that influence the final decisions (Busenitz & Barney, 1997; Gudmundsson and Lechner, 2013; Van Knippenberg et al., 2015; Zhang et al., 2020). The emerging stream of data-driven decision-making (DDD)—urged by unprecedented availability of data and data technologies, such as Artificial Intelligence (AI) and Machine Learning (ML)—has gained special attention within strategic management research as it could complement the logical type of humans in decision-making (Van Knippenberg et al., 2015; Choudhury et al., 2020; Shrestha et al., 2020). Hence, it is important to distinguish between the so-called “substitutional” view that explores the extreme of AI or ML algorithms and their ability to replace humans in autonomous decision-making agents either as algorithms or robots (Duan et al., 2019), and the data-driven decision-making that is focused on “supportive” functions of data algorithms and humans being the main decision-makers (Raisch and Krakowski, 2020).

Thus, with all promises of data-driven decision-making, algorithms that suppose to address imperfections and biases of human decision-making (Lindebaum et al., 2020) still coexist with humans’ cognition and intuition. Moreover, algorithms can be the subject of biases themselves. Specifically, recent research has witnessed the presence of biases that arise not directly from data but are rather based on how humans deal with algorithms (Choudhury et al., 2020). On top of this, there is the issue of managerial trust in data that is critical for the success of integrating data algorithms into decision-making (Surbakti et al., 2020; Glikson and Woolley, 2020).

This complex dynamic and the dual nature of DDD creates a particular dilemma for decision-makers. While the recent research has outlined different sides of the DDD phenomenon’s dual nature separately, the relationship between its main constructs, namely, cognitive and data biases and trust in data, is not yet well understood.

This paper aims to better understand the cognitive and data biases and how they influence trust in data when data-driven decision-making is first introduced. Based on an ethnographic approach, I explore the case of a Commercialization team of the traditional transport authority in the North of the UK applying data to drive strategic decisions for the first time. Additionally, to support qualitative findings with data-driven insights, I performed the new text mining quantitative analysis.

2.2. THEORETICAL BACKGROUND

This paper adopted the parallel-competitive theory to look through its lens at data-driven decision-making (Hodgkinson and Sadler-Smith, 2018). Based on the assumptions of the traditional school of managerial decision-making, two fundamental types of decision-making, namely, judgmental and logical (Simon, 1987), participate in decision-making. The parallel-competitive theory suggests that the two types are working simultaneously and in parallel so that a final decision is taken when intuitive and analytical types have each responded (Hodgkinson and Sadler-Smith, 2018). Specifically, the judgmental type is characterized by the rapid and intuitive nature of judgments and often by the absence of a comprehensive analysis of the situation. At the same time, decision-makers usually have a great confidence level due to the high level of experience. These circumstances can be the perfect ground for the presence of cognitive biases. The logical type is characterized by the explicit nature of taking decisions based on logic, goals, and evaluated consequences of taking alternative decisions. Despite the different nature of the two types of decision-making, not only do they naturally coexist within the decision-making process and do not exclude each other but also it is unlikely to find a strictly "logic" or "intuitive" manager. The latter will more likely combine both styles (Hodgkinson and Sadler-Smith, 2018).

In the next subsections, the author will look at the two types of decision-making competencies closer, where the judgmental type of decision-making will first be explored through the lens of cognitive biases; the author will adopt a logical perspective on data-driven decision-making and its challenges.

2.2.1. Cognitive Biases

Cognitive biases refer to the deviation decision mechanisms and subjective opinions individuals use in decision-making, especially under uncertainty, and play an essential role in the cognition ability of individuals (Busenitz & Barney, 1997; Haselton et al., 2005; Zhang & Cueto, 2017; Kahneman & Tversky, 1977; Stubbart, 1989). Different contexts imply different cognitive biases occurrence. Specifically, the initiative on testing data as the new supportive tool for data-driven decision-making implies two contexts, first, digital transformation and, second, new information adoption. Digital transformation is accompanied by the mass hype towards technologies and the belief that everyone should technologically transform. Managers willing to employ new data technology features into the decision-making process most probably will experience herd behavior bias (Banerjee, 1992; Bohnsack et al., 2018; Ding & Li, 2019). Moreover, information avoidance bias (Golman et al., 2017; Lallement et al., 2020) characterizes the specific context of new information adoption, mainly because the data-driven decisions' main component is data or information (Baesens et al., 2016). Therefore, despite the vast amount of cognitive biases discussed within decision-making literature, in this study, the author focus on the two specific cognitive biases that are more likely to occur during the first introduction of data-driven decision-making. These biases are herd behavior and information avoidance biases (Ding & Li, 2019; Lallement et al., 2020).

2.2.1.1. Herd Behavior

When individuals tend to mimic others' actions, it is called herding or herd behavior (Banerjee, 1992; Ding & Li, 2019). Specifically, a decision-maker tends to follow or imitate his/her fellow's actions and neglect his/her own opinion (Banerjee, 1992; Sgroi, 2003). Herding is more likely to have negative implications for firms (Lieberman & Asaba, 2006). For instance, managers influenced by herd behavior may make wrong and inefficient choices, while their analysis suggests that other organizations' actions are incorrect (Banerjee, 1992; Devenow & Welch, 1996; Scharfstein & Stein, 1990; Barreto & Baden-Fuller, 2006).

The process of following herd decisions is also explained by the information cascade theory (Welch, 1992; Banerjee, 1992). According to this perspective, herding's occurrence requires two conditions: product uncertainty and the observed repeatable actions of others (Banerjee, 1992; Ding & Li, 2019). There is the third condition within information cascade theory, which plays a vital role in herding, known as the shared-identity phenomenon (e.g., Berger et al., 2018). In complex and dynamic environments overloaded with information, managers tend to use cognitive simplifications to cope with ambiguity (Barreto & Baden-Fuller, 2006). Decision-makers create cognitive categories of firms combined with relevant similarities. The categorization influences their taken decisions in the way that a firm tends to converge with a category in a reference point (Porac & Thomas, 1990). Not only herding more likely occurs when actors share social identity (Weatherall & O'Connor, 2020), but also shared social identity itself promotes herd behavior (Berger et al., 2018).

When it comes to technology innovation and digital transformation, herd behavior occurs, especially when choosing between alternatives in adopting new technologies exists (Bhattacherjee & Premkumar, 2004; Sun, 2013). In the circumstances of choosing between two or more technologies, herd behavior plays the role of the mechanism that leads later technology adopters to reduced willingness to invest in making a serious choice between technologies (Geroski, 2000; Duan & Whinston, 2009). Interestingly, if others prove one technology to work better than the existing technology, firms have no motive to invest in testing other technology (Geroski, 2000).

2.2.1.2. Information Avoidance Bias

Information avoidance bias leads to neglecting of information, although it could potentially improve decision-making (Golman et al., 2017; Lallement et al., 2020) and sticking to the "comfortable" beliefs until individuals experience the negative consequences of the avoidance if the information (Zhang & Cueto, 2017). Selective information acceptance plays a vital role in information avoidance bias. It works as the deliberate choice mechanism, where individuals tend to accept positive evidence and avoid neutral or negative facts (Golman et al., 2017; Zhang & Cueto, 2017). Among the most occurring reasons why individuals avoid information (Sweeny et al., 2010), there are motives of natural resistance to change existing beliefs, avoiding undesired actions, and "risk, loss, and disappointment aversion" (Golman et al., 2017). For example, investors avoid checking their financial portfolios when the stock market is down (Sicherman et al., 2016); managers avoid recognizing arguments that contradict their prior beliefs and decisions, even when such arguments could help them avoid implementing measures that are insufficiently substantiated (Deshpande & Kohli, 1989; Schulz-Hardt et al., 2000).

There exist various tactics to avoid information. *Physical information avoidance* is the basic one, as individuals choose simply not to consider information (Golman et al., 2017). *Inattention*, another tactic, does not presuppose active avoidance; however, individuals tend to allocate their attention in a most efficient for a decision-maker way, therefore, presupposing rational inattention to some pieces of information (Golman et al., 2017; Caplin & Dean 2015; Sallee, 2014). *Biased interpretation of information* is conceived to avoid drawing the most logical picture from received information (Golman et al., 2017). In this case, individuals interpret evidence in the most comfortable way to keep holding wanted beliefs. In this way, managers may not draw obvious conclusions from data (Golman et al., 2017). Research suggests that against logic, more intelligent individuals tend to misinterpret information and locate their intelligence to support what they want belief, or that creative individuals find it easy to be dishonest in data interpretation (Gino & Ariely, 2012; Kahan et al., 2012).

2.2.2. Data as The Supportive Tool in Decision-Making

Data-driven decision-making is understood as the practice to inform decisions with data analysis rather than base them solely on intuition (Provost & Fawcett, 2013). The emerging within strategic management topic of data-driven decision-making with some exceptions has been focused on exploring mostly the narrow topic of data algorithms, machine learning, Big Data analytics technologies, and their futuristic opportunities for better decision-making so far (Bertsimas & Kallus, 2020; Camiña et al., 2020; Mandelbaum et al., 2020). Only recent research has started exploring the risk of human biases being imported to machines due to increased interactions between managers and algorithms, as well as the ways to mitigate them (Raisch & Krakowski, 2020).

Specifically, several scholars pointed out the importance of the data input that feeds algorithms and indirectly influences decisions. Lindebaum et al. (2020) highlighted that while data algorithms are automatic and follow programmed rules, there is room for data input vulnerability, as data input depends mostly on humans. Further, Choudhury et al. (2020) have proposed the term input incompleteness bias—referred to as the situation when ML algorithms are not initially provided with all relevant information required for deriving results due to the human intentional or not intentional behavior. As the mitigating power here, Choudhury et al. (2020) suggested the importance of the two aspects, namely, managerial domain-expertise and the knowledge of how to deal with the data technology. Domain expertise is crucial as managers set inputs and outputs based on their domain knowledge for supervised learning and decide inputs for unsupervised learning (Raisch & Krakowski, 2020). When it comes to know-how, research has already highlighted the need to acquire data technology knowledge about the whole chain of the new decision-making process moving from precise data analysis goals (Vidgen et al., 2017) to specific knowledge on data capture and correct incorporation into decisions (Waller & Fawcett, 2013; Baesens et al., 2016). However, it results in a dilemma for managers at the initial stage of incorporating data in decision-making – working with data while not acquiring knowledge yet. Therefore, the systematic issue here is that managers "outsource" the whole data analytics process to IT departments or data scientists, mainly as data analysts are still the primary data knowledge holders (Vidgen et al., 2017; Watson, 2014). This can become the source of another human bias imported to machines.

Further, as DDD is data-dependent, it ultimately depends on how data is treated (Kattel et al., 2019). Specifically, although managers do feel very positive about data when it comes to incorporating it into decision-making, they still might believe that their intuition is more accurate than data analytics (Mikalef et al., 2018) or that algorithms cannot replace the instinct of humans (Surbakti et al., 2020). Thus, managerial cognition and emotions can leave out of consideration the Big Data insights (Hodgkinson & Healey, 2011). More broadly, Baesens et al. (2016) discussed the effect of trust in data and analytics as one of the critical factors of the effective DDD and derived the success formula of the power of data as "Information (Data) + Trust." Authors suggest that even with correctly derived data insights, a firm will not benefit if managers do not trust data and analytics techniques. More specifically, though, the recent study of Glikson and Woolley (2020) suggests that trust works differently for different types of AI. In the case of AI presented as embedded data technology (not robots or autonomous online agents) and that is the closest context for DDD, initial trust in data is high, while the further use of data trust intends to reduce. Additionally, trust is quite a sensitive substance. Even statistically significant algorithms and models can generate a lack of trust. This situation occurs if managers that traditionally base their decision-making on economic returns meet analytical insights that are just the result of a statistical exercise and are not adding any business value (Baesens et al., 2016).

2.3. METHODOLOGY

The present study applies an ethnographic methodology as the phenomenon of the investigation lies on the verge of managerial cognition and data technology. As such, it requires full immersion into the environment and context of those under observation for a long time to be comprehensively explored (Sanday, 1979; Van Maanen, 1979). In this way, the author could explore more in-depth the new phenomenon's dynamics (Schensul et al., 1999).

2.3.1. Setting

In this paper, I examine the context of the first use of data to improve strategic managerial decision-making. I chose the case of the leading firm responsible for coordinating transport services in North West England as it is the representative situation for most industries when a traditional company with low knowledge about data technology decides to test new for the firm data-technology and assess new commercial opportunities of data-driven decision-making. Moreover, the trial was run by the two managers of the Commercialization division of the firm; therefore I was able to be focused on cognitive and data biases at the managerial level.

The general context of the case is that the Commercialization Division of the transportation firm called for changes for their outdoor advertising services. They recognized the need to replace the old-fashioned way of price formation and switch to the social media formula but for outdoor advertising. Hence, form the prices for outdoor advertising according to an engagement rate or the number of pedestrians seeing the advertising. The trial project's general objective was to understand and provide pedestrian behavior analytics using and visualizing Big Data. Three schemes to test were defined: how to help outdoor advertising agencies working with the transportation company to plan effective ads campaigns; how to diversify services portfolio and build smarter advertising revenue streams; how to optimize and commercialize the company's spaces. However, the high-level idea was to try how Big Data would work for enhanced and data-driven decision-making and new revenue stream creation.

Four managers were involved in the test project: two managers of the transportation firm's Commercialization Division and two managers of the Big Data technology provider. The Division was new (two years), and it was responsible for the commercialization of projects that lay apart from the main transport activities, such as outdoor advertising services delivery. The technology firm was a two years old startup using WiFi sensors to generate Big Data about pedestrians. The main startup specialization was the transportation industry, with a few projects in other spheres of cities' services.

As the Commercialization Division almost did not have experience working with Big Data, the whole technical and strategic assistance was provided by the data technology supplier company. Thus, plans, strategy, KPIs, and goals were developed by the technology provider's project lead. The project's sensor network was installed across an indicative study area; it was used to capture real-time data and examine historical data (within the project duration) on pedestrian movement and dwell-time, including origin/ destination, journey time statistics, footfall, duration. The technology solution was WiFi sensors that search for WiFi signals from mobile devices. The data generated was then visualized on the dashboard, making a clear picture of how citizens move around a city. Sensors used ML algorithms to catch a unique media access control (MAC) address of a mobile device, track origin and destination, time spent in an area of a mobile device. The existing dashboard was used to visualize data; however, the technology provider managers downloaded, analyzed, and visualized all data manually due to the project's specificity.

2.3.2. Data Collection and Analysis

The ethnographic methodology presupposes building an evidence base and collecting and analyzing a vast amount of data to immerse into the phenomenon. All these require the use of different methods to analyze data. The present research consists of the three mains phases guiding the researchers in the multi-sources and methods research. They are displayed in Table 6.

Phases	Steps	Data Collection and Analysis Methods
I. Formative Research	1. Meeting with the team responsible for the project: on the side of the transportation company and the technology provider 2. Formalizing the role of the researcher in the participatory observation 3. Developing semi-structured interviews with all research team members 4. Developing observation goals 5. Getting constant access to the flow of project-related documentation (KPIs, timelines, plans, responsible, budgeting, Legal Documents, chats, etc.)	• Preparation for data collection: • Guides for interviews (Morris (2015) • Developing observation forms • Collection of the initial information on the firm, setting
II. Ethnographic case study research	1. Conducting semi-structured interviews at each core phase of the project: 1) planning and initiating, 2) data technology assessments, 3) final report 2. Participatory observing participants of the project, including dials and assessments at the facilities of the firm 3. Formalizing observations in the detailed observation forms during and after observations 4. Reading all incoming and outcoming communication via email in real-life,	Data collection through data triangulation (Yin, 2013): • Participatory observation (Carr et al., 2008; Bryman, 2012). • In-depth semi-structured interviews (Morris (2015) • Internal documents and communication gathering

	documenting considerations 5. Reading and scanning all incoming and outcoming official documentation in real-life, documenting considerations	
III. Qualitative Data Analysis	Analyzing detailed observations forms, communication, and documents.	Inductive approach (Gioia et al., 2013).
IIIa. Quantitative Data Analysis	Analyzing all raw contents quantitatively (emails and interviews)	Text Mining: Sentiment Analysis (Pandey et al., 2017; Moe & Schweidel, 2017)

Table 6. Phases, Steps, and Data Collection and Analysis Methods

In this vein, Phase I, formative research, focused on the initial information on the setting, questions, and guides for the interviews prepared. The interview guides were prepared following Morris's (2015) practical suggestions. There were prepared observation forms to fill in during the meetings, calls, and data technology setups, data visualization, and assessments to collect information on data issues and cognitive biases that may occur during the project. All co-authors participated in this stage of the research.

Phase II focused on the data collection through triangulation of data. I used participatory observation to ground the findings; interviews and internal documents served as important data sources. Firstly, the participatory observation was conducted as it is recognized as effective to assess the antecedents and consequences of the specific behavior within a particular context (Carr et al., 2008; Bryman, 2012). Moreover, the participatory observation method allows to build non-hierarchical relationships (Clark et al., 2009); I found it very important for sensitive topics, such as cognition. Thus, one of the co-authors was directly involved as the Digital Transformation manager working on the technology provider firm’s side and running on the project from March 2019 to August 2019. The co-author participated in all steps of the data technology test starting from the first initiating email until the last assessment meeting in the office of the firm, including project's time, finance, and strategic plans design, technology installs, Big Data analysis, and visualization (40 hours in total). This was specifically helpful to follow the data technology part and reveal data traps. Further, during the five months, the co-author participated in all phone calls between the technology provider and the firm (1,5 hours in total), attended all meetings where results of Big Data generation and analysis were presented and discussed in real-time with managers of the firm (3 meetings, 4,5 hours in total). Secondly, the co-author was put in a copy in all email communication between the technology provider and the firm (overall, 50 emails). The flow of incoming and outcoming documents was collected and screened in real-time, including plans, the project strategy, KPIs, goals, raw and visualized Big Data. Thirdly, four in-depth semi-structured interviews were conducted with the firm's three managers (5,2 hours in total). The interviews were conducted after the four main parts of the project in the firm's office, specifically, initiating and planning, two assessments, final report meeting. It became possible to get more informal information on project flow and technology performance. One extensive interview with the technology supplier firm representative was conducted (1 hour in total). The dataset comprehends 26,5 hours of the transcript. Overall, it allowed getting the firm's natural and real-time reaction on data, results, new technology, and understanding the cognitive aspects. Therefore, one co-author was involved in data collecting.

Phase III and IIIa were dedicated to the data analysis and were performed simultaneously. As the research analyzes both sides of DDD, data and cognition, I decided to add qualitative analysis quantitative elements to derive insights using the same contents. Thus, firstly, all three co-authors interpreted results by qualitatively analyzing contents following the Gioia et al. (2013) approach for inductive data analysis. Specifically, researchers qualitatively interpreted large datasets, including observation field notes, managerial discussions during meetings, interviews, and emails by continuously and iterative reading and deriving key topics, aspects, and items that were guiding the decision-making of the managers. I firstly performed open coding and derived initial topics grouped in the categories. I then performed detailed coding—searching for connections between first-order topics based—and derived high-order themes. Finally, I aggregated high-order topics into the overarching dimensions. The data structure is presented in Figure 2.

Additionally, in Phase IIIa the co-author who conducted data collection run a quantitative analysis. Specifically, by running sentiment analysis on the same content of interviews and emails used for qualitative research, I derived topics related to positive, negative, and neutral emotions. Sentiment analysis is a new text mining technique that identifies the polarity, namely, positive, negative, or neutral sentiment of text using Natural Language Processing and categorizing text based on the sentiments into categories (Pandey et al., 2017; Moe & Schweidel, 2017). Moreover, although very new, this method has already been used in technology innovation research as the new method to assess a technology customer value perception (Yoon et al., 2020). The research uses a specific approach to derive the sentiment based on non-lexicon methods and employs unsupervised machine learning of the Python programming language.

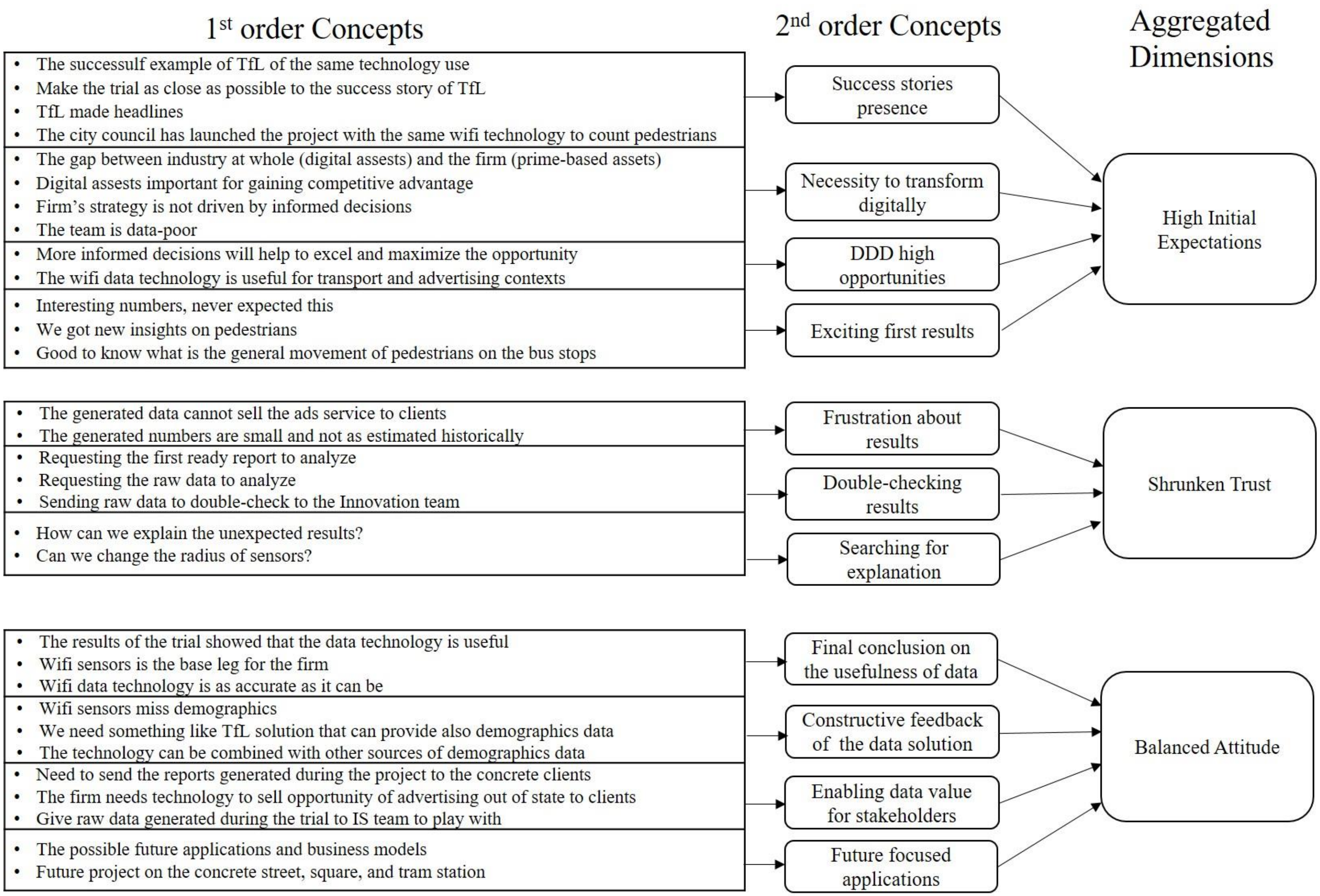

Figure 2. Data Strucutre

2.4. FINDINGS

The three aggregated dimensions of the qualitative content analysis are related to the trial phases. Moreover, they form the so-called zones with the defined order, beginning, and the end (Figure 3).

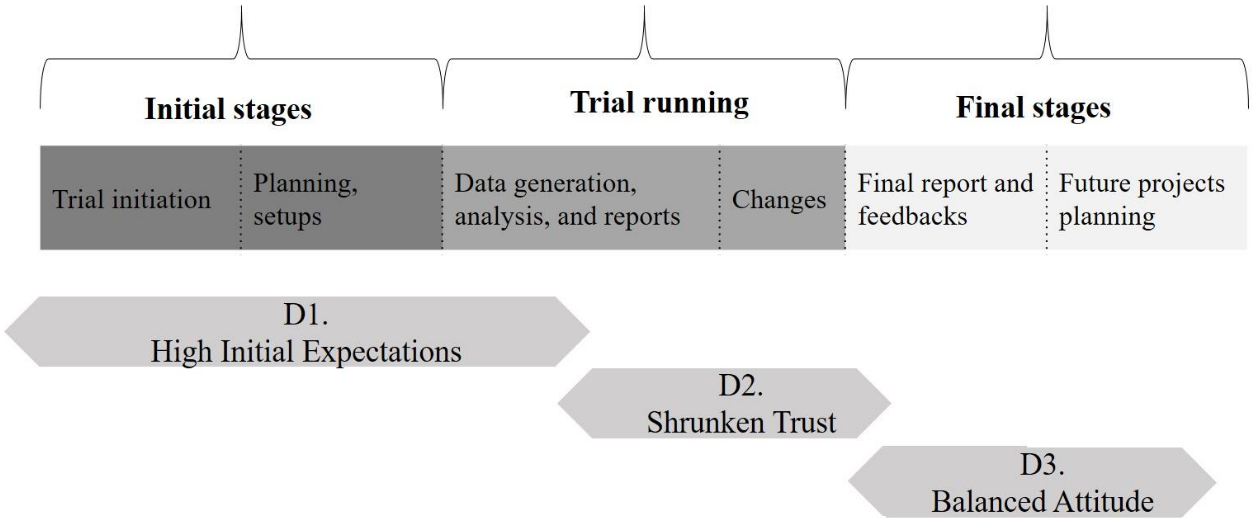

Figure 3. Mapping of the Dimensions on the Trial Timeline

Moreover, the three dimensions are unique in their nature—they disclose different degrees of trust in data. The first dimension (D1) that was observed at the beginning of the project indicates a very positive managerial attitude to data and its ability to support decision-making. The second dimension (D2) that includes contents of the trial's actual running shows a loss of trust in data-driven results. While the final stage of the project was characterized by the grown trust in data again, which is possible to spot in the third aggregated dimension (D3). In the next sections, I go deeper into investigating what caused those fluctuations of trust and what was the main driver of the decision-making in three different zones based on the ethnographic data and participatory observations. Additionally, to double-check the findings on the three different and time-related sentiments occurring during the project, I performed sentiment analysis.

2.4.1. Traps Zones

2.4.1.1. Zone of High Expectations or Cognition Traps Zone

The author observed that the project's initial stages were characterized by a very strong positive position towards data technology. However, as revealed, the decision-making was influenced by cognitive biases, namely, *herd behavior* and *information avoidance traps*. Therefore, The author labeled the situation as a cognition trap zone.

In this very case, the *herd behavior trap* was grounded on two main motives. First, the feeling that everyone in the transportation industry is digitally transforming, and second, the success story of the company's shared identity. Thus, the perception that almost all industry players are performing or have already performed digital shift was expressed during interviews. "*You can see that in the industry is already a lot of that stuff anyway, and yeah, it is happening there, and we're quite interested in this*" (Commercialization Delivery Manager, interview). Not only the general feeling but also a clear understanding of why, what, and how exactly to change: "*a lot of our assets are prime-based, whereas in the industry now a lot of industry's assets are digital and they can generate a lot more revenue. As for advertising in digital, we've got a look a sort of strategy on trying gain as more digital assets as we can, so we can compete on that market*" (Commercial Sales and Sponsorship Manager). Moreover, answering the questions on why the firm decided to test specifically WiFi sensors to generate data, managers answered highlighted that many technological alternatives are not necessarily stable on the market and immediately added: *"we know that the [...] City Council have used WiFi sensors to gather analytics on people's movement trends*." (Commercialization Delivery Assistant Manager). This was the first sign that the herd behavior took place in the case under investigation. Moreover, herd behavior trap involved the presence of the success story from partners, Transport for

London (TfL), who adopted the same technology in the same advertising context. TfL is the leading London-based company in the transportation sector, and it sets the trends for the whole transportation sector. Indeed, TfL " *[…] in terms of automation systems and technology […] is seen as a leader."* – said the CEO technology provider firm. The firm under exploration and TfL share main features, such as working in the same transport sector, operating in big cities that are industrial hubs, having the same market issues, users, and assets that need to be used most efficiently (transport, advertising assets). Thus, not only the effectiveness of the employment of the data technology to drive decisions in advertising was proved from the very beginning; I observed that the trust in the data technology was reinforced by the shared identity aspect. Exemplarily, managers were literally calling the solution implemented by TfL the "*holy grail*" several times. In this vein, it was evident that this enthusiasm led to the initially inflated trust towards the new data technology.

The information avoidance trap was observed as managers were focused on avoiding disconfirming facts about data usefulness. The main logic driving this bias was that the managers lacked their own data experience while they were confident in their domain-expertise. The author discovered managers were using two tactics of information avoidance, firstly, biased interpretation of the information and, secondly, inattention. Thus, biased interpretation appeared following the two aspects. On the one hand, managers expressed a high level of trust in data and the improvements data could bring: "*I think, one of the challenges […] about planning assets, is that […] we don't know where to put things. Cause it's not driven by any informed process, so this type of insight […] is quite informative for us to then be able to position assets correctly rather than doing it on assumptions, which is what we are doing at the moment"* (Commercialization Delivery Manager, interview). *"We are data-poor; we need to be making* *more informed decisions on how to excel and maximize the opportunity*" (Commercialization Delivery Assistant Manager). On the other hand, managers had a very basic experience of working with data: *"[…] for us, we are low entry-level, we wow any data*" (Commercialization Delivery Manager). With this respect, we observed that managers interpreted the disputable situation in a way that confirms their trust in data. Specifically, data showed that the peak-time during working days in the specific location was not as estimated and for years perceived the only correct by the firm (8-10 AM instead of 7-9 AM). Managers interpreted this fact as the confirmation of the good performance of data technology. According to them, the not severe mismatch between data and experienced-based results only confirms that the general line of peak-times was correct. The data results were not questioned, although the previous experienced-based estimates were different. The peaky choice to interpret positively not expected data results became evident as, during the next assessments, the same fact on the peak-time was taken negatively. Indeed, later managers even requested the raw data to double-check the results.

Further, inattention was observed as managers avoided paying attention to the results that could disconfirm data potentiality until the same results were observed repeatedly. Specifically, data showed a very low number of pedestrians (12-3 PM, 115 pedestrians) on the train station where commuters' flow should have been massively more at this time. While managers did accept the fact of the mismatch ("*interesting number*" Commercialization Delivery Manager), they put the new fact in a "blind zone" and moved their attention to the discussion of the confirmative results. The inattention became very prominent for the research performing observations only at the end of the whole project, as after data showed the same results over again, managers started expressing doubts on the overall effectiveness of the WiFi data technology for advertising purposes; thus, the technology was reset up. This situation will be described in the next section.

2.4.1.2. Zone of Shrunked Trust or Data Technology Traps Zone

As the project moved on, The author observed a change in the level of trust in data. Managers were expressing doubts in data ability to drive the decisions. It became especially prominent not only because managers were expressing frustration and loss of trust, but also because managers decided to double-check data-driven insights using the third source of knowledge, the firm's innovation team, and sent all data outputs produced so far to managers who already had the experience of applying data into decision-making. The author found that the skepticism was based on the wrong data technology setups and wrong data inputs, which became evident only in the later stages of the trial. The project was mainly performed by the technology provider, including developing the three so-called "packs" together with KPIs, goals, and setups. However, several aspects, which at the first side seemed just technical setups, turned out to be strategic actions that influenced the decision-making later. Initially, the firm was quite vague even in describing their concrete project's goals. There was an overall understanding that *"this trial is while interesting, to try to inform some of the decisions that we make*." (Commercialization Delivery Manager). Comparing with the CEO of the technology provider goals description: "*We are trying to understand if we can create an offline experience to online metrics. So, we are trying to mirror footfall and offline analytics to street analytics*". Therefore, the data technology provider played the role of expertise carrier and influenced the data input and the way to interpret data output. This caused several data technology traps that I labeled as *data tuner trap*, *visual angle trap*, and *cause-effect trap*. Accordingly, I labeled the period when the traps became evident as a data technology traps zone.

The data tuner trap was grounded on the fact that a data tuner persona, which in this particular case was IoT firm manager, influenced data inputs and variables. Specifically, managers of the firm aiming at performing Digital Transformation lacked expertise and knowledge in data technology aspects and outsourced all technology set up decisions and the project's design plan to the IoT firm. Managers played a more passive role at the very beginning by mainly approving or rejecting specific steps. Thus, it was chosen to run the project only during specific time-periods (e.g., only mornings 8 AM – 10 AM), which were picked up by the technology provider intuitively as analyzing all peak and off-peak hours would have required time and resources. Further, the technology provider chose the locations for sensors based on their understanding of outdoor advertising services' specifics (i.e., the tram stop in front of the shopping mall was considered over a specific place of the main square). Thus, the TC's domain-expertise was not included in planning and performing the trial initially. Although being reasonable for the technology company, the locations would have been changed by the firm later. During the final interview, there was expressed even a concrete location where it would be good to run another pilot: "*I'd like to do one pilot in one-stop […], for example from […] Street or […] Square*" (Commercialization Delivery Manager). Further, the data visualization was performed manually by the IoT firm's staff, as the existing dashboard was not able to provide the level of data visualization needed for the advertising specifics. There were developed Excel templates to quickly run some data analysis and designed custom templates of what exactly to visualize on the mid-term reports and the final report. While data analysis models and templates were initially set up to provide cumulative numbers, e.g., the whole number of pedestrians for defined hours (e.g., 8 AM – 10 AM), it became evident that advertising specificity required a more precise trend line, where there are not only peak-hours but also all data per weekdays and the whole data history (30 days). All the aspects that were not initially recognized by managers due to the low knowledge of data technology and cognitive biases further led managers to frustration. On the one hand, the technology seemed to perform well and deliver results as initially set up. On the other hand, it took some time to realize that new data results were misleading anyway. As I observed before, to understand that the results were not correct and were produced being influenced by the wrong data inputs and variables setup but not by the data technology poor performance in general, the trust in data dropped. The resets caused time-waste to rearrange data analytics and visualization and big confusion in the perception of the overall data technology performance.

Visual angle trap appeared when data being utterly correct in general and for the transportation industry was inappropriate for use in the given advertising context. As in geography, a visual angle changes the perception of the same picture. Specifically, I discovered that both the firm's and the technology provider's managers having experience only in the transportation industry were not clear in distinguishing between two different streams, transport industry and advertising in the transport context at the beginning. Exemplarily, commenting on the level of data, which would be possible to generate with WiFi sensors, managers wrote, "*This info is both valuable for transport planning and commercial activities*" (email). Indeed, being aware of the need "*to look after our home advertising contracts and trying to extract value through them*" (Commercialization Delivery Manager), managers were confused with the overall transport context. Furthermore, being used many times for cities and transportation, the technology was set up mainly in the same way for outdoor advertising, as the overall context and even the target was completely the same: understanding pedestrians' behavior using the same algorithms, WiFi technology, data visualization. For example, the new context of use required a more precise definition of the sensors' working radius. As expected initially, the best radius to catch the number of pedestrians was 10 meters – this is what was estimated as the optimum visibility distance to measure advertising engagement rate accurately, but also this is the most commonly used radius to measure the number of pedestrians for transportation purposes. This occurred to be misleading information that almost caused the collapse of the whole project. Specifically, managers were frustrated by the too low number of pedestrians in the train station area, while their expectations and expertise were saying that the number of pedestrians should have been at least double higher. Indeed, after careful elaboration on what caused an unexpectedly low number of potential advertising viewers, managers of the firm suggested that advertising context amplifies different radius that just transportation context, the radius of sensors cover was enlarged, as the majority of the advertising dashboards being big enough is better seen from 60, and some of them even from 150 meters. Until the data started producing outputs in correspondence with the advertising context, the positivism about the data technology was low.

Cause-effect trap accompanied previously discussed data technology traps. The author observed, as data and applied algorithms were providing statistics and facts that had meaning for algorithms mainly, data *per se* was not able to answer the question "why" neither provides answers on the causes of some events. Thus, an unexpected peak-time period in one of the working days could not be explained by data. While managers accepted that results generated by data were inappropriate to base on their strategic decisions *("This data cannot promote our services,"* Commercialization Delivery Manager), they were not searching for causes that triggered not expected results. Without managers critically challenging the results, even meaningful and accurate data that were generated during the project could lead to the wrong direction, like in the case with 10 meters range and visual angle trap mentioned above. This situation did not last for a long time, however, managers started looking for explanations before taking the final decision, and this led them to the third zone described in the next section.

2.4.1.3. Zone of Balanced Attitude or Traps Recognition Zone

The further transcript analysis of direct participation revealed that while some of the results produced by data in cognition and data technology traps zone were questioned relatively in traps zones, they were more precisely examined later when they accumulated in the zone the author labeled traps recognition zone. The author observed that the managers recognized that both data technology and intuition were not performing well, and there are exact reasons for this. Managers initiated the changes in sensors' setups. Specifically, they requested to enlarge the sensors' radius, changed visualization options. Compare, the first report exhibits results of the 10 meters radius on the train station, while the final report already shows the radius of 60 and 150 meters, the whole visualization has been changed too. The author also observed that managers acknowledged that their intuitive estimates might be wrong, while data could provide new unexpected but correct insights. Specifically, the case of mismatch between expected and data-driven insights on the peak-time is discussed in the information avoidance trap section. Moreover, unlike at the beginning, where the trust in data was groundlessly high, now managers recognized the role of managerial cognition in DDD as the translator of the data results. "*Overall, data that we have seen is particularly useful to assess trends and peak flows and we are able to translate it into how we might better utilize our advertising assets*" (Commercialization Delivery Manager). Comparing with the initial situation when data were trusted unconditionally, managers now highlighted the need "*to make sure we calibrate our numbers, to make sure we are selling on an accurate data*" (Assistant Commercialization Delivery Manager) at the same time recognizing the value of data. Further, managers gave constructive feedback on the drawbacks of the data without diminishing its value for decision-making., *i.e.*, "*It is base leg data set for us […]. The only puzzle which is missing is demographics*" (Commercial Delivery Manager).

2.4.2. Changes of Trust in Data. Sentiment Analysis

To add to the analysis data-driven facts on the level of data trust, the author measured the changes in positivism degree during the project by analyzing the sentiment of the in-depth interviews and emails that produced managers. I found three positivism levels with a relatively equal degree that managers were expressing, namely, positive, neutral, negative, that is depicted on the heat-map (see Figure 4).

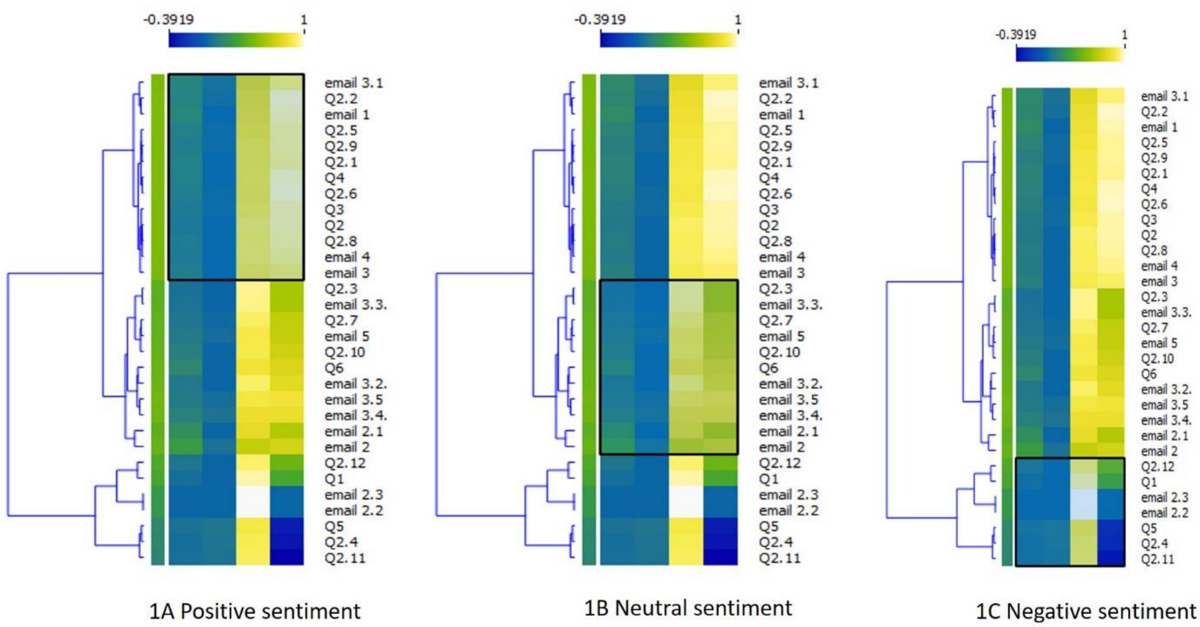

Figure 4. Clusters Based on Sentiment Analysis Scores

The analysis of all interviews and emails revealed that all three levels of the sentiment are almost equal in terms of the number of content items within each cluster. It can signalize the equal role of those emotions in the project too. Moreover, the set of specific items in each cluster was strongly time-related to the project. For example, positive sentiment consists mainly of content items (emails and interview answers) from the project's beginning. Negative sentiment is composed of items from the middle of the project, while neutral – from the end. In the next Table 7, I present the details on the correlation between project steps and sentiments degree. The compound feature summarizes all possible sentiments of the email or an answer to a question, therefore giving the mean sentiment of a specific content item.

title true	**Positive**	**Negative**	**Neutral**	**Compound**	**Project steps**
email 1	0.179	0	0.821	0.9709	Initiating
email 2	0.27	0.055	0.675	0.7263	Planning
email 3	0.117	0	0.883	0.9178	Planning
email 4	0.095	0	0.905	0.9445	Planning
email 5	0.094	0.031	0.876	0.6747	Planning
Q1	0.037	0	0.963	0.34	First meeting
Q2	0.089	0	0.911	0.9657	First meeting
Q3	0.098	0.019	0.883	0.9652	First assessment
Q4	0.131	0	0.869	0.974	First assessment
Q5	0.051	0.077	0.871	-0.3182	First assessment
Q6	0.163	0	0.837	0.7469	First assessment
email 2.1	0.203	0	0.797	0.6249	Between assessments
email 2.2	0	0	1	0	Between assessments
email 2.3	0	0	1	0	Between assessments
Q2.1	0.133	0	0.867	0.95	Final report
Q2.2	0.156	0.036	0.808	0.9786	Final report
Q2.3	0.05	0	0.95	0.5927	Final report
Q2.4	0.039	0.07	0.891	-0.3421	Final report
Q2.5	0.142	0.018	0.839	0.9511	Final report
Q2.6	0.121	0.013	0.867	0.9827	Final report
Q2.7	0.075	0.017	0.908	0.6705	Final report
Q2.8	0.101	0	0.899	0.9602	Discussing future options
Q2.9	0.118	0.02	0.862	0.9538	Discussing future options
Q2.10	0.116	0	0.884	0.7003	Discussing future options
Q2.11	0.04	0.057	0.903	-0.3919	Discussing future options
Q2.12	0.077	0	0.923	0.4457	Discussing future options
email 3.1	0.168	0.054	0.778	0.9272	Planning next tests
email 3.2.	0.086	0	0.914	0.775	Planning next tests

email 3.3.	0.06	0	0.94	0.5994	Planning next tests
email 3.4.	0.126	0.057	0.818	0.8094	Planning budget
email 3.5	0.094	0.037	0.868	0.8434	Discussing next steps

Table 7. Sentiment Analysis. Time-Line

The author will present the sentiments and content items following this project timeline logic. Specifically, the author found that several positive topics accompanied positive sentiment at the initial stage of the project. Firstly, the other company's success story employed the same technology within the same advertising for transport context. "*Below are a few screenshots from a system that was developed by TfL, Exterion Media, and Telefonica fusing various data sets – it's the holy grail in terms of audience data, we'd love to have access to this level of data*." Secondly, I found that managers were very optimistic about the data technology itself; this positivism was not yet connected with a personal experience of the data technology use. Phrases like "*We are low entry-level, we wow any data*" or "*I don't care what the solution is, is the data that is to improve. I think WiFi data for me is as accurate as it can be in terms of a raw count of people*" illustrate the situation. Finally, overall, managers expressed trust in data technology that should dramatically increase their advertising revenue stream with relatively low efforts. "*Think, with a bit of figure from us, developing a strategy and putting things digitally; you could see that [advertising revenue] grows 2-3 fold, 4 fold with a little bit of afford from our side*". Additionally, the sentiment analysis revealed a positive sentiment at the final stage of the project. It was connected with the potentiality of the data technology that was discovered during the whole project. Overall, the first cluster's positivism was expressed in the form of a high level of trust towards data and technology.

The revealed negative sentiment was reflected in doubts about data technology and the results it generated. Negative sentiment was observed mainly in the middle of the project. The project managers requested, "*Is it possible to get all locations for the last month?*" and "*Is it possible to get the raw data from the pilot in the interim*?" or "*I think we should send this one of our advertising partners to see. I'd send then this data set and I'd give them what has been done so far*". Finally, the ML algorithms reasonable indicated the statement on "*We have so many technologies nowadays, and we really do not know which one to invest in. They can die tomorrow*" this concern pointing out the necessity to try before to adopt data technology was the only negative sentiment message expressed at the beginning of the project. Overall, the level of negative sentiment was connected with the distrust towards results generated by data technology.

Finally, the neutral sentiment was observed mainly on the content from the final stage of the project. I did not find any polar negative or positive messages as in previous clusters. Instead, managers were confident in their expectations from the data technology as now it is based on their own experience and knowledge about data technology and experts' opinion. "*I have shared the cost information with colleagues who are reviewing [our company] holistic approach to data; the initial feedback is that we need to run some manual counts to validate the accuracy of the data before we can determine wider deployment of sensors. Once we start to receive data from the sensors at [the street], we can set up manual counts*". Managers were discussing technical aspects of future projects, expressing very clear goals rather than expectations. "*We don't need [data] constantly, we need the data to be able to quote what footfall is for one-stop or advertising side and then in six-month time we might check and calibrate our numbers to make sure we are selling on accurate data*" or "*For me, it is [...] a test to understand what level of data can be greened and to shape our strategy*". Therefore, the neutral sentiment is characterized by an informed trust in data and is observed mainly at the project's final stage.

2.5. DISCUSSION

By using an ethnographic approach, the present research explored cognitive and data biases when data are first introduced to improve decision-making and how they influence trust in data. I captured granular details on how both cognition and data drive decision-making by influencing trust in data. Specifically, I found that different types of traps zones drove managerial trust in data. The cognition traps zone that the author observed at the initial stages of the tests captures the strong trust in data effectiveness built upon someone else's data experience and defended by cognitive biases. I found that the data-based knowledge source is suppressed, as the firm relied more on intuition and beliefs than on new data results. Some intuition-based facts contradict data-based results, but managers do not recognize it due to cognitive biases. Data technology traps zone appeared after the cognition traps zone. I found that the level of trust in data effectiveness was low, as managers were misled by the data tuner persona, wrong data input, and data setups. The cognition knowledge source was suppressed. Data technology traps had a strongly-pronounced socio-technology nature. Thus, the IoT system and data it generated were initially programmed with algorithms designed for the specific case, set up to produce specific insights in the context of the transportation and later advertising industry, all these based on specific experience and expertise of the data technology provider and company testing the concept. Therefore, Big Data, which is supposed to be unbiased, could not be completely neutral as they were affected by human touch. Traps recognition zone followed data technology traps zone and was characterized by the acknowledgment of the two knowledge sources results, intuitive and data. In this zone, none of the knowledge sources was dramatically suppressed. Indeed, managers were reasoning, interpreting data results while questioning both their prior estimates and data insights. All these led to the growing trust in data again.

To the best of the author's knowledge, this is the first research that comprehensively explored the dual nature of biases, cognitive and logical, of data-driven decision-making with the focus on trust in data. Taking this into account, in this research, I contributed to the managerial decision-making and emerging data-driven decision-making streams within strategic management literature in several ways. Firstly, the research contributed to understanding the parallel-competitive theory (Hodgkinson and Sadler-Smith, 2018), working on the new phenomenon of data-driven decision-making. The findings showed that in the case of DDD, both cognition and analytics each independently participates in decision-making. Moreover, I also observed the intriguing synergy between cognition and data that occurred only after each type of decision-making contributed separately. This dynamic explored through the lens of biases results in the three traps zones. Thus, secondly, the research shed light on the role of cognitive biases that compose the cognitive traps zone. In line with the scholars, the researcher found that cognitive biases play beliefs defensive role and, firstly, influence decisions (Golman et al., 2017; Lallement et al., 2020). However, once cognitive biases lose their active defending function, they supported a better understanding of the technical aspects and applicability of data and activated the questioning mechanism for the new data-driven insights, not matching wanted beliefs. Thirdly, the study contributed to the emerging discussion on data biases that occur due to humans' behavior (Lindebaum et al., 2020; Choudhury et al., 2020) and introduced data technology traps zone related *data traps*, on the analogy with cognition traps. The findings suggest that "human-related" aspects can be understood as a strongly-pronounced socio-technology nature of data. Specifically, data has a strong dependency on a data technology tuner persona and industry specificity that might influence data input. Moreover, the findings showed that data results initially have meaning for algorithms and not obligatory have meaning for managers, as in the *visual angle trap*. It has been recognized that the lack of knowledge about the whole chain of the decision-making process based on data and the issue of outsourcing data to IS teams can influence negatively DDD (Waller & Fawcett, 2013; Baesens et al., 2016; Vidgen et al., 2017), the study crystallized how this situation can become a burgeoning milieu of cognition and especially technology traps occurrence. Finally, the study contributed to the discussion on the data trust for effective decision-making (Surbakti et al., 2020; Glikson and & Woolley, 2020). So far, data trust has been comprehensively discussed concerning the substitutional role of AI in decision-making (Glikson and & Woolley, 2020). The preset research breaks into an understanding of how trust in data works when data has a supportive role for decision-making. Specifically, the findings showed that data trust being influenced by cognitive or data traps relatively follows the positive-negative-neutral scenario. Specifically, the fluctuations I observed were strongly related to trap zones order. Thus, high trust was observed in the cognition traps zone, low trust in the data technology traps zone; finally, the neutral level of trust was observed in the traps recognition zone.

In this vein, the research bridges two streams of literature that is related to managerial decision-making, namely, managerial cognition (Busenitz & Barney, 1997; Haselton et al., 2005; Kahneman & Tversky, 1977; Zhang & Cueto, 2017;) and the emerging stream on data-driven decision-making (Baesens et al., 2016; Choudhury et al., 2020; Shrestha et al., 2020). The author suggests that managerial decision-making literature should consider not only managerial cognition but also data the pillars of strategic decision-making. The author argues DDD as the research stream should go beyond purely data technologies and Bid Data analytics scope (Bertsimas & Kallus, 2020; Camiña et al., 2020; Mandelbaum et al., 2020) and include cognition as the integral factor of enhanced decision-making. Data-based decisions are not free of cognitive biases as managers can fall victim to cognitive traps even being informed about them (Kahneman & Tversky, 1977) and being very positive about DDD in general (Mikalef et al., 2018; Surbakti et al., 2020) especially when first data technology advances are attempted. Moreover, DDD can have biases too that are caused by humans; this can add additional noise and influence the trust in data. Finally, The author contributed to the discussion on trust in data by introducing the importance of trust when data plays a supportive role in decision-making (Baesens, 206; Glikson and Woolley; 2020).

2.6. MANAGERIAL IMPLICATIONS, LIMITATIONS, AND FUTURE RESEARCH DIRECTIONS

Managers of traditional companies willing to employ data for enhanced decision-making for the first time will meet a specific challenge – estimating the new way of decision-making using new for data analytics and combining it with managerial cognition. This is the paradox of the new data and digital era. With this respect, they should be aware that the first steps in trying data-driven decision-making are characterized by low technology awareness and lack if no absence of domain-specific technology experience, which can fall into the cognition and technology traps. Firstly, the knowledge that cognition and technology traps could reinforce or even trigger each other will help managers set up preventive strategies to run more unbiased assessments and tests. Alongside this, previously made calculations based on the managerial intuition should not be set up as the success criteria for the new data technology, while constructing specific and not biased performance measures for Big Data analytics is particularly crucial (Mikalef et al., 2018). The better assessment strategy will be the employment of an external and reliable knowledge source to double-check both cognitive estimations and data results. Further, accepting that data is the new source of knowledge and it can confront intuition, will help to mitigate the consequences in case of the mismatch between those two. It is a good approach to use managerial intuition to explain the results generated by data technology and adjust the technology if needed. Moreover, managers, knowing that the level of trust in data performance differs from how data technology performs, can plan stages of assessment strategies in time, putting initially focus more on acquiring technology awareness to avoid snap judgments. Finally, trying to acquire as much knowledge as possible related to the whole data analytics chain before introducing data-driven decision-making will help to grow the trust in the data insights.

Although I performed quantitative sentiment analysis, the research has a mainly qualitative nature. Therefore, it could have some biases, such as the results could be too dependent on data and lack generalization. I focused on the situation in which managers already had positive expectations of data technology. Even though I believe that this is the bias many firms can fall into on the way to digital transformation (Bohnsack et al., 2018), the results might not work for the “skeptical” scenario. Moreover, the data technology employed in the case (IoT) might also be the subject of the specific traps. The research can also have cognitive biases in explaining cause-effects dependencies or the tendency to find facts in line with the author’s view. As I was focused on the limited number of cognitive biases that arose from the literature as the most suitable situation of the first data-driven decision-making trial, I believe that not all existing cognitive biases that suit the situation might have been reflected. However, as the research also has a pioneering nature, it can be the first immersive step to the emerging and fascinating topic of cognition and data's combinatorial power.

With this respect, for future research, I suggest exploring empirically other types of traps occurring while introducing data-driven decision-making and when data technology initially is perceived positively. Specifically of great interest might be exploring different cognitive biases that might arise in other contexts. Moreover, it would be important to investigate cases where initial expectations about the data technology and benefits it could bring are low. Empirical studies on more mature stages of data-driven decision-making employment within a firm could highlight a new understanding, e.g., what happens after the new data-based mode to inform decisions is adopted. Finally, focusing on exploring how cognition and technology traps influence the adoption of data-driven decision-making and why after successful trials, new data technology is not adopted.

3. BIG DATA INSIGHTS WANTED: DIMENSIONS, EFFECTS, AND MECHANISM FROM A PILOT PROJECT

ABSTRACT

The paper aims at opening up the black box of Big Data by taking a close look at their characteristics and effects that influence the ability of Big Data to create different types of insights. Yet researchers have mostly focused on the role of Big Data analytics in the insights creation, neglecting the better understanding of the underlying mechanism of different insights creation at the level of Big Data. Through a participatory observation research project the author disentangled Big Data dimensions and built a matrix that explains this mechanism via what the author defined as *Proliferation* and *Additive* effects. In this vein, the study provides a bridge between technology innovation and information systems literature and allows a better understanding of the dynamics of insights creation by focusing on Big Data dimensions and their fine-grained distinct effects. It also contributes to the discussion on the usefulness of Big Data. Finally, several managerial implications are highlighted and further discussed.

Keywords:

Big Data, Insights Creation, Big Data Dimensions, Digital Technologies, Proliferation Effect, Additive Effect, Participatory Observation.

3.1. INTRODUCTION

Big Data is the next milestone to advance innovation, performance, and competition (Chen et al., 2012; McAfee et al., 2012; Braganza et al., 2017; Mikalef et al., 2020). Existing research has shown that top-performing firms use data in their decision-making process five times more than low-performing firms (LaValle et al., 2011). Notwithstanding the significance of Big Data, recent research has started putting in question the given usefulness of Big Data for firms (Cappa et al., 2020; McAfee and Brynjolfsson, 2012). Yoo (2015) argues that having large volumes of data does not necessarily mean having interesting data, nor does it guarantee efficient decision-making (Williams et al., 2013; Vidgen et al., 2017), while converting data into meaningful insights is an effective use of Big Data (Chen et al., 2015; Mikalef et al., 2020; Johnson et al., 2017).

The phenomenon of Big Data has been primarily explored in terms of the different dimensions such as *Volume, Velocity,* and *Variety* (Khan et al., 2014; Gani et al., 2016; Hashem et al., 2016). Recently, the new tendency, especially with practitioners, has become adding new strategic-oriented dimensions such as *Veracity,* turning the focus of the discussion towards quality (Simsek et al., 2019; Urbinati et al., 2019). Moreover, recent studies have started questioning the importance of the "big" aspect of Big Data, arguing that the size side is an inclination that misleads researchers (George et al., 2014; Yoo, 2015). The research focus has shifted to those aspects of Big Data dimensions that influence insights creation (Yoo, 2015; Sivarajah et al., 2017; Larson and Chang, 2016; Conboy et al., 2020).

While several pieces of evidence show that Big Data dimensions influence the nature and quality of created insights (Aaltonen and Tempini 2014; Constantiou and Kallinikos, 2015; Yoo, 2015; Johnson et al., 2017), the mechanism behind this process remains mainly unexplored

(Ghasemaghaei and Calic, 2019). The majority of answers on Big Data insights creation result in exploring Big Data analytics technologies and algorithms (Sivarajah et al., 2017; Van Rijmenam et al., 2019; Dremel et al., 2020). At the same time, it has been recognized that Big Data Analytics is only a sub-process in the complex process of meaningful insights extraction (Gandomi & Haider, 2015). Thus, there exists the situation where Big Data analytics as the tool to create insights has been widely discussed, while Big Data has not been covered as the core component of Big Data insights creation mechanism. The exception is various research that explore Big Data as the socio-technological phenomenon that has an impact on businesses, industries, and society in general. Therefore, the prevailing statistically and technologically heavy lens used to look at Big Data insights creation complicates the understanding of the usefulness of Big Data for managers (Court, 2015; Yadegaridehkordi et al., 2018; Zhang et al., 2019). Indeed Big Data insights creation is a black box.

This paper takes the challenge and investigates the core mechanisms of Big Data insights creation and disentangles it on the level of Big Data dimensions. Based on a participatory observation research project of the pilot Active Travel Insights project launched in four European cities, the paper explores the role of Big Data dimensions and their characteristics to generate insights and build a conceptual framework that explains this mechanism.

The research makes several contributions. Firstly, it is the first research that sheds light on the mechanism of producing different types of insights based on the role of Big Data dimensions, sub-dimensions, and effects they produce, rather than Big Data analytics. The author found that all dimensions, namely, *Volume, Variety, Veracity,* and *Velocity*, contribute to the comprehensive quality of Big Data insights; it is not possible to claim that one dimension is more important than another for insights creation. That is in line with the existing view that a merely computational view on Big Data dimensions depreciates its strategic significance for insights creation (Dijcks, 2012; Schroeck et al., 2012; Yoo, 2015). Specifically, the author supported and enlarged fragmented research (Yoo, 2015; Sivarajah et al., 2017; Gandomi and Haider, 2015; Lee, 2017; Surbakti et al., 2020) on the quality side of Big Data dimensions by introducing the sub-dimensions notion. The author found that sub-dimensions shape the traits of the Big Data dimensions: an understanding of the *Volume* and *Variety* cannot be narrowed down to purely size and variety of data sources or *breadth*, as it is the *depth* of Big Data that enables more sophisticated insights (Yoo, 2015; Günther et al., 2017; Pröllochs and Feuerriegel, 2020). Moreover, *Veracity* and *Velocity* cannot be narrowed to the speed of data generation or the reliability of data sources only as they include other sub-dimensions that play an independent role in Big Data insights creation. Secondly, the author discovered that Big Data sub-dimensions produce two effects that work simultaneously at different degrees. This performs as the mechanism of the different types of Big Data insights creation. In so doing, the paper responds to the call for more research on Big Data insights with a focus not only on the issues of data analytics (Bharadwaj et al., 2013; Wessel, 2016; Vidgen et al., 2017) but also on Big Data *per se* and contributes to the discussion on the usefulness of Big Data (Johnson et al., 2017; Cappa et al., 2020). The research also supports the understanding of Big Data as the essential source to create relevant insights, and paying attention only to firm capabilities (*i.e.*, Big Data Analytics capabilities) could lead to a partial view of the overall phenomenon (Chen et al., 2012; McAfee et al., 2012; Wamba et al., 2017; Mikalef et al., 2020).

3.2. LITERATURE BACKGROUND

Big Data is commonly defined as large structured and unstructured data sets characterized by high volume, variety, and speed, that require the use of new computational techniques to discover trends and patterns within large datasets to enable better decision-making (Chen et al., 2012; George et al., 2014; Kwon et al., 2016). Although there is no clear definition of Big Data insights, there is an understanding that outcomes of the Big Data analysis are useful information, such as hidden patterns and unknown correlations (Chen et al., 2015; Waller & Fawcett, 2013; Vidgen et al., 2017). Therefore, in this paper, Big Data insights are conceived as an outcome of the extraction process of useful meaning from Big Data, which potentially leads to gaining firms' competitive advantage (Waller and Fawcett, 2013; Chen et al., 2015; Vidgen et al., 2017).

In the attempt to investigate the phenomenon, existing Big Data research conceived Big Data in terms of its primary features or dimensions, which the author summarized below in Table 8.

Dimension	**Studies**	**Conceptualization**
Volume	Laney, 2001; Chen et al., 2012; Gandomi, & Haider, 2015; Ghasemaghaei et al., 2018; Kwon et al., 2014; Al-Nuaimi et al., 2015	The magnitude of the size of the data. *E.i.*, Facebook stored 260 billion photos using a storage space of over 20 petabytes.
Variety	Laney, 2001; Gandomi, & Haider, 2015; Sivarajah et al., 2017	Structural heterogeneity in a dataset. *E.i.*, text, audio, video, numbers, images, weblogs, scientific data, user-generated content.
Velocity	Laney, 2001; Gandomi, & Haider, 2015; Sivarajah et al., 2017; Johnson et al., 2017	The speed at which the firm processes and analyzes data. *E.i.*, Wal-Mart, for instance, processes more than one million transactions per hour.
Veracity	Schroeck et al. 2012; Gandomi, & Haider, 2015; Sivarajah et al., 2017	Represents the unreliability inherent in some sources of data. *E.i.*, copying data with biases, imprecision, fabrications, messiness, and misplaced evidence.

Table 8. Big Data Dimensions Conceptualization

Starting from the seminal work of Laney (2001), a "3Vs" approach where *Volume*, *Variety*, and *Velocity* are the core characteristics of Big Data and quantitative view on Big Data was overriding (Chen et al., 2012; Kwon et al., 2016; Gandomi and Haider, 2015; Al Nuaimi et al., 2015; Ghasemaghaei et al., 2018; Urbinati et al., 2019). However, recently, researchers started focusing more on those aspects of Big Data dimensions that influence insights creation, highlighting the lack of knowledge on the impact of Big Data on creating insights (Ghasemaghaei and Calic, 2019). Thus, while *Volume* conceived as the magnitude of Big Data (Gandomi, & Haider, 2015) was mainly perceived as a purely quantitative dimension, this now has started being considered an inclination that misleads researchers (George et al., 2014; Yoo, 2015), as it came from the misnomer of Big Data, where "Big" reflects only the size of data. Indeed, large datasets *per se* do not make Big Data interesting and insightful (Chen et al., 2012; Yoo, 2015; Sivarajah et al., 2017), while increasing large-scale data opens up new opportunities for digging into the qualitative side of the *Volume* of Big Data. While, *Variety* that reflects the structural heterogeneity of a dataset cannot be underestimated in terms of data sources richness, as it leads to more sophisticated insights, the shift in the focus from just the number of Big Data sources towards the so-called granularity of Big Data insights was observed (Yoo, 2015; Günther et al., 2017; Pröllochs and Feuerriegel, 2020). Moreover, while *Velocity* is a crucial point in terms of the speed of data generation (Sivarajah et al., 2017) that makes Big Data different from statistics, surveys, and archival data sources that remain mainly static (George et al., 2014), the exponential growth of real-time generated Big Data has triggered the need in enhancing real-time analytics (Gandomi and Haider, 2015). Therefore, researchers pointed out the importance of exploring other kinds of *Velocity*, such as the speed of taking actions upon Big Data Analytics (Gandomi and Haider, 2015) or the issues of Big Data decay or the declining value of data over time (Lee, 2017). Furthermore, recent studies refer to the speed of Big Data analytics not only as a characteristic but also as an asset that is more important than *Volume* of data (Conboy et al., 2020), also highlighting the importance of fast data visualization (Larson and Chang, 2016). *Veracity* of Big Data becomes even more salient as the quality of insights is strictly connected with the quality of the analyzed data (Schroeck et al., 2012; Gandomi and Haider, 2015; Côrte-Real et al., 2020), the presence or absence of high-quality data in terms of timing and relevance (Sukumar and Ferrell, 2013; Lee, 2017; Ghasemaghaei, 2019), or technical point of view (Côrte-Real et al., 2020). It is believed that poor *Veracity*, which also might relate to poor data quality, might affect the credibility of generated insights (Surbakti et al., 2020).

Therefore, current studies have shown the shifts of the focus from the phenomenon of Big Data towards Big Data insights (Urbinati et al., 2019; Surbakti et al., 2020). There is an understanding that literally, any firm can benefit from Big Data as Big Data enables insights by extracting well-ordered information from unstructured data (Van Rijmenam et al., 2019; Vahn, 2014). Indeed, insights have the potential to transform key business functions and improve the competitiveness and market position of a firm (LaValle et al., 2011; Milliken, 2014; Chen et al., 2015; Gretzelet al., 2015; Erevelles et al., 2016; Akter and Wamba, 2016; Appioet al., 2019). Moreover, it has been witnessed that the modern data-driven mode of decision-making is outperforming the classic intuitive school. Specifically, LaValle et al. (2011) argued that so-called top-performers firms tend to apply Big Data Analytics in their key business activities, from financial management and budgeting to brand or market management, while low-performers mostly apply intuition as the key source of managerial knowledge. Big Data insights can support not only day-to-day managerial operations, such as supply-chain performance but also strategic decisions, better use of resources and assets, meaningful insights creation is also associated with business growth (LaValle et al., 2011; Chen et al., 2012). Moreover, Big Data being the source of new knowledge and insights can provide a certain level of managerial confidence to managers and help them to overcome emotional issues while coping with market uncertainty, especially in fast-changing and developing markets (Chen et al., 2012). However, the conditions where insights produced via Big Data can trigger changes across a firm are that Big Data insights should be related to a business-strategy, easy to understand and consume, and should be timely (LaValle et al., 2011; Wang et al., 2018). Effective insights are not only those created with meaning but also those created at a needed time, leading to fast decision-making. Therefore, concepts as “actionable”, “timely”, insights became buzzwords in defining the new practicability aspect of Big Data insights (LaValle et al., 2011; Chen et al., 2015; Sivarajah et al., 2017; Davenport and Harris, 2017).

Summarizing the discussed aspects, the existing theoretical background on Big Data allowed us to observe a clear move from exploring just Big Data dimensions *per se* towards Big Data insights (Bharadwaj et al., 2013; Vidgen et al., 2017; Wessel, 2016). Although current studies evidence that Big Data dimensions influence Big Data insights, the research is still fragmented and lacks understanding of how Big Data dimensions participate in insights creation. Therefore, there is a call for an in-depth and not abstract exploration of the role of Big Data dimensions in Big Data insights creation. To understand the dynamics, the researcher adopted a case study approach as it is more suitable to explore not only the inputs (Big Data dimensions) and output (Big Data insights), but also to look inside the black box of the Big Data within the real business setting.

3.3. METHODOLOGY

3.3.1. Research Method

As the present research aims at exploring the phenomenon of Big Data insights creation not from the conventional Big Data analytics side but the side of Big Data *per se* with the focus on practical aspects, the author decided to adopt a participatory observation as the main method (Bryman, 2012). Participatory observation is mostly focused on getting “knowledge for action” rather than “knowledge for understanding” (Clark et al., 2009) and, therefore, corresponds to the practical nature of the study. Moreover, as the research explores the phenomenon in a real business setting, it presupposes the involvement of various stakeholders from Big Data scientists to Managers, decision-makers, and analysis of various aspects that contribute differently to the insights creation. With this respect, the participatory approach that is recognized as an efficient method to understand complex situations and relationships and to achieve more significant and less hierarchical research practice (Clark et al., 2009) works the most to tackle this issue. Participant observation usually presupposes developing generalizable connections based on rich empirical observation (Van de Ven and Poole 1995; Street and Meister, 2004). The author drew on the three stages participatory research process developed by Street and Meister (2004) based on the participatory action research (Baskerville and Wood-Harper, 1998; Baskerville, 1999) widely used in the IS field when the deep dive into the practical aspects of the phenomenon is needed. Moreover, the three stages of research are especially helpful when there is a need to move from case observations to generalizable to a certain level of findings (Street and Meister, 2004). Table 9 presents the research design and summarizes the methods and approaches the author used in the research.

<table>
<tr><td colspan="2">Step 1: Learning about Big Data dimensions characteristics</td><td colspan="2">Step 1A: Learning about two Big Data effects on Big Data insights creation</td></tr>
<tr><td>Methodological tool</td><td>Description</td><td>Methodological tool</td><td>Description</td></tr>
<tr><td>Main:
Participatory observation</td><td>- Deriving the key points of the pilot move, changes, or issues connected with Big Data dimensions in real-time;
- Enabled to follow official milestones and spot natural reactions and key topics discussed in the formalized meetings.</td><td>Main:
Participatory observation</td><td>- Deriving the key points of the pilot move, changes, or issues connected with Big Data insights creation in real-time;
- enabled following the process of Big Data insights creation from the beginning until the end of the Pilot.</td></tr>
<tr><td>Supportive:
Slack channel content analysis</td><td>- An extension of the participatory observation;
- Enabled observation of non-formalized pilot dynamics between meetings, to spot stressing points for the project participants in setting up Big Data.</td><td>Supportive:
Slack channel content analysis</td><td>- An extension of the participatory observation;
- Enabled of observation of non-formalized pilot dynamics between meetings, to spot stressing points for the project participants in designing and generating Big Data insights.</td></tr>
<tr><td>Supportive:
Internal official documents analysis</td><td>- Analysis of unstructured data following the general inductive approach to analyze qualitative data.</td><td>Supportive:
Internal official documents analysis</td><td>- Analysis of unstructured data following the general inductive approach to analyze qualitative data.</td></tr>
<tr><td colspan="2">Step 2: Specifying findings into Big Data dimensions</td><td colspan="2">Step 2A: Evaluating two Big Data effects notion on Big Data insights creation</td></tr>
<tr><td colspan="2">Description</td><td>Methodological tool</td><td>Description</td></tr>
<tr><td colspan="2">Conceptualization of the specific findings on the Big Data dimensions into sub-dimensions.</td><td>Big Data Analytics</td><td>- Running Big Data analytics to create new Big Data insights using basic software; (Excel) and programming language Python
- Analysis of the created insights to explain the mechanism of generation of different types of Big Data insights.</td></tr>
<tr><td colspan="2">Step 3: Conceptualizing Big Data dimensions</td><td colspan="2">Step 3A: Conceptualizing the two Big Data effects into the matrix</td></tr>
<tr><td colspan="2">Description</td><td colspan="2">Description</td></tr>
<tr><td colspan="2">Deriving key similarities and differences between Big Data sub-dimensions and conceptualizing them two Big Data effects.</td><td colspan="2">Conceptualizing Big Data insights matrix based on the 1A and 2A stages.</td></tr>
</table>

Table 9. Research Steps and Methodological Tools

The author implemented two iterative in a certain way rounds of the participatory observation research, where the second round of participatory observations starts with the new gained from the previous round of knowledge while being focused on grasping new knowledge too. Each of the two rounds was composed of three steps. As typically, the first step of the participatory observation research aims at building a robust evidence base on Big Data dimensions, in step 1, the author used the triangulation principle of data collection and analysis (Yin, 2013). Specifically, the author triangulated participatory observations' notes with other data sources to enable reliable evidence-base and facts production. Thus, the author collected and qualitatively analyzed internal chat messages where all communication between three technological companies concerned the Pilot and data generation and a vast amount of official documents. This was particularly helpful to follow the overall flow of the project from its start to its end, to understand practical issues of Big Data concerning dimensions, enable collecting more pieces of evidence on sub-dimensions, and other aspects, which were not or could not be explicitly reflected during calls, meetings. In step 2, the author reflected and specified what the author learned during step 1 into the Big Data sub-dimensions notion. During step 3, the author moved to the abstraction and conceptualized findings into the two effects notion working on Big Data sub-dimensions. Specifically, the author derived key similarities and differences between Big Data dimensions by looking at their respective sub-dimensions and grouped four dimensions into two distinct groups. The author then challenged the similarities and differences by crosschecking them. Thus, the researcher conceptualized the findings on the sub-dimensions and derived two distinct effects.

The author repeated the research process using gained knowledge to grasp new aspects of the Big Data insights creation mechanism during the second round of the research. Thus, at step

1A the author was gathering data on Big Data insights creation using the notion of the two effects from step 3 and collecting new evidence base by triangulating participatory observation data with internal chat and internal documents analysis. During step 2A, the author evaluated findings from step 1A performing her own Big Data analytics to better explore the two effects notion on Big Data insights creation. Using raw Big Data of the Pilot and drawing on the conceptual basis from the second step, the author produced and analyzed different types of Big Data insights. The author applied both basic analytics (Microsoft (R) Excel) and more sophisticated analytics (Python programming language). Descriptive, exploratory, predictive, and time-series analyses were run. Finally, in step 3A the author conceptualized findings on the two effects into Big Data insights creation matrix.

The 3 stages framework enabled us to gradually move from participatory observations and learning from the case to conceptualizing findings. Although the steps of the research occurred subsequently, the boundaries were blurred, as the researchers were repeatedly going back to observations, documents analysis, and conceptualization. This reflective and iterative research approach was important not to lose crucial aspects and deal with the vast amount of the data collected.

3.3.2. Data Collection and Analysis

In steps 1 and 1A, a big amount of primary data based on participatory observations was collected. The researchers took part and followed one specific case of Big Data generation and Big Data insights creation from the initial phase till the end. Further, as a supportive knowledge source, the researchers got access to the internal chat logs (Slack App), where the main communication between the technology companies responsible for the installs and the whole Pilot design occurred. The Slack logs transcript took more than 85 pages. The author also collected internal official documents, including project milestones and deadlines, strategic plans, goals setting, monthly reports, final closing up report, and others. All data sources of the first steps are specified in Table 10. More than 200 pages were collected throughout the Pilot.

Data collection started in April 2019 and ended in August 2019 with the direct involvement of the authors as the Coordination Manager of one of the technology provider companies of the Pilot the author gathered ethnographic data through a period of 5 months. The author participated firstly in the action planning stage of the Pilot (April 2019) and later in the action taking the stage of the Pilot, which consisted of the sensors' location choosing, technology settings, dashboard designs, data generation, data analysis (May-August 2019). Finally, the author participated in the action evaluation that is reflected in the final report. Thus, the co-author attended weekly online calls (via Google Hangouts), real-life meetings of the Pilot Projects Managers and technology suppliers (London office). As the project was international, all meetings were held online, with some offline meetings between different stakeholders. The author documented formal main topics and the agenda of the discussion and meetings at all stages of the Pilot and planned and unplanned changes and issues of the installations, designs, data generation, data analysis, re-scoping the Pilot to explore the nature and characteristics of Big Data dimensions within a concrete business setting.

Method/Step	Data source	Source details	Quantity
Participatory observation	**Online weekly standup calls of technological companies**	During the calls, technology companies discussed the flow, timing of the project, project, technological, data gathering and analysis challenges, dashboard design, and development, urgent needs, mutual steps to perform, plans for the next week, and results of the previous week.	13 hours in total of 15 calls 20 pages of the notes
	Online standup calls cities and Project Lead	During the calls were discussed challenges, issues, needs, changes to implement.	3 calls for Helsinki 2 calls for Manchester city 10 pages of the notes
	Offline meetings: 1. Meeting with the Project Lead team during the weekly call 2. Meeting with the Funding Body team, the weekly call	1. London office of the Project Lead technology company. 1 weekly call meeting, discussion on the location choose and online visualization of the final 2. London office of the Funding body team. Discussion on the re-scoping of the pilot goals according to the new GDPR requirements expressed by cities.	1. 1.13 hour 2. 40 mins. 10 pages of the notes
	Visualization Dashboard development	Visualized Big Data both in real-time and historical data. The research was observing 1. The design and development process of the dashboard 2. The process of testing the dashboard 3. Changes implemented	2 months of the dashboard design and development
	Raw Big Data, CSV files	CSV files with all Big Data gathered during the project run. The researcher was observing 1. Volume, Variety, Velocity, Veracity dimensions, and their characteristics and changes reflected in changes of single variables, single data sources, timing, interruption in data generation, and other aspects.	WiFi sensors (Helsinki, Antwerp) CCTV (Antwerp, Helsinki, Manchester)
Qualitative content analysis	**Internal documents:** 1. Goals settings from cities' side 2. Scope of the pilot for all participants 3. Re-scoping document 4. Project management documents, GANNT charts 5. Project Risks analysis 6. Mid Term reports 7. Final report	1. The document indicated needs in launching the pilot of each city 2. The document with clear goals for each participant, technical specifications, data, and privacy protection rules and regulations 3. Changed Scope document according to the new GDPR requirements expressed by cities, GANTT chart 4. Managerial guidelines for the implementers of the pilot: milestones, phases, deadlines, KPIs, responsible persons 5. Document indicating technological and managerial risks of the pilot 6. Reports from the technology companies reporting reaching milestones – for funding body. 7. Final report communicating the final results of the project for cities: number of sensors, insights gained, participants, KPIs reached, deadlines met/failed, challenges and issues, suggestions for the future Big Data projects.	1. 3 pages 2. 15 pages 3. 23 pages 4. 13 documents (80 pages) 5. 2 excel files 6. 3 reports in average 15 pages each 7. 29 pages
	Internal communication chat Slack	All communication in the Slack App (internal chat) between all members of the technology firms involved in different positions (i.e., Managers, Data Scientists, Technicians). It served as the platform to communicate, plan, discuss issues, and share solutions.	4 months of gathering data 85 pages transcribed according to dates and phases of the Pilot

Table 10. Summary of Data Sources of the 1 and 1A Steps

As the data collected at the first step were heterogeneous in nature, the researcher developed an ad-hoc data analysis protocol based on the general inductive approach for qualitative data analysis (Thomas, 2006). The approach is especially suitable for condensing various raw contents into a summary (Thomas, 2006). The protocol aimed at helping, firstly, to unify in a single format facts and knowledge gained through participatory observations and texts, secondly, to ease follow-up analysis of the vast amount of various data, thirdly, to derive key practical aspects that are related to Big Data dimensions and that occurred during the whole period of the Pilot run. Specifically, the authors were rigorously and systematically reading all contents familiarizing with the context and some hidden aspects. Then, following the protocol format, the author systematized contents from meetings, calls, Slack daily messages, reports, and plans focusing on Big Data dimensions (step 1) and insights creation (step 1A). Final items were screened to triangulate data sources and group pieces of evidence in logic and concise blocks related to the specific issue or event. Overall, more than 200 items of the Pilot were produced, the final 52 groups were derived (see an example in Appendix A). Data analysis at the first step of the research was a reflective and iterative process, as managers and decision-makers did not mention directly Big Data dimensions, therefore the researcher was ought to find hidden but important aspects of Big Data dimensions. The reflective practice and participatory observations coupled with the expertise of the author in participating and coordinating Big Data pilot allowed to derive the key points of the project move, changes, and issues connected with Big Data dimensions and build a large empirical knowledge base on the fine-grained sub-dimensions and Big Data insights.

The step 2A included mainly analytics of Big Data generated during the Pilot. Thus, the author had access not to all raw Big Data (all raw Big Data sets were displayed on the dashboard) but specifically data generated in three cities (Manchester, Helsinki, and Antwerp) during one month – August 2019. CSV files with the raw Big Data consisted of single data variables, such as geolocation, date, time, category, and a number of road users, and single data sources. Using only available raw Big Data the author produced and analyzed different types of Big Data insights. Using both basic analytics (Microsoft (R) Excel) and the Python programming language.

3.4. ACTIVE TRAVEL INSIGHTS PILOT STUDY

The present study examines the Active Travel Insights (ATI) case, a large-scale pilot project that amalgamated different data feeds from various sensors in four European cities: Antwerp (Belgium), Eindhoven (Netherlands), Helsinki (Finland), and Manchester (United Kingdom). The ATI project aimed at providing evidence for each municipality for new cycling infrastructure investment. More specifically, ATI aimed at helping cities gain a deeper understanding of their cycle network, including how many people use cycle routes, deep insights into the interaction between different types of road users, the impact of the road network on the environment, and where citizens are traveling to and from. The pilot project also aimed at educating cities to use Big Data and Big Data technologies for building long-term strategies.

The Pilot, a multinational project, involved different stakeholders and participants. The funding body, a London-based firm, had the controlling function, the Project Lead with managerial function, three technology providers (Wi-Fi sensors technology company, CCTV cameras technology company, environmental sensors technology company) responsible for the whole Pilot planning and implementation, and four European cities municipalities with the customer role. Figure 5 presents the three Internet of Things (IoT) technologies that produced different types of Big Data in real-time and accumulated historical Big Data. Furthermore, there was tailored the dashboard with the use of API of technological companies, open-source data managing frameworks. The dashboard enabled raw Big Data accumulation and visualization either in real-time (the new data are displayed every 15 minutes) or in historical perspectives (daily, weekly, or defined time-period).

TYPE OF A SENSOR	**WIFI SENSOR**	**CCTV CAMERAS**	**ENVIRONMENTAL SENSORS**
FUNCTION OF A SENSOR	Origin-Destination Monitoring	Classified Counts	Environmental monitoring
DESCRIPTION	Sensors search for WiFi signals from mobile devices to help to visualize how citizens move around a city. Sensors use Machine Learning algorithms to catch unique MAC number of a mobile device, track origin and destination, time spent in an area of a mobile device within the radius of 300 meters.	Sensors can classify between nine types of road users, including cyclists, pedestrians, vans, tracks, cars, buses. The sensors use Machine Learning algorithms to recognize, count and track different types of vehicles within the field of view.	The sensors measure air quality (CO, NO, NO2 measured in parts per million (ppm)) and temperature.

Real-time data + Historical data

Figure 5. Types of Big Data Technologies in ATI Pilot

The smart mobility and better quality of life context of the ATI Pilot was one of the criteria to choose the case. Smart mobility is an exponentially growing direction, and there is a big need for smart extraction of data on how people move and their road behavior (Uras et al., 2020). Indeed, Big Data, Artificial Intelligence (AI), and Internet of Things (IoT) are playing a crucial role in turning cities into Smart Cities and the application of these technologies will be only growing in the foreseeable future (Vaidya et al., 2018). However, on the one hand, IoT technologies that produce Big Data are one of the most commonly used technologies to generate

Big Data, not only in smart cities context. On the other hand, IoT produces various types of data distinct from, for example, Big Data from social media or medical systems. The ATI Pilot used three IoT technologies that were generating data in different forms and shapes in real-time but also accumulating historic data, exemplarily, classified counts, location of road users, origin-destination matrix, new/returning visitors, dwell times, air quality measurements, video, and images. This made the ATI Pilot especially interesting to take as the basis of the research.

The participants of the ATI Pilot were another important reason to choose the case. Thus, although, there were data scientists in the team, the main stakeholders guiding the goals of the Big Data use, assessing the accessibility, quality, and usefulness of Big Data insights created were municipalities' managers and decision-makers. The latter was interested in the real applicability of Big Data insights but had little experience in working with Big Data. Therefore, this was particularly helpful to derive key aspects of Big Data dimensions related to insights creation on the managerial and decision-making level. Moreover, the large-scale project allowed exploring different degrees of insights created, from simple to complex ones. Finally, the pilot nature of the case made it possible to investigate the insights creation process and pitfalls from the very beginning, from goal settings and installation of sensors, choosing the correct data analytics techniques to the visualization of the results.

Following the literature background, the author observed the four Big Data dimensions, namely *Volume, Variety, Velocity*, and *Veracity,* as the four buckets that the author opened up to explore in-depth characteristics and features that shape these four dimensions. Based on the triangulated data analysis, the author disentangled dimensions into respective sub-dimensions exposed in the next session.

3.5. FINDINGS

3.5.1. Big Data Dimensions and Sub-dimensions

Among the Big Data dimensions, the dimension of *Volume* identifies the quantity of data. By analyzing and comparing raw and analyzed Big Data sets, the author found that a large amount of Big Data is produced via the process of data fusion[2] that allows the generation of new data from existing datasets. The process of fusion occurs not only when two or more single variables are combined but also when mixed or single variables are fused, thus, proliferating the volume of data. Both single and fused variables can include real-time and historical data, thus, ultimately, increasing *Volume.* The data fusion process delineates two sub-dimensions, the *breadth* and the *depth* of *Volume*. The *breadth* refers to the number of single variables available (e.g., time, date, pedestrians, geolocation, in/out, car, van, etc.), whereas the *depth* reflects the number of fused variables (e.g., number of pedestrians in specific geolocation in specific date and time; air pollution level compared with the traffic flow in specific geolocation and date; trend line of cyclists over a weekend, etc.). The importance of having big enough *Volume* of data in terms of *breadth* was mentioned several times during the Pilot, for instance, "*Will the dashboard have historic data on AQ sensor?*" or "*This should then allow us to build a picture of what [...] additional datasets can be brought in to supplement the dashboard*" (Slack channel). However, the *breadth* of *Volume* dimension was perceived by data scientists involved in the Pilot as raw data that serve to produce more *depth* by fusing data. "*Can the metadata such as lat/longs, descriptions, photos, ID numbers be sent through for everything [...] that was installed, so that we can start to build the databases required to feed into the marketplace?*" (Slack channel). Indeed, the significance of the *depth* was especially highlighted. As stated at the planning stage

[2] Data fusion is the process of integration and processing of different data sources or variables by applying analytics algorithms after the first data sets generation.

of the Pilot, "*It's at this point that we can begin to build an understanding of what we are looking to achieve and display as an amalgamation of the different datasets"* (Slack channel). The author observed that data fusion helps to transform a large amount of not connected datasets into one single insight, such as types of road users on roads, air pollution level, or origin-destinations of pedestrians, as well as into more complex insights picture, such as the correlation of peak-times of motorized vehicles and air quality. Hence, on the one hand, the *breadth* provides a broader scenario of the phenomenon under investigation by increasing the number of single variables, while, on the other hand, the *depth* offers a more specific and synthetic picture.

Variety dimension identifies the number of different types of data sources (*i.e.,* Wi-Fi sensors, CCTV cameras, environmental sensors). The author found that similarly to *Volume, Variety* grows with data fusion and is characterized by two sub-dimensions, *breadth* and *depth*. Here, the *breadth* refers to the number of single sources[3] available, whereas the *depth* reflects the number of fused data sources. For example, Wi-Fi sensors are a specific data source that produces one type of output, nevertheless, if Wi-Fi sensors are combined with other types of data sources, such as CCTV cameras they can be considered a hybrid data source that produces more in-depth outputs. In this case study, the author noted that the datasets on the number of motorized vehicles and the level of environmental pollution separately give only numerical insights, while these combined datasets provide a deeper understanding of the air pollution at a certain area and at certain a time-period. While the importance of the *Variety* of Big Data sources *per se* cannot be questioned, the author observed that Pilot participants were paying particular attention to data fusion as something even more valuable. "*Having data analysis gathering Wi-Fi, CCTVs, and Air quality sensors [...] would be really valuable*" (Slack channel). *Or "As a basic example [of*

[3] For example, a single Wi-Fi sensor in a road can be considered a single data source and as such it can be combined with other data sources of the same or different type.

the amalgamation of datasets] the open-source events and Wi-Fi data combined could provide an interesting 'event popularity index'" (Slack channel). Moreover, by analyzing raw Big Data the author observed, the higher the number of data sources is, the higher the *breadth* of *Variety* will be. Analysis of already fused Big Data revealed the consistent pattern, the higher the number of hybrid data sources is, the higher the *depth* of *Variety* will be.

Velocity is conceived in terms of the speed of the key Big Data processes. While, following the literature, *Velocity* reflects the speed of Big Data generation, the author observed that three other characteristics compose the *Velocity* feature. Specifically, the author delineated four sub-dimensions: *data generation speed, data fusion speed, data visualization speed,* and *data use speed*. Indeed, the analysis of the working group meetings and slack channel conversations reveals that *data generation speed* was crucial as data gathering regards actual and relevant data on the road traffic participants and environmental conditions; sensors were producing data constantly with the speed of 3G/4G connection[4]. It was also important to generate as fast as possible enough fresh data to analyze the situation on the streets from historical perspectives and to build a comprehensive evidence base for infrastructure investments. Mainly, for this reason, the dashboard was programmed to show also historical data from the beginning of the project.

Furthermore, the author found that the speed of *data fusion* and *data visualization* is as important as *data generation speed*. Either historical or real-time data were fused and visualized automatically in real-time, using API models and schemes. The author unveiled that automatic data processing enabled producing insights from historical and real-time generated data at a higher speed. The author found that several times during the Pilot technology providers were referring to ready data models as helpful tools while working with Big Data. Although not

[4] 3G/4G was the highest quality of connection available at the moment of piloting.

directly, this dynamic is also highlighted in the ATI middle report in terms of the importance of the API and data analytics models: "*Integration into the API is very simple. It's just a matter of taking the data from our sensors, converting it to a data model, and pushing the data to an instance with a simple HTTP request*". Technology companies highlighted the special role of APIs in the Final Report "*having standardized data models, and a standardized API for pushing and pulling data simplified things greatly and saved a lot of time*." Furthermore, fast *data visualization* was considered by all participants one of the most crucial parts of the Pilot. The high speed of *data visualization* was critical, as it meant immediate access to insights to decision-makers. The author discovered that the role of the dashboard and *data visualization* was equalized to the role of Big Data *per se*. Exemplarily, "*before the pilot, [cities] did not have consistent data on travel behaviour along the routes of interest. The dashboard created as part of this project provides each city with insights into how the road space is used by both motorised and non-motorised modes of transport*" (Final Report). Thus, the speed with which real-time outputs were visualized on the dashboard was every 15 minutes, close to real-time. This speed was optimal to present data without overloading the dashboard, providing a picture in real-time. The last sub-dimension, *data use*, refers to the rate of data useful time or data "best before." As the Pilot did not presuppose the immediate use of gained insights, the author could not fully capture the data use facet. However, the Final report allowed us to conclude that the ATI solution will make an influence in the long run if insights are used immediately. Specifically, from an economic point of view, the use of the produced insights can influence "*Reduced operating costs for the municipality*" and "*Improved service efficiency for the municipality*", as well as "*Travel time savings, Reduced congestion, Increased local spend in the economy, Increased tourist attractiveness and spend*".

Veracity dimension reflects the degree of reliability of data that is composed of different aspects. While all technologies were proved to be a very trustable source of data and technology providers were selected to participate in the Pilot as reliable partners—therefore, the credibility of the Big Data sources was not questioned—the author identified additional steps performed by the Pilot managers to avoid misleading data. More specifically, the author identified five main critical sub-dimensions enabling trustful Big Data, namely, *targeting, representative sample, technical specification, up-to-date-data,* and *error-free data manipulation.* Thus, *targeting* regards the stage before the actual piloting and it was focused on the specific goals of each city. For example, in the case of Helsinki, the ATI City Evaluation Report pointed out that "*their aim is to reduce traffic-related emissions by promoting cycling and non-motorized forms of transport. They want to do this through the use of real-time data and real-time traffic control*". The author observed that this target dictated technology to use, locations (roads) of sensors to install, time of the Pilot running, and real-time data generation or historical data analysis. However, the real needs of cities became narrower and more specific with deeper engagement in the project. Thus, in the anonymized questionnaire for cities run by the technology companies, cities indicated three main needs: "*Get empirical evidence on how the tech works. Have some data on what % of people have Wi-Fi on (compare Wi-Fi-sensor data with the camera-feed data)"; "more insights in travel behaviour; average speed, amount of cyclists, preferred routes, etc. with the main focus on cyclists; to inform the cycle infrastructure planning*".

Furthermore, to enable *Veracity* of data, a *representative sample* that accurately represents the population should be analyzed. In the particular case, sensors positioning, which was essential to provide a *representative sample,* was chosen carefully. Specifically, routes included those that "*connect businesses to residential hubs, new residential developments, or* *between major transport hubs*” (Scope of the Pilot). Indeed, choosing locations that would “*enable correct sample*” (planning meeting of technology providers) was the key target at the initial stage of the Pilot. The time for each city for the “*acceptance of proposed locations*” (Timeframe document per each city) that were chosen by technology providers was two weeks and more. Therefore, for the sake of a representative sample, even initial carefully chosen locations and supposed by technology companies to be the correct ones were changed in some cities. As indicated in the second Midterm Report from technology companies, “*due to requests from [two cities], the function and location of the [Wi-Fi] sensors in these cities does not match our original proposal*.” When it comes to the *representative sample*, data saturation level is an important aspect. For example, based on the expertise of the technology’s providers and the average Pilot’s time, it was decided to gather data for two months to achieve the desired data saturation level for the Pilot aimed more at introducing new technology for enhanced decision-making to cities.

Surprisingly, the author found that respecting *technical specifications* was another key point affecting *Veracity,* as it enables the constant gathering of good quality data. The author observed that at the initial phase of the project design, special attention was put on preventive measures to avoid technical issues that might negatively affect results. Specifically, the consortium of technology companies made all possible to prevent technical errors by creating Install Risks Register plan per each city, assessing potential issues and possible solutions, as well as the likelihood of an issue to occur and its potential impact on the project. For instance, Helsinki was indicated as having the lowest likelihood for the issues related to sensors functioning. However, if they occur, they will affect the project negatively at the highest grade; the preventive solution was to test sensors before installs. The author found that being in control of all sensors' status in different countries and being able to spot an error and respond quickly was crucial for the Pilot. As the pilot specialist said: "*It would also be useful [...] to be able to query the status of each device to determine if it is up, down, whatever useful feedback. Could you include this in Product Board*?" (Slack Channel). Thus, sensor-information API was included to react in real-time in case of technical issues, such as unplugged sensors in different locations: "*Antwerp sensor 4 is offline. Can someone check power, please?*" (Slack Channel). Moreover, even errors during installs might influence the results. For example, CCTVs have to be installed at a height between 4m and 8m, with a clear view of the road from the side of the carriageway as the best position to not lose any road user. Moreover, power and internet connectivity were fundamental for the whole project as well. Once a sensor is disconnected, some data are missed, as happened with one city during the project.

The fourth sub-dimension of the Big Data Veracity is *up-to-date data*. The level of "up-to-date" is unique for each use-case; thus, the author observed that using real-time data for the Pilot was crucial as analyzing road traffic even from one year ago could have caused radical bias. At the same time, two-month-old historical data visualized on the dashboard were as well up-to-date for the project.

The fifth sub-dimension is *error-free data manipulation* that refers to (un)intentional error that occurs while analyzing. This situation might deal with the use of wrong algorithms and models, especially when taken from open source libraries. Analysis of the internal channel Slack content showed that the discussion around choosing appropriate data models and APIs was a key topic. "*I think they are looking for some certainty on what the outputs from the Wi-Fi API will be*" (Slack channel). As in ATI were used pre-developed open-source data models, they had to be tested to work with traffic efficiently, the crowd flows, and environmental counts within the transportation context. Specifically, the validation service provided by the funding body indicated where and how many times specific data models were used and validated in other pilot projects[5]. Moreover, the APIs were initially developed to apply for data generated by IoT and sensors. This allowed more precise and error-free data manipulation. Finally, analytics models and algorithms were set up following the mobile penetration rate in countries, including the error rate on the number of people who might have switched off mobile devices or turned off Wi-Fi search signal (for Wi-Fi sensors).

By zooming in Big Data dimensions, the author was able to derive and classify their related characteristics that work as sub-dimensions. The author built a framework that represents Big Data dimensions and their respective sub-dimensions (see Figure 6). Thus, Big Data dimensions and sub-dimensions have a hierarchical nature. In the next section, the author is going to conceptualize how and to what extent the discovered sub-dimensions shape the traits of Big Data dimensions.

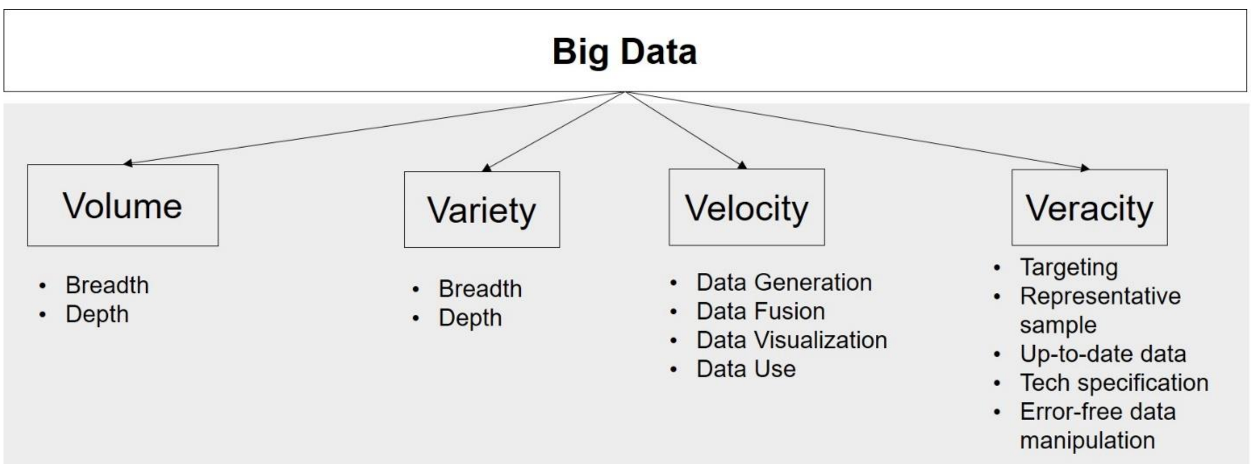

Figure 6. Dimensions and Sub-Dimensions of Big Data

[5] https://validation.services.synchronicity-iot.eu/table/

3.5.2. Big Data Effects

Using a new knowledge base, the author conceptualized the findings on the sub-dimensions presented above. Thus, the dimensions *Volume* and *Variety* have the same sub-dimensions, namely, *breadth* and *depth,* which cannot work for the other two dimensions. In its turn, *Velocity* and *Veracity,* while not having the same sub-dimensions, still share common features. Specifically, sub-dimensions shape the overall traits of their head dimensions negatively or positively. This prompted us to go deeper into the shared similarities and differences between Big Data dimensions and conceptualize them into two effects, namely, *Proliferation* and *Additive*. Once discovered, the two effects were crosschecked to find any possible rejecting the conceptual findings facts.

Specifically, the author observed that *Volume* and *Variety* can generate a *Proliferation* effect (see Figure 7), which the author defined as an exponential growth of Big Data that goes through a process of data fusion boosted by division and recombination of different levels of data source(s). Indeed, division and recombination are two core properties of the *Proliferation* effect that distinguishes it from the *Additive* effect. The *division property* refers to the ability of Big Data to be split into autonomous portions of data that can be manipulated for producing new output. In its turn, *recombination property* refers to the ability to recombine various autonomous portions of data in a meaningful way. In other words, *Volume* and *Variety* grow or proliferate with the division and recombination of datasets. The *Proliferation* effect composes the ground for the *Volume* and *Variety* dimensions' performance, not only influencing the breadth of Volume and Variety of Big Data (number of data variables and sources) directly but also the *depth*, as with data division and recombination more sophisticated and synthetic outputs come out.

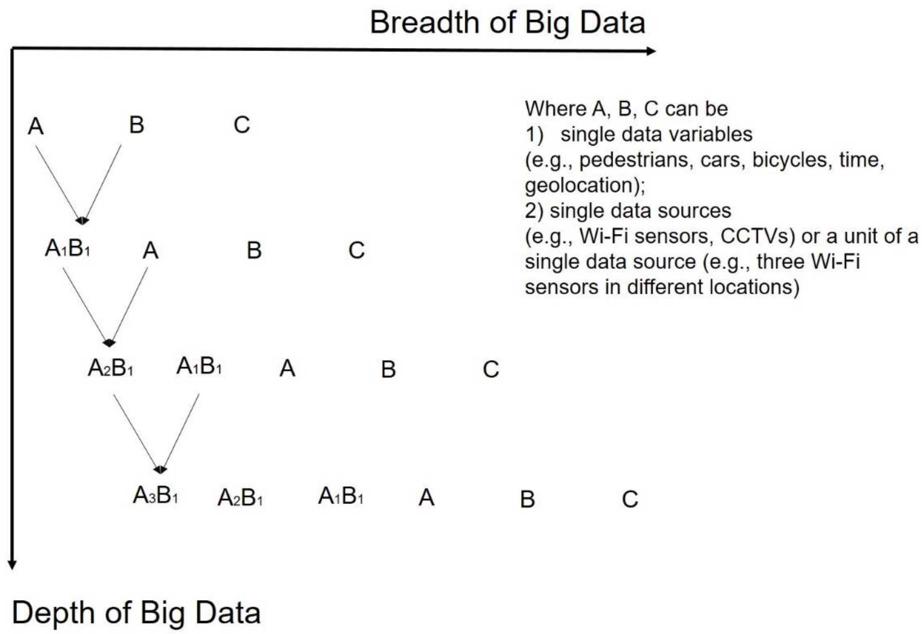

Figure 7. *Proliferation* Effect of *Volume* and *Variety*

Conversely, *Velocity* and *Veracity* have an *Additive* effect defined as the performance of all sub-dimensions acting together is equal to the sum of each sub-dimension taken separately. Adding or disregarding one, several, or all facets of *Veracity* or *Velocity* affects in a positive or negative way, respectively, the overall performance of *Veracity* and *Velocity*. Exemplarily, by respecting all sub-dimensions of the *Veracity*, it is possible to reach a high level of Big Data reliability, while disregarding at least one of the facets leads to the lowering of Big Data trustworthiness. Similarly, having all four facets of *Velocity* (data generation, fusion, visualization, and use speed) at a high-speed rate will increase the overall speed of *Velocity* dimension. The logics of *Additive* effect are represented in Figure 8.

Figure 8. *Additive* effect on the Example of *Velocity* Dimension

3.5.3. Big Data Insights

Using the findings and conceptualization of the two effects of Big Data dimensions presented above, the author runs the second round of the participatory observations. Thus, although all Big Data dimensions were the point of critical attention of the ATI pilot technical specialists, the author found that dimensions were not the same during the pilot, they were evolving, and this fact triggered different insights creation.

Specifically, the author observed four types of insights produced during the pilot. The first group of insights was able to answer simple numerical or categorical questions, such as “What are the types of vehicles using a road?”, “How many bicycles are on a road?”, “What are environmental conditions close to a road area?”. However, it is not only about the depth of data fusion. The author observed that insights were not highly reliable as, exemplarily, not all sensors were set up correctly yet, the target of the data gathering and analysis was defined very generally, the dashboard was on a testing phase, and data were not visualized automatically. These insights were used to prove the need to perform changes on a very general level and to further set up the pilot project. For this reason, the technology providers and cities involved in the pilot tended to not spend much time on assuring the Big Data *Veracity* and *Velocity*. The second group consists of reliable insights although not deeply fused. Exemplarily, insights could prove that “Types of vehicles used on roads are diverse”, “There are fewer bicycles on roads than motorized vehicles”, and “Weekday's traffic trends are clearly defined and visible to build new traffic strategies”. The distinguishing feature here was that insights were fused and visualized fast and automatically using reliable algorithms and technology setups were respected. In ATI Pilot, these insights proved that types of vehicles used on roads are diverse, there are not a lot of bicycles on roads in comparison with motorized vehicles, and weekdays traffic trends are clearly defined and visible to build new traffic strategies. The third group of insights includes the ones that were very sophisticated in terms of data sources and data fusion. It was possible to answer comparative questions, such as, “What is the level of air pollution with all road users or with only cyclists and pedestrians?”, “What is the cyclists’ peak hour over weekends?”, “What is the main destination point of pedestrians on Wednesdays?”. However, these insights lacked reliability and speed features, especially data visualization was slow. As the dashboard had certain visualization functionality restrictions, cities couldn't get cumulative and automatically visualized answers on the busiest days in the cities (weekdays trends) or sum of median cyclist/pedestrians counts across all count lines, and other insights but only via manual data fusion and visualization. Finally, the fourth group consists of granular insights based on reliable real-time data gathered, fused, and visualized at a high speed. It becomes possible to answer questions, such as “Can data confirm that removing all motorized vehicles from roads will have a direct positive impact on air pollution?”, “What will be the picture of air pollution like if the author removes vans from a data set?” etc.

To minimize case dependencies and to evaluate correctly the role of different degrees of *Proliferation* and *Additive* effects on Big Data insights creation, the author runs Big Data analytics. For this, the author used raw Big Data generated during the ATI Pilot. The goal was to get insights on road users’ traveling patterns in Manchester and predict road users’ behavior based on the historical counts, as it was not included as options in the original ATI pilot’s dashboard but might be particularly interesting for the future planning of cities’ infrastructure. The author did not have access to all sets of raw Big Data, specifically not to all Wi-Fi sensors, however, the author proceeded with what the author was given to simulate a real-life situation.

Firstly, the author run descriptive analytics that aimed at answering the simple question, "*What is the total number of vehicles per each street in Manchester per week?*" using data from one data source – CCTV cameras (see Figure 9). The author had access to one week of data, from 12th August 2019 to 18th August 2019, and analyzed only them.

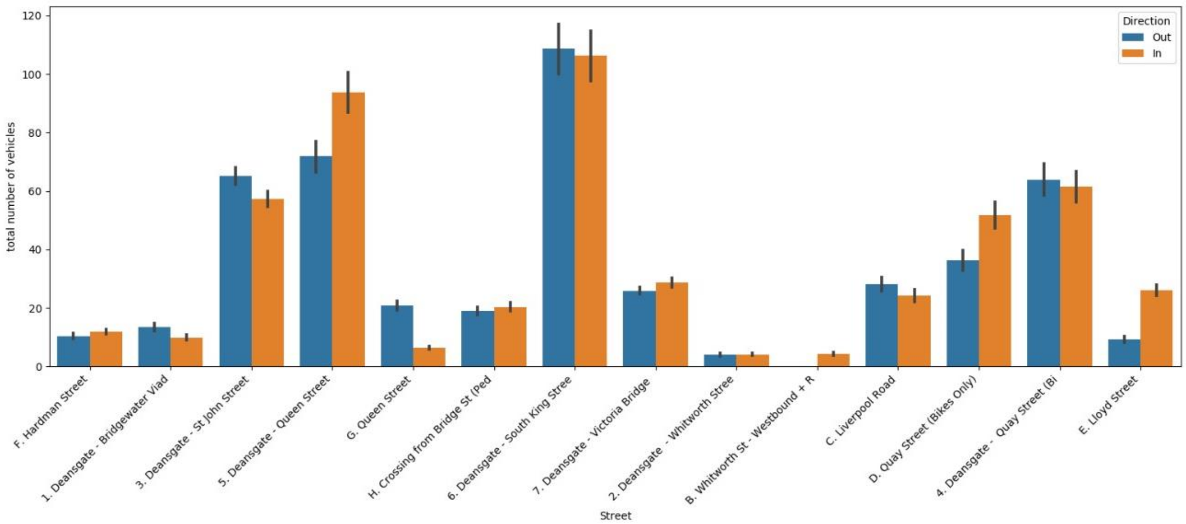

Figure 9. The Total Amount of Vehicles Using on Manchester per Each Street, In/Out

Thus, the author found that both *breadth* in terms of data sources (only CCTVs data source and single cameras across all streets of Manchester) and data variables (streets, in/out direction, date, and not defined road users) and *depth* of data were at a low level (not fused deeply). Indeed, the *breadth* of Big Data was enough to show general traffic patterns in the city in comparison per road. Data was not deeply fused, it was a general level of simple calculations. That means they have a low degree of *Proliferation* effect. Furthermore, the author did not profile data to enable error-free data manipulation, as it was rather difficult to do manually. This resulted in the issue that the results presented all road users, including pedestrians, therefore, the sample was not representative enough to conclude on the vehicles' number. One CCTV camera stopped gathering data (14th August 2019 at Deansgate/Victoria Bridge location) due to the electric power blackout; no alternative energy source was provided immediately. Moreover, data was analyzed and visualized semi-automatically that took us relatively a lot of time to put them in a file, apply appropriate formula and visualize only one week of data from one city (using Excel package). Therefore, *Veracity* and *Velocity* were at a low degree of *Additive* effect too. The author found that insights produced at a low degree of both *Additive* and *Proliferation* effects are simple and not sophisticated, although they enabled the first look at traffic and vehicle movement without deep data analysis.

Secondly, the author runs an exploratory analysis of the Big Data set to find the answer to the question to test the hypothesis, such as "*Is the number of motorized vehicles higher than the number of non-motorized ones in Manchester per week?*". The author analyzed the counts of motorized vs. non-motorized vehicles during the week from 12th to 18th August 2019 in Manchester; the author analyzed data from two sources – Wi-Fi sensors and CCTV cameras (see Figure 10).

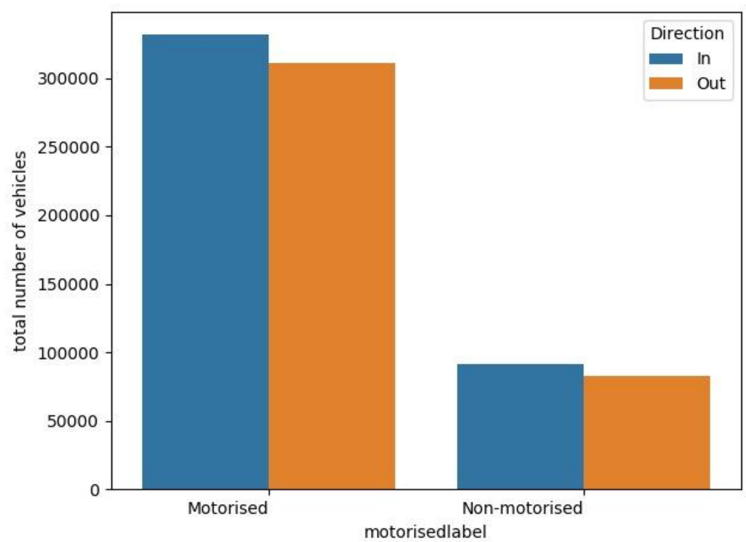

Figure 10. The Counts of Motorized vs. Non-Motorized Vehicles in a Week in Manchester

Although, in the counts of motorized vs. non-motorized vehicles, the author analyzed two data sources – CCTVs (for both motorized and pedestrians) and Wi-Fi sensors (for pedestrians only), data at this level of analysis does not have a high degree of *Proliferation* effect. Specifically, the author was working with the high level (abstract level) of data with larger *breadth* (the author analyzed all streets CCTVs' data from, the author also included the second data source – Wi-Fi sensors) and still relatively low *depth* (mostly single variables). However, the data has a high degree of *Additive* effect because of *Velocity* and *Veracity* characteristics of the data. To generate an answer for such questions, the author used Python software packages, so the author was able to deal with the big size of data easily. Therefore, the author used the Python Pandas library for data preprocessing and Matplotlib for data visualization not only to minimize errors in data manipulation but also to increase the *Velocity* dimension, specifically, *data fusion* and *data visualization* speed. Finally, *Veracity* dimension was respected at the levels of its all sub-dimensions, all technical specifications were respected, no hardware was performing poorly, high precision in representative sampling was reached by employing two complementary Big Data sources – CCTVs and Wi-Fi sensors. The author found that in the case of a low degree of *Proliferation* and a high degree of *Additive* effects, insights generated at two diametrically-opposed degrees of the effects are high-level and general, but fast to get and reliable.

Thirdly, the author runs a time series analysis to answer the explorative question, "*What is the cyclist patterns across a day in Manchester?*". For this type of question, the author aggregated time slots (every 15 minutes) data from all count lines in Manchester and analyzed cyclists' trends across a day in two directions – in and out. The author analyzed data from 31st July 2019 to 6th August 2019 week. there are two clear spikes in the data, one around 8:00 to 9:00 am and then the higher spikes around 4:00 to 5:00 p.m. (see Figure 11). Moreover, the very precise cyclists' movement trend line across the city was discovered.

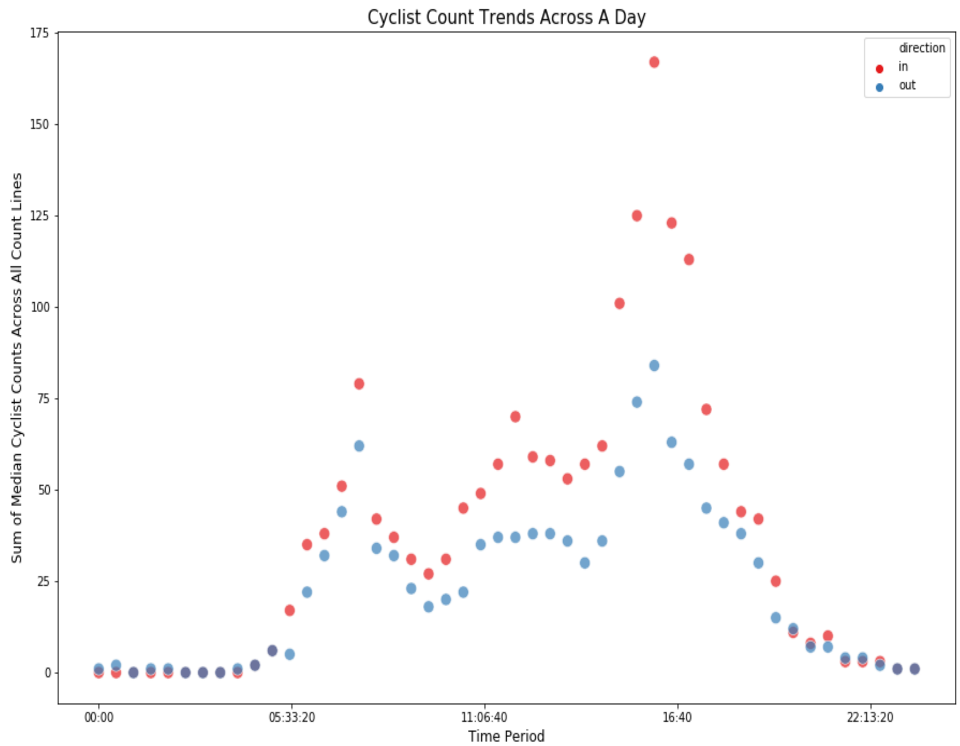

Figure 11. Cyclists Count Trends Across a Day in Manchester

This level of analysis has a high degree of *Proliferation* effect. It is not anymore just a high abstract level of data but very precise and narrow counts and trends search, as the author analyzed not only single variables (just dates as previously) but deeper levels of fused data (specific time slots during specific dates – every 15 minutes together) to find precise and very granular traveling pattern. Although *Variety* is presented in terms of single CCTV cameras located at all counts lines in Manchester and not in terms of *Variety* of the data sources, it became possible to go more in-depth and fuse data. The time-series analysis was run using Python programming language and Matplotlib for the bar chart creation as more advanced data visualization is appreciated in this case. Despite this, *Velocity* of data visualization was not very high, as it required an ad-hoc timely approach. *Veracity* was affected by the fact that data from one CCTV camera was missed (Deansgate/Queen street; 2nd August 2019; around 15:00 – 19:45) due to technical issues. Therefore, not it is not possible to claim that data are representative (especially it is harmful to short time-period data such as one week), nor that the overall *Additive* effect is high. The author found that data with a high degree of *Proliferation* effect and a low degree of *Additive* effect provide sophisticated insights with high *depth* and *breadth* but with small reliability, insights that are slowly to analyze and apply.

Finally, the author runs predictive analytics, specifically, decision tree, to predict the fellow of traffic for Manchester and the preferable street for motorized and non-motorized vehicles. The author analyzed CCTVs data from 1st August 2019 to 9th August 2019 and Wi-Fi sensors from 22nd June 2019 to 30th September 2019. The street names were used as target variables while features include: country, timeslot (morning/afternoon); direction (in/out), weekend (true/false), motorized vehicles (cars, trucks, motorbikes, vans, taxis, and emergency cars), and non-motorized (cyclists and pedestrians)**.** The decision tree shows patterns for the UK, highlighting the most probable directions for motorized road users in the UK (see Figure 12).

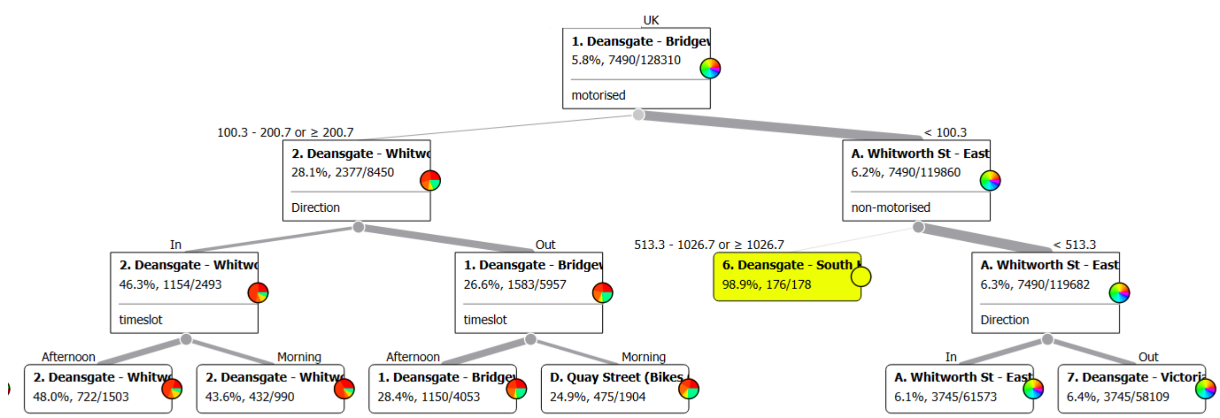

Figure 12. The Travel Direction for the UK

For instance, if you are in 1.Deansgate-Bridgwater Viad, then motorized vehicles could lead you to the next street. If the average number of motorized vehicles is less than 100 (<100.3), then the next street is A.Whitworth St-East bound; otherwise, it leads to 2.DeansgateWhitworth Street. While from A.Whitworth St-East bound, the author can see that majority of non-motorized (non-motorized>=513) will head toward 6. Deansgate-South bound. Otherwise, if the average non-motorized is less than 513.3, then the next stop is the same street: A.Whitworth St-East bound. From this point, it depends on the direction of traffic. For the "out" direction, the destination street is 7. Deansgate-Victoria Bridge. This level of analysis has both effects at a high degree. Insights produced at this level are highly sophisticated and granular due to the high degree of *Proliferation* effect. Moreover, *Velocity* and *Veracity* features are respected. *Velocity* characteristic is at the highest degree, as the author run sophisticated predictive analytics using ready data models and visualizing the decision tree automatically, and no ad-hoc approach was used. Most importantly, *Veracity* is respected too, as the author took an existing reliable dataset from the Pilot, no technical disruptions occurred during this one month. Moreover, to enable the highest level of data reliability, the author analyzed data on non-motorized vehicles from both CCTVs and Wi-Fi sensors. Specifically, Wi-Fi sensors are responsible for accurate counts of pedestrians in ATI case, while considering data on pedestrians only from CCTV might put down representativeness of the data set. Finally, the target was very specific – the author wanted to predict patterns in Manchester for the same month (August) for the next year, as one month of historical data analysis is too small to predict for the whole year ahead or another month. The author found that a high degree of *Proliferation* and *Additive* effects offer highly reliable synthetic insights that are generated, fused, visualized at a fast pace.

3.6. DISCUSSION

Although the existing literature reasonably states that Big Data insights should be timely, useful, actionable, etc., (LaValle et al., 2011; Chen et al., 2015; Sivarajah et al., 2017; Davenport and Harris, 2017), it does not give any tools other than Big Data analytics that explain how a firm can generate those insights, thus making the statements rather superficial and Big Data usefulness rather vague for managers. To the author's knowledge, the research is the first that sheds light on the mechanism of producing different types of insights based on the role of Big Data dimensions and their distinct effects. The author conceptualized the findings by drawing an empirically grounded matrix that builds on the notion of *Proliferation* and *Additive* effects working simultaneously either at a low or high degree. Thus, Figure 13 presents the matrix depicting four types of Big Data insights created according to the degree of *Proliferation* and *Additive* effects.

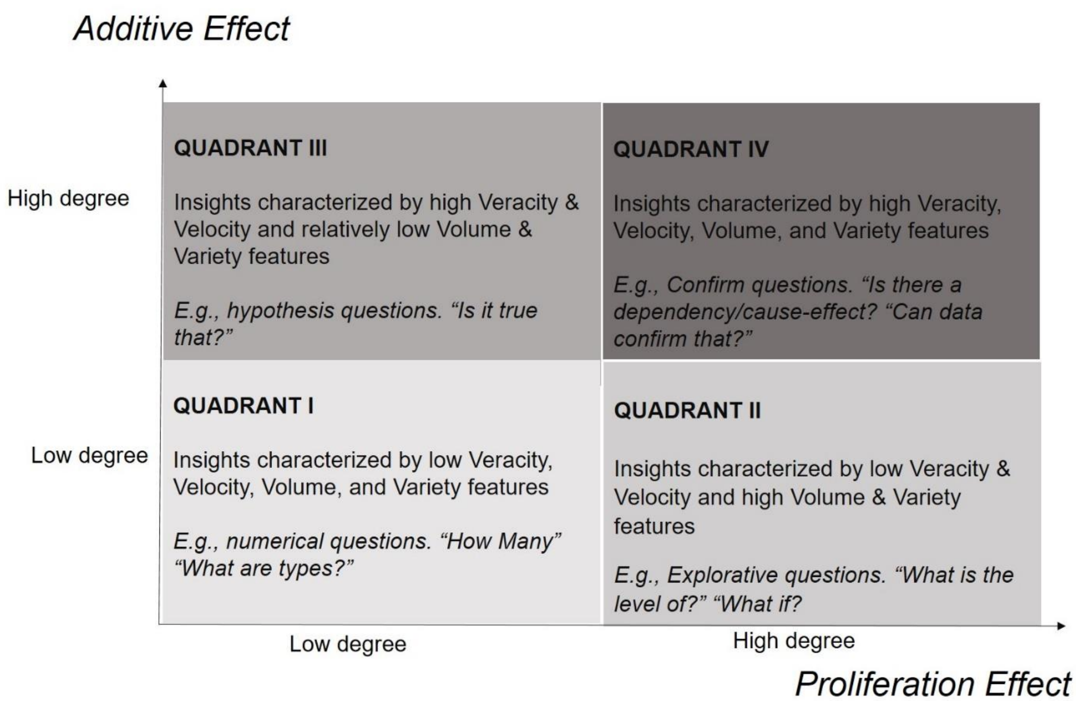

Figure 13. Big Data Effects and Insights Matrix

Specifically, Quadrant I depicts how the low degree of both *Proliferation* and *Additive* effects enables simple numerical insights that can be even single variables of Big Data or results of a basic level of data fusion with the low level of data *Veracity* and *Velocity*. This type of insights can be characterized by the absence of hidden patterns or trends search.

In Quadrant II the insights produced at a high degree of *Additive* and low degree of *Proliferation* effects have a high level of reliability, but they do not, however, provide a deep understanding of the phenomenon under investigation as they lack *Proliferation* effect, more precisely, the *depth* of Big Data. For example, hypothesis questions might be answered at this level of analysis, as the high speed of data generation, fusion, and visualization makes them useful to build reliable high-level reports at a fast-pace to prove that change is needed.

Quadrant III shows how a high degree of *Proliferation* and low degree of *Additive* effects enable insights that seek hidden logics, dependencies using computational tools. However, the level of data reliability might not be sufficient to base any strategic decision, change, initiative, without risk of failure and make a critical mistake. The speed is the critical characteristic here, as the *depth* of data will pose an issue for fast analysis, visualization, or use of data. This type of analysis could be seen as a follow up for numeric questions where the author can rely on high *Proliferation,* but the *Additive* Effect is still low.

Quadrant IV synthesizes how a high degree of both *Proliferation* and *Additive* effects enables insights that not only unveil dependencies, correlations, hidden trends but also can identify proving or rejecting facts for strategic decisions. The high degree of both *Proliferation* and *Additive* effects *enables* the most synthetic and reliable insights and evidence that can be used to build long and short-term strategies, in the case, in the sphere of transport and city planning.

The degree of *Proliferation* effect influences the level of sophistication of insights. Thus, insights produced at a low degree of *Proliferation* effect can answer simple numerical or categorical questions. *Additive* effect affects the level of usability of Big Data insights. Insights produced at a lower degree of the *Additive* effect cannot be used immediately, as they are not reliable and require a higher level of *Veracity.* Moreover, they are not generated, or fused and visualized fast enough for immediate and effective use. While a high degree of *Additive* effect enables use in fast-pace mode highly reliable data.

The author argues that insights are generated at the level of Big Data, and only then can they be translated into value at a firm level. Therefore, in this paper, the author focused on Big Data as the most essential and the prime source to create relevant insights that can be translated into a potential advantage over competitors. Neglecting this source and paying attention only to firm capabilities such as Big Data Analytics capabilities (Van Rijmenam et al., 2019) could lead to a partial view of the overall phenomenon; this view is consistent with the existing Big Data studies (Chen et al., 2012; McAfee et al., 2012; Wamba et al., 2017; Mikalef et al., 2020). A deeper understanding of the insights creation mechanism is especially relevant for firms that tend to rely more on the managerial intuition in decision-making as they do not fully understand the value and perceived benefits of Big Data due to its statistically and technologically heavy nature (Court, 2015; Yadegaridehkordi et al., 2018; Zhang et al., 2019).

In the present research, the author enlarged and specified the role that Big Data dimensions play in Big Data insights creation mechanism and explained the role through the two effects, namely, *Proliferation* and *Additive*. Thus, while it is not disputable that the merely computational view on Big Data depreciates its strategic significance (Dijcks, 2012; Schroeck et al., 2012; Yoo, 2015), the author crystallized how exactly all dimensions, namely, *Volume, Variety, Veracity,* and *Velocity*, contribute to the comprehensive quality of Big Data. Specifically, an understanding of the *Volume* and *Variety* cannot be narrowed to purely size and variety of sources or to the *breadth* of data. The author found that the *depth* of Big Data could be even more important for meaningful insights creation than the size and variety of Big Data sources *per se* (Yoo, 2015; Günther et al., 2017; Pröllochs and Feuerriegel, 2020). The case of

ATI showed that the quality and quantity of *Volume* of Big Data increases with data fusion; moreover, it is due to this *depth* produced by data fusion that more sophisticated insights were created. *Variety* of data sources does not guarantee meaningful insights, while the right combination of data sources does. It has become evident that variable data sources without their further deepening and sophistication have a lower capability of producing meaningful insights. With this respect, *Proliferation* effect that was observed on the two dimensions affects the level of sophistication or granularity of the insights created. The author also found that although granular insights are wanted, two other dimensions influence the insights creation mechanism. Indeed, while the four *Velocity* sub-dimensions are case-related and there is no unified model of time-use or visualization speed rate, having all four at a high-speed rate enables an *Additive* effect on the value creation ability of Big Data, as the Active Travel Insights analysis revealed. Each sub-dimension has an independent impact on the meaningful insights' creation. Specifically, treating *Velocity* as an asset rather than just a characteristic (Chen et al., 2012) can improve the ability of Big Data insights to be timely. While data generated at a fast speed can have more potential to bring value, the author found that fast data analytics and data visualization can unlock the value potentiality (Larson and Chang, 2016). Fast generated and analyzed data can lose value if not used timely (Gandomi and Haider, 2015; Lee, 2017). Poor data *Veracity* might affect the credibility of generated insights (Surbakti et al., 2020). Furthermore, respecting or disregarding one, several, or all sub-dimensions of *Veracity* can affect positively or negatively, respectively the ability of Big Data to produce valuable and reliable insights. Not only the data source but also time and relevance (Sukumar and Ferrell, 2013; Lee, 2017; Ghasemaghaei, 2019) or technical setups (Côrte-Real et al., 2020) influence the quality of Big Data and, consequently, Big Data insights. Therefore, even with a low level of *Proliferation* effect, insights produced with the high *Additive* effect degree could bring a certain value to organizations.

The present research does not aim at explaining the role of the two effects on producing “good” or “bad” Big Data insights but rather at providing an understanding of how to produce correct and needed insights. Thus, as the author observed during ATI Pilot, all created insights served a certain goal during the Pilot. Moreover, Big Data analytics performed by the author confirmed that even insights with a low degree of *Proliferation* (not sophisticated insights) are needed, as well as a high degree of *Additive* effect is not always the key target when insights need to be sophisticated and not with the high level of usability.

The present work provides a bridge between technology innovation and information systems literature and responds to the call to go beyond the boundaries of the IS research and offer new theoretical insights that guide management practices in the digital age (Yoo, 2012). Specifically, the author responds to the call for more research on Big Data insights with a focus not only on the issues of data analytics (Bharadwaj et al., 2013; Wessel, 2016; Vidgen et al., 2017; Pröllochs and Feuerriegel, 2020). Thus, the author initiated a first to the author’s knowledge complex research to understand the role of Big Data dimensions in the insights creation mechanism following the fragmented research (George et al., 2014; Yoo, 2015; Lee, 2017). Moreover, the author contributed to the discussion on the usefulness of Big Data (Johnson et al., 2017; Cappa et al., 2020). In line with the scholars assuming that Big Data is useful when they generate valuable information and insights (McAfee and Brynjolfsson, 2012; Marshall et al., 2015; Cappa et al., 2020), the author put the break into understanding how Big Data sub-dimensions influence the nature and quality of created insights, that was a missing point (Aaltonen and Tempini 2014; Constantiou and Kallinikos, 2015; Yoo, 2015). Finally, Big Data represents a strategic resource for firms but is an emerging paradigm within technology innovation literature, therefore, understanding the mechanisms of creation of different insights and developing credible sources of information from Big Data to include in firms' digital transformation activities is crucial (Sivarajah et al., 2017; Chen et al., 2012; McAfee et al., 2012; Braganza et al., 2017; Wamba et al., 2017).

3.7. MANAGERIAL IMPLICATIONS, LIMITATIONS, AND FUTURE RESEARCH DIRECTIONS

The research has several implications for managers. It is not possible to generalize what is useful, timely, valuable Big Data insights, as they are unique for each firm and project. However, the notion of Big Data dimensions and sub-dimensions can become an assessment tool for any Big Data project, guiding managers and assuring necessary for the quality, depth, and speed of insights creation. Moreover, a manager can use the Big Data insights matrix to map which kind of insights they aim to get, and what are the main dimensions they should target to create timely and useful insights. Secondly, firms should adopt a smart resource allocation; this does not mean using many computational tools if they need simple numerical insight or adopting very advanced visualization tools to get fast insights instead of deep but slow ones. As the research showed, organizations might already intuitively implement elements of the wise resources alignment, however, doing it in a more conscious and strategic way can lead to better value creation. Thirdly, firms can better control their expenses, as they can increase their ability to improve the quality of insights raising the degree of the effects. Specifically, they might choose to employ more hardware to generate various data (*breadth*), which is particularly costly, or to deepen insights via algorithms (*depth*) instead; invest in real-time data generation or to buy historical data from the third party, as either source of the data can fit a project, but the latest option costs less. This can lead to better returns on investments. Finally, having more knowledge on Big Data and its produced effects, firms and managers can enhance their planning activities and craft long-term Big Data strategy and data-driven organizational culture.

The main limitation of the study regards its explorative nature of the study; this can limit its potential for generalizability. To address this issue, the author applied the data triangulation principle; moreover, the author implemented a sort of validation of the author's conceptual and empirical findings by applying Big Data analytics. To cover all possible aspects of the Big Data insights creation, the author analyzed the pilot project, thus making it possible to follow all pitfalls from the very beginning. However, the pilot nature of the case did not allow us to fully explore a theoretically derived sub-dimension of *Velocity*, namely, *use speed*, which the author believes is crucial not only for Big Data insights creation but also for value generation. The author also suggests that some other sub-dimensions might arise while exploring cases at the mature stages or within different contexts. Finally, the limitation posed by the smart mobility context and the peculiarities of IoT technologies used in the Pilot can restrict the generalizability of the findings. As it seems almost not possible to exclude the bias of the context while performing the participatory observation study, the author put special attention on the case choice. Thus, although there is presented only IoT technology to generate Big Data, the shape of the data was diverse (videos, measurements, geolocations, etc.).

Therefore, future works might offer studies that address the generalizability issue of the current research by employing other methods, specifically, testing the effects notion using multiple case study approach or validating matrix using a quantitate approach on more massive data sets. Findings should be tested on the applicability in other fields, for example in e-commerce, advertising, or medicine. An important aspect is to investigate the boundaries of the two effects, specifically, if a high degree of both effects always has a positive sign or to explore what happens with the insights when the degree of the effects is not strictly polar. Moreover, as the research focused solely on the Big Data level, future studies might explore the conjoint effects of Big Data as an essential resource of competitive advantage and the firm's capabilities to transform them into useful and high-quality insights. Specifically, exploring how the firm's capabilities influence which dimensions they will focus on and how does it affect the sub-dimensions.

CONCLUSIONS

The present thesis investigated three particular phenomena of the earliest Digital Transformation (DT) stage, namely, proof-of-concept (PoC), data-driven decision-making (DDD), and Big Data insights creation. The study aimed at disentangling technological uncertainty and technology awareness at the level of PoC, at developing a better understanding of the cognitive and data biases and their influence on trust in data when DDD is first introduced, and investigating the core mechanisms of Big Data insights creation on the level of Big Data dimensions. Focusing on the three aspects allowed me to build a research agenda that consists of complementing each other parts. Furthermore, grounding the pioneering study on both the existing research body and research frontiers allowed to establish the research program that contributes to the existing literature in technology innovation and opens up new research avenues within the emerging topic of Digital Transformation. Next, the chapter will summarize the main findings, contributions to theory and methods, concluding with the limitations, and future research directions.

Summary of the Main Findings

The essay on the PoC—the initial step for firms willing to digitally transform—disentangled technology uncertainty at the level of PoC and found how cognitive biases influence the dynamics of PoC. I found that cognitive biases drove what I labeled as *perceived technology potentiality*. I delineated three degrees of technology awareness that developed as PoC is run: *borrowed technology awareness, minimum acquired technology awareness,* and *enhanced technology awareness.* Finally, I drew a PoC dynamic explaining how the two constructs, *perceived technology potentiality,* and *technology awareness* strengthen or balance each other in the PoC milieu.

Further, I explored the first use of data-driven-decision-making and the role of trust in data. I maintained the focus on the topic of cognitive biases within the context of DT's initial steps and introduced the new verge, the data biases. I found that different types of traps drove managerial trust in data. I conceptualized traps into three zones that are related solely to the DDD type that supports decision-making. The *cognition traps zone* that I observed at the initial stages of the tests captures the strong trust in data effectiveness built upon someone else's data experience and defended by cognitive biases. *Data technology traps zone* appeared after the cognition traps zone. I found that the level of trust in data effectiveness was low, as managers were misled by the data tuner persona, wrong data input, and data setups. Therefore, Big Data, which is supposed to be unbiased, could not be completely neutral as it was affected by data biases. *Traps recognition zone* followed data technology traps zone and was characterized by acknowledging the two knowledge sources results, intuitive and data. All these led to the growing trust in data again.

Finally, by focusing on the other side of the DDD – data, I shed light on the mechanism of producing different types of insights based on the role of Big Data dimensions and their distinct effects. Specifically, I disentangled Big Data dimensions into sub-dimensions and conceptualized differences and similarities into the notion of the two effects, namely, *Proliferation* and *Additive*. Ultimately, by using this conceptualization effort, I showed how four different types of insights could be generated as *Proliferation,* and *Additive* effects work simultaneously either at a low or high degree.

Contributions to Theory

The first essay bridges two streams of literature to initiate the discussion on the phenomenon of PoC that was surprisingly by-passed in the technology innovation literature.

Firstly, the study contributes to the understanding of PoC by defining its main constructs. Further, it contributes to the managerial cognition literature by focusing specifically on beliefs and attitudes in the technology innovation, foresight, and adoption context and combining it with fragmented research on technological uncertainties (Kahneman and Tversky, 1977; Banerjee, 1992; Busenitz and Barney, 1997; Ragatz et al., 2002; Zhang and Cueto, 2017). Moreover, the introduction of the notion of technology awareness degrees in the PoC phase, the present work enlarges the existing research that has covered mainly the technology product innovation context so far (Fleck, 1994; Bohn et al., 2005; Roca et al., 2017).

The second essay connects core streams in innovation studies, namely, managerial cognition (Busenitz & Barney, 1997; Haselton et al., 2005; Kahneman & Tversky, 1977; Zhang & Cueto, 2017;) and the emerging stream on data-driven decision-making (Meyer et al., 2014; Baesens et al., 2016; McAfee & Brynjolfsson, 2012). Firstly, the work enlarged the understanding of how the traditional theory of decision-making, namely, Parallel-competitive theory (Hodgkinson and Sadler-Smith, 2018), works on the new phenomenon of DDD. Thus, surprisingly but in line with the traditional view on strategic decision-making, both cognition and analytics in DDD independently participate in decision-making. I also observed the intriguing synergy between cognition and data that occurred only after each type of decision-making contributed separately. Secondly, the research contributed to the managerial cognition literature by focusing on the cognitive biases that arise even with DDD promises to address imperfections and biases of human decision-making (Lindebaum et al., 2020). I also argue DDD as the research stream should go beyond purely data technologies and Big Data analytics scope (Bertsimas & Kallus, 2020; Camiña et al., 2020; Mandelbaum et al., 2020) and include cognition as the integral factor of enhanced decision-making. Thirdly, the work contributed to the emerging discussion on data biases that occurred due to the humans' behavior (Lindebaum et al., 2020; Choudhury et al., 2020) and introduced data traps on the analogy with cognition biases. Finally, the research contributed to the discussion on the trust in data for effective decision-making (Surbakti et al., 2020; Glikson and & Woolley, 2020) by focusing on the specific situation of data as a supportive humans' decisions tool; trust in data is influenced by cognitive and data traps.

The third essay provides a bridge between technology innovation and information systems (IS) literature and responds to the call to go beyond the IS research boundaries and offer new theoretical insights that guide management practices in the digital age (Yoo, 2012). Specifically, the research responds to the call for more research on Big Data insights, focusing not only on data analytics issues (Bharadwaj et al., 2013; Wessel, 2016; Vidgen et al., 2017; Pröllochs and Feuerriegel, 2020). Thus, to my best of knowledge, I initiate the first research to understand the role of Big Data dimensions in the insights creation mechanism following the fragmented research (George et al., 2014; Yoo, 2015; Lee, 2017). I put the break into understanding how Big Data sub-dimensions influence the nature and quality of created insights; that was a missing point (Aaltonen and Tempini 2014; Constantiou and Kallinikos, 2015; Yoo, 2015).

Contributions to Method

The second area of the contributions is to the research methods for exploring the early stages of DT. Due to the novelty of the topic and its position on the research frontiers of the technology innovation field, the first reasonable impulse is to apply qualitative methods that aim at providing fine-grained findings on the phenomenon. However, when it comes to generating knowledge that aims at real-world applicability, the number of methods reduces.

Further, the technological nature of the research topic urges to use of quantitative methods to enable a greater level of reliability and generalizability of findings. However, running mix-methods studies is not always an option. With this respect, two methodological contributions arose.

Firstly, among the variety of the qualitative methods, the participatory observation is particularly interesting as it allows deep dive into the topic, establishing non-hierarchical relationships with the participants and non-passive observations in the heart of the novel topics (Clark et al., 2009; Bryman, 2012). Moreover, the method is of special interest for researchers with a practical background in the topic and enables generating "knowledge for action" (Clark et al., 2009). However, I faced the desert of methodological guidelines and papers describing the method used within technology innovation literature and active method's application within information systems (IS) literature (Baskerville & Wood-Harper, 1998; Baskerville, 1999). The solution was by drawing on the three stages participatory research process developed by IS scholars Street and Meister (2004) to build an ad-hoc comprehensive framework of reliable data generation via participatory observation and transformation of data in knowledge within the technology innovation context. Therefore, the present research was the first that brought the method to the technology innovation field, to the best of my knowledge. Together with the traditional methods provided by technology innovation studies, it allowed a better understanding of the phenomenon under investigation.

The second contribution is related to the use of novel computational content analysis techniques to validate qualitative findings and enable more nuanced results rather than come up with new findings. The use of quantitative methods in this mode opens up new exciting possibilities to address the generalizability issue of qualitative studies, and the issue of the methods lack to explore emerging topics, such as Digital Transformation (Hanelt et al., 2020). Specifically, Latent Dirichlet analysis applied to a text corpus uses an unsupervised Bayesian machine learning algorithm to discover context-specific hidden topics (Blei et al., 2002). The method has started gaining attention in strategic management related to sensitive cognitive topics to discover latent themes in the collection of documents (i.e., Kaplan and Vakili, 2015). Another very new text mining technique, namely, Sentiment Analysis (Pandey et al., 2017; Moe & Schweidel, 2017), identifies the polarity of the sentiments, namely, positive, negative, or neutral. It uses Natural Language Processing and categorizes text based on the sentiments (Pandey et al., 2017; Moe & Schweidel, 2017). The technology innovation field uses a combination of quantitative and qualitative methods in the mix-methods studies (Venkatesh et al., 2013); however, the direction I applied in the present study - the use of the very new quantitative content techniques to provide more nuanced qualitatively derived findings - is a quite high potentiality combination to inform existing research methods.

Limitations

Despite the mentioned above contributions, several limitations as well as the ways I addressed them in work, are worthy of highlighting. Firstly, while the novelty of the topics suggested the qualitative nature of the research, as it allows relying upon particularly fine-grained data and bringing an understanding of the phenomenon under investigation to a deeper level, qualitative research generates issues of restricted generalizability. To address the issues, every essay uses the triangulation principle of data collection and analysis, thus, enabling profound evidence-based and relying upon different data sources. Moreover, each essay applies an ad-hoc quantitative approach to validate qualitatively derived findings. Thus, I applied the two novel content analysis techniques (Sentiment Analysis and Latent Dirichlet Analysis) in the first two essays and Big Data Analytics in the third essay. The second limitation touches upon the first chapter and stems from the restricted ability to investigate all possible behavioral scenarios. I chose a specific "positivistic" scenario to address the issue as it is discussed as the most probable situation in the technology innovation literature. In this scenario, managers already had positive expectations towards technology to test (Bohnsack et al., 2018). A similar limitation comes to the second essay, where I chose only one scenario data playing the "supportive" role in decision-making; however, the "substitutional" function of data and AI is missing in the research.

Future Research Directions

As the current work is pioneering and offers novel contributions to the emerging topics of the first steps within Digital Transformation, it is especially promising to continue the research efforts in the indicated directions. By picking up the baton of this study, future works might offer studies that focus on other technological contexts. It would be of great interest and importance to explore if other dynamics and effects are working in different technological contexts. With this respect, another sub-direction could test the findings of this research on the applicability in other technological fields and industries to derive fine-grained aspects within the ocean of emerging technologies. Thus, it would be fascinating to understand if some specific technology-dependent dynamics could not be observed in this research. Further, researchers might step into the promising field of algorithms that replace humans in decision-making and investigating what influences the trust in data in the case when human cognition theoretically is completely excluded. Additionally, as the second essay is grounded on the parallel-competitive theory assuming that data and cognition both equally participate in decision-making, it would be of great interest to look at the DDD through the lens of other decision-making theories, where, for example, only one, either intuition or logical type (data), plays a more important role in decision-making. Finally, although in the era of DT technologies are mostly positively met by managers, it could be relevant to explore the scenario when managers are initially skeptical towards technologies and the need to transform digitally.

REFERENCES

Aaltonen, A., & Tempini, N. (2014). Everything counts in large amounts: a critical realist case study on data-based production. *Journal of Information Technol*ogy, 29 (1), 97–110.

Airehrour, D., Gutierrez, J. A., & Ray, S. K. (2019). SecTrust-RPL: A secure trust-aware RPL routing protocol for Internet of Things. *Future Generation Computer Systems*, *93*, 860-876.

Akter, S., & Wamba, S. F. (2016). Big data analytics in E-commerce: a systematic review and agenda for future research. *Electronic Markets*, *26*(2), 173-194.

Akter, S., Bandara, R., Hani, U., Wamba, S. F., Foropon, C., & Papadopoulos, T. (2019). Analytics-based decision-making for service systems: A qualitative study and agenda for future research. *International Journal of Information Management*, *48*, 85-95.

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, *6*(1), 25.

Andriole, S. J. (2017). Five myths about digital transformation. *MIT Sloan management review*, *58*(3).

Appio, F. P., Lima, M., & Paroutis, S. (2019). Understanding Smart Cities: Innovation ecosystems, technological advancements, and societal challenges. *Technological Forecasting and Social Change*, *142*, 1-14.

Athilingam, P., Jenkins, B. A., Zumpano, H., & Labrador, M. A. (2018). Mobile technology to improve heart failure outcomes: A proof of concept paper. *Applied Nursing Research*, *39*, 26-33.

Bacon-Gerasymenko, V., & Eggers, J. P. (2019). The dynamics of advice giving by venture capital firms: Antecedents of managerial cognitive effort. *Journal of Management*, *45*(4), 1660-1688.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Transformational issues of big data and analytics in networked business. *MIS quarterly*, *40*(4).

Banerjee, 1992; https://www.behavioraleconomics.com/resources/mini-encyclopedia-of-be/herd-behavior/;

Barreto, I., & Baden-Fuller, C. (2006). To conform or to perform? Mimetic behaviour, legitimacy-based groups and performance consequences. *Journal of management studies*, *43*(7), 1559-1581

Baskerville, R. "Investigating Information Systems with Action Research," Communications of the Association for Information Systems (2:19), 1999.

Baskerville, R., and Wood-Harper, A. T. "Diversity in Information Systems Action Research Methods," European Journal of Information Systems (7), 1998, pp. 90-107.

Baum JA, Li SX, Usher JM. Making the next move: How experiential and vicarious learning shape the locations of chains' acquisitions. Administrative Science Quarterly. 2000 Dec;45(4):766-801.

Berger, S., Feldhaus, C., & Ockenfels, A. (2018). A shared identity promotes herding in an information cascade game. *Journal of the Economic Science Association*, *4*(1), 63-72.

Bernardo, A. E., & Welch, I. (2001). On the evolution of overconfidence and entrepreneurs. *Journal of Economics & Management Strategy*, *10*(3), 301-330.

Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, *66*(3), 1025-1044.

Bharadwaj, A., El Sawy, O. A., Pavlou, P. A., & Venkatraman, N. (2013). Digital business strategy: toward a next generation of insights. *MIS quarterly*, 37(2), 471-482.

Bhattacherjee, A., & Premkumar, G. (2004). Understanding changes in belief and attitude toward information technology usage: A theoretical model and longitudinal test. *MIS quarterly*, 229-254.

Bhattacherjee, A., & Premkumar, G. (2004). Understanding changes in belief and attitude toward information technology usage: A theoretical model and longitudinal test. *MIS quarterly*, 229-254.

Blank, S. (2013). Why the lean startup changes everything. *Harvard Business Review*, *91*(5), 63-72.

Bohn, R. E. (2005). From art to science in manufacturing: The evolution of technological knowledge. *Foundations and Trends in Technology, Information, and Operations Management*, *1*(2), 1-82.

Bohnsack, R., Hanelt, A., Marz, D., & Marante, C. (2018, July). Same, same, but different!? A systematic review of the literature on digital transformation. In *Academy of Management Proceedings* (Vol. 2018, No. 1, p. 16262). Briarcliff Manor, NY 10510: Academy of Management.

Bonaccorsi, A., Apreda, R., & Fantoni, G. (2020). Expert biases in technology foresight. Why they are a problem and how to mitigate them. *Technological Forecasting and Social Change*, *151*, 119855.

Braganza, A., Brooks, L., Nepelski, D., Ali, M., & Moro, R. (2017). Resource management in big data initiatives: Processes and dynamic capabilities. *Journal of Business Research*, *70*, 328-337.

Bryman, A. (2012): Social Research Methods, 4th edition, Oxford university press, Oxford.

Busenitz, L. W. (1999). Entrepreneurial risk and strategic decision making: It'sa matter of perspective. *The Journal of Applied Behavioral Science*, *35*(3), 325-340.

Busenitz, L. W., & Barney, J. B. (1997). Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making. *Journal of business venturing*, *12*(1), 9-30.

Camiña, E., Díaz-Chao, Á., & Torrent-Sellens, J. (2020). Automation technologies: Long-term effects for Spanish industrial firms. *Technological Forecasting and Social Change*, *151*, 119828.

Camuffo, A., Cordova, A., Gambardella, A., & Spina, C. (2020). A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. *Management Science*, *66*(2), 564-586.

Caplin, A., & Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, *105*(7), 2183-2203.

Cappa F, Oriani R, Peruffo E, McCarthy I. (2020). Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety and Veracity on Firm Performance. *Journal of Product Innovation Management*. https://doi.org/10.1111/jpim.12545

Carr, E. G., Ladd, M. V., & Schulte, C. F. (2008). Validation of the contextual assessment inventory for problem behavior. Journal of Positive Behavior Interventions, 10(2), 91-104.

Cennamo, C., Dagnino, G. B., Di Minin, A., & Lanzolla, G. (2020). Managing Digital Transformation: Scope of Transformation and Modalities of Value Co-Generation and Delivery. *California Management Review*, *62*(4), 5-16.

Chemla, G., & Tinn, K. (2020). Learning through crowdfunding. *Management Science*, *66*(5), 1783-1801.

Chen, D. Q., Preston, D. S., & Swink, M. (2015). How the use of big data analytics affects value creation in supply chain management. *Journal of Management Information Systems*, *32*(4), 4-39.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 1165-1188.

Chen, M. J., & MacMillan, I. C. (1992). Nonresponse and delayed response to competitive moves: The roles of competitor dependence and action irreversibility. *Academy of Management Journal*, *35*(3), 539-570.

Chen, M. J., & MacMillan, I. C. (1992). Nonresponse and delayed response to competitive moves: The roles of competitor dependence and action irreversibility. *Academy of Management Journal*, *35*(3), 539-570.

Choudhury, P., Starr, E., & Agarwal, R. (2020). Machine learning and human capital complementarities: Experimental evidence on bias mitigation. *Strategic Management Journal*.

Clark, A., Holland, C., Katz, J., & Peace, S. (2009). Learning to see: lessons from a participatory observation research project in public spaces. *International journal of social research methodology*, *12*(4), 345-360.

Conboy, K., Dennehy, D., & O'ConnorO'Connor, M. (2020). 'Big time': An examination of temporal complexity and business value in analytics. *Information & Management*, *57*(1), 103077.

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: big data and the changing context of strategy. *Journal of Information Technology*, *30*(1), 44-57.

Contigiani, A., and Levinthal, D. A. (2019). Situating the construct of lean startup: adjacent conversations and possible future directions. *Industrial and Corporate Change*, *28*(3), 551-564.

Côrte-Real, N., Ruivo, P., & Oliveira, T. (2020). Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value?. *Information & Management*, *57*(1), 103141.

Court, D., 2015. Getting big impact from big data. McKinsey Q. https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/getting-big-impact-from-bigdata (January).

Damanpour, F., & Wischnevsky, J. D. (2006). Research on innovation in organizations: Distinguishing innovation-generating from innovation-adopting organizations. *Journal of engineering and technology management*, *23*(4), 269-291.

Davenport, T., & Harris, J. (2017). *Competing on analytics: Updated, with a new introduction: The new science of winning*. Harvard Business Press.

Deshpande, R., & Kohli, A. K. (1989). Knowledge disavowal: structural determinants of information-processing breakdown in organizations. *Knowledge*, *11*(2), 155-169.

Devenow, A., & Welch, I. (1996). Rational herding in financial economics. *European Economic Review*, *40*(3-5), 603-615.

Dijcks J. P. Oracle: Big data for the enterprise. Oracle White Paper, 2012.

Ding, A. W., & Li, S. (2019). Herding in the consumption and purchase of digital goods and moderators of the herding bias. *Journal of the Academy of Marketing Science*, *47*(3), 460-478.)

Dremel, C., Herterich, M. M., Wulf, J., & Vom Brocke, J. (2020). Actualizing big data analytics affordances: A revelatory case study. *Information & Management*, *57*(1), 103121.

Duan, W., Gu, B., & Whinston, A. B. (2009). Informational cascades and software adoption on the internet: an empirical investigation. *MIS quarterly*, 23-48.)

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63-71.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory Building from Cases: Opportunities and Challenges. Academy of Management Journal, 50(1), 25-32.

Eisenmann, T. R., Ries, E., and Dillard, S. (2012). Hypothesis-driven entrepreneurship: The lean startup. *Harvard Business School Entrepreneurial Management Case*, (812-095).

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*(2), 897-904.

Fleck, J. (1994). Learning by trying: the implementation of configurational technology. *Research policy*, *23*(6), 637-652.

Fossas-Olalla, M., Minguela-Rata, B., López-Sánchez, J. I., & Fernández-Menéndez, J. (2015). Product innovation: When should suppliers begin to collaborate?. *Journal of Business Research*, *68*(7), 1404-1406.

Fosso Wamba, S., & Boeck, H., 2008. Enhancing information flow in a retail supply chain using RFID and the EPC network: A proof-of-concept approach. Faculty of Engineering and Information. Sciences - Papers: Part A. 2520.

Gaba, V., & Terlaak, A. (2013). Decomposing uncertainty and its effects on imitation in firm exit decisions. *Organization Science*, *24*(6), 1847-1869.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, *35*(2), 137-144.

Gani, A., Siddiqa, A., Shamshirband, S., & Hanum, F. (2016). A survey on indexing techniques for big data: taxonomy and performance evaluation. *Knowledge and information systems*, *46*(2), 241-284.

George G., Haas M.R., Pentland A. (2014). Big Data and management. *Academy of Management Journal, 57 (2), 321-326*

Geroski, 2000). Models of technology diffusion. *Research policy*, *29*(4-5), 603-625.).

Ghasemaghaei, M. (2019). Are firms ready to use big data analytics to create value? The role of structural and psychological readiness. *Enterprise Information Systems*, *13*(5), 650-674.

Ghasemaghaei, M., & Calic, G. (2019). Does big data enhance firm innovation competency? The mediating role of data-driven insights. *Journal of Business Research*, *104*, 69-84.

Ghasemaghaei, M., Ebrahimi, S., & Hassanein, K. (2018). Data analytics competency for improving firm decision making performance. *The Journal of Strategic Information Systems*, *27*(1), 101-113.

Gino, F., & Ariely, D. (2012). The dark side of creativity: original thinkers can be more dishonest. *Journal of personality and social psychology*, *102*(3), 445.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational research methods*, *16*(1), 15-31.

Glikson, E., & Woolley, A. W. (2020). Human trust in Artificial Intelligence: Review of empirical research. *Academy of Management Annals*, (ja).

Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information avoidance. *Journal of Economic Literature*, *55*(1), 96-135.

Gretzel, U., Sigala, M., Xiang, Z., & Koo, C. (2015). Smart tourism: foundations and developments. *Electronic Markets*, *25*(3), 179-188.

Gudmundsson, S. V., & Lechner, C. (2013). Cognitive biases, organization, and entrepreneurial firm survival. *European Management Journal*, *31*(3), 278-294.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191-209.

Gupta, S., Kar, A. K., Baabdullah, A., & Al-Khowaiter, W. A. (2018). Big data with cognitive computing: A review for the future. *International Journal of Information Management*, *42*, 78-89.

Hanelt, A., Bohnsack, R., Marz, D., & Antunes, C. A systematic review of the literature on digital transformation: insights and implications for strategy and organizational change. *Journal of Management Studies*.

Haselton, M., Nettle, D., & Andrews, P. (2005). The handbook of evolutionary psychology. In D. Buss (Ed.), The evolution of cognitive bias (pp. 724–746). Hoboken, NJ: Wiley

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., ... & Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management*, *36*(5), 748-758.

Helfat, C. E., & Raubitschek, R. S. (2018). Dynamic and integrative capabilities for profiting from innovation in digital platform-based ecosystems. *Research Policy*, *47*(8), 1391-1399.

Herterich, M., Brenner, W., & Uebernickel, F. 2016. Stepwise evolution of capabilities for harnessing digital data streams in data-driven industrial services. MIS Quarterly Executive, 15(4): 297-318

Hess, T., Matt, C., Benlian, A., & Wiesböck, F. (2016). Options for Formulating a Digital Transformation Strategy. *MIS Quarterly Executive*, *2*(15).

Hodgkinson, G. P., & Healey, M. P. (2011). Psychological foundations of dynamic capabilities: Reflexion and reflection in strategic management. *Strategic management journal*, *32*(13), 1500-1516.

Hodgkinson, G. P., & Sadler-Smith, E. (2018). The dynamics of intuition and analysis in managerial and organizational decision making. *Academy of Management Perspectives*, *32*(4), 473-492.

Horton, G. I., & Radcliffe, D. F.,1995. Nature of rapid proof-of-concept prototyping. *Journal of Engeering Design*, *6*(1), 3-16.

Jakkhupan, W., Arch-int, S., & Li, Y. (2011). Business process analysis and simulation for the RFID and EPCglobal Network enabled supply chain: A proof-of-concept approach. *Journal of Network and Computer Applications*, *34*(3), 949-957.

Johnson, J. S., S. B. Friend, & H. S. Lee. 2017. Big data facilitation, utilization, and monetization: Exploring the 3Vs in a new product development process. *Journal of Product Innovation Management,* 34 (5), 640–58.

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature climate change*, *2*(10), 732-735.

Kahneman D, Tversky A. Intuitive prediction: Biases and corrective procedures. Decisions and Designs Inc Mclean Va; 1977 Jun.

Kaplan, S., & Tripsas, M. (2008). Thinking about technology: Applying a cognitive lens to technical change. *Research Policy*, *37*(5), 790-805.

Kaplan, S., & Vakili, K. (2015). The double‐edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, *36*(10), 1435-1457.

Kasurinen, J., & Knutas, A. (2018). Publication trends in gamification: A systematic mapping study. *Computer Science Review*, *27*, 33-44.

Kattel, R., Lember, V., & Tõnurist, P. (2019). Collaborative innovation and human-machine networks. *Public Management Review*, 1-22.

Kendig, C. E. (2016). What is proof of concept research and how does it generate epistemic and ethical categories for future scientific practice?. *Science and Engineering Ethics*, *22*(3), 735-753.

Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Ali, M., Kamaleldin, W., ... & Gani, A. (2014). Big data: survey, technologies, opportunities, and challenges. *The Scientific World Journal*, 2014. https://doi.org/10.1155/2014/712826

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review*, *94*(2), 211.

Klayman, J., & Ha, Y. W. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 596.

Kretschmer, T., & Khashabi, P. (2020). Digital Transformation and Organization Design: An Integrated Approach. *California Management Review*, *62*(4), 86-104.

Kumar, S., & Terpstra, D. (2004). The post mortem of a complex product development—lessons learned. *Technovation*, *24*(10), 805-818.

Kuusi, O., Cuhls, K., & Steinmüller, K. (2015). The topical collection on quality criteria for futures research: a short introduction.

Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International journal of information management*, *34*(3), 387-394.

Lallement, J., Dejean, S., Euzéby, F., & Martinez, C. (2020). The interaction between reputation and information search: Evidence of information avoidance and confirmation bias. *Journal of Retailing and Consumer Services*, *53*, 101787.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*, 6(70), 1.

Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, *36*(5), 700-710.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review*, *52*(2), 21-32.

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, *60*(3), 293-303.

Leventhal, L. M., Teasley, B. E., & Rohlman, D. S. (1994). Analyses of factors related to positive test bias in software testing. *International Journal of Human-Computer Studies*, *41*(5), 717-749.

Lieberman, M. B., & Asaba, S. (2006). Why do firms imitate each other?. *Academy of management review*, *31*(2), 366-385.

Lindebaum, D., Vesa, M., & den Hond, F. (2020). Insights from "the machine stops" to better understand rational assumptions in algorithmic decision making and its implications for organizations. *Academy of Management Review*, *45*(1), 247-263.

Lounsbury, M. (2008). Institutional rationality and practice variation: New directions in the institutional analysis of practice. *Accounting, organizations and society*, *33*(4-5), 349-361.

Mandelbaum, A., Momčilović, P., Trichakis, N., Kadish, S., Leib, R., & Bunnell, C. A. (2020). Data-driven appointment-scheduling under uncertainty: The case of an infusion unit in a cancer center. *Management Science*, *66*(1), 243-270.

Martínez-Román, J. A., Gamero, J., Tamayo, J. A., & Delgado-González, L. (2020). Empirical analysis of organizational archetypes based on generation and adoption of knowledge and technologies. *Technovation*, 102145.

Martino, J. P. (2003). A review of selected recent advances in technological forecasting. *Technological forecasting and social change*, *70*(8), 719-733.

Matt, C., Hess, T., & Benlian, A. 2015. Digital transformation strategies. Business & Information Systems Engineering, 57(5): 393-34

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, *90*(10), 60-68.

Meyer, G., Adomavicius, G., Johnson, P. E., Elidrisi, M., Rush, W. A., Sperl-Hillen, J. M., & O'Connor, P. J. (2014). A machine learning approach to improving dynamic decision making. *Information Systems Research*, *25*(2), 239-263.

Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management*, *57*(2), 103169.

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and e-Business Management*, *16*(3), 547-578.

Mikalef, P., Pappas, I. O., Krogstie, J., & Pavlou, P. A. (2020). Big data and business analytics: A research agenda for realizing business value. *Information & Management,* 57(1), 103237.

Milliken, A. L. (2014). Transforming big data into supply chain analytics. *The Journal of Business Forecasting*, *33*(4), 23.

Moe, W. W., & Schweidel, D. A. (2017). Opportunities for innovation in social media analytics. *Journal of Product Innovation Management*, *34*(5), 697-702.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological review*, *115*(2), 502.

Morris, A. (2015). A Practical Introduction to In-depth Interviewing. Sydney: SAGE.

Pandey, A. C., Rajpoot, D. S., & Saraswat, M. (2017). Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing & Management*, *53*(4), 764-779.

Papadopoulos, G. Z., Gallais, A., Schreiner, G., Jou, E., & Noel, T., 2017. Thorough IoT test-bed characterization: From proof-of-concept to repeatable experimentations. *Computer Networks*, *119*, 86-101.;

Parviainen, P., Tihinen, M., Kääriäinen, J., & Teppola, S. (2017). Tackling the digitalization challenge: how to benefit from digitalization in practice. *International journal of information systems and project management*, *5*(1), 63-77.

Pérez-Luño, A., Wiklund, J., & Cabrera, R. V. (2011). The dual nature of innovative activity: How entrepreneurial orientation influences innovation generation and adoption. *Journal of business Venturing*, *26*(5), 555-571.

Porac, J. F., & Thomas, H. (1990). Taxonomic mental models in competitor definition. *Academy of management Review*, *15*(2), 224-240.

Pröllochs, N., & Feuerriegel, S. (2020). Business analytics for strategic management: Identifying and assessing corporate challenges via topic modeling. *Information & Management*, *57*(1), 103070.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data*, *1*(1), 51-59.

Ragatz, G. L., Handfield, R. B., & Petersen, K. J. (2002). Benefits associated with supplier integration into new product development under conditions of technology uncertainty. *Journal of business research*, *55*(5), 389-400.

Raisch, S., & Krakowski, S. (2020). Artificial Intelligence and Management: The Automation-Augmentation Paradox. *Academy of Management Review*, (ja).

Remencius, T., Sillitti, A., & Succi, G. (2016). Assessment of software developed by a third-party: A case study and comparison. *Information Sciences*, *328*, 237-249.

Ries, E., 2011. *The lean startup: How 'today's entrepreneurs use continuous innovation to create radically successful businesses*. Currency.

Ries, E., 2017. *The startup way: how modern companies use entrepreneurial management to transform culture and drive long-term growth*. Currency.

Roca, J. B., Vaishnav, P., Morgan, M. G., Mendonça, J., & Fuchs, E. (2017). When risks cannot be seen: Regulating uncertainty in emerging technologies. *Research Policy*, *46*(7), 1215-1233.

Rosales-Morales, V. Y., Sánchez-Morales, L. N., Alor-Hernández, G., Garcia-Alcaraz, J. L., Sánchez-Cervantes, J. L., & Rodriguez-Mazahua, L. (2020). ImagIngDev: A New Approach for Developing Automatic Cross-Platform Mobile Applications Using Image Processing Techniques. *The Computer Journal*, *63*(5), 732-757.

Sallee, J. M. (2014). Rational inattention and energy efficiency. *The Journal of Law and Economics*, *57*(3), 781-820.

Sanday, P. R. (1979). The ethnographic paradigm (s). *Administrative Science Quarterly*, *24*(4), 527-538.

Scharfstein, D. S., & Stein, J. C. (1990). Herd behavior and investment. *The American economic review*, 465-479.

Schensul, S. L., Schensul, J. J., & LeCompte, M. D. (1999). *Essential ethnographic methods: Observations, interviews, and questionnaires* (Vol. 2). Rowman Altamira.

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). Analytics: the real-world use of big data: How innovative enterprises extract value from uncertain data, Executive Report. *IBM Institute for Business Value and Said Business School at the University of Oxford*.

Schulz-Hardt, S., Frey, D., Lüthgens, C., & Moscovici, S. (2000). Biased information search in group decision making. *Journal of personality and social psychology*, *78*(4), 655.

Sebastian, I. M., Ross, J. W., Beath, C., Mocker, M., Moloney, K. G., & & Fonstad, N. O. (2017). How Big Old Companies Navigate Digital Transformation. *MIS Quarterly Executive*.

Sgroi, D. (2003). The right choice at the right time: A herding experiment in endogenous time. *Experimental Economics*, *6*(2), 159-180.

Shrestha, Y. R., Krishna, V., & von Krogh, G. (2020). Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. *Journal of Business Research*, *123*, 588-603.

Simon, H. A. (1987). Making management decisions: The role of intuition and emotion. *Academy of Management Perspectives*, *1*(1), 57-64.

Simon, H. A. 1991. Bounded Rationality and Organizational Learning. Organization Science, 2(1): 125-134.

Sims, C. A. (2003). Implications of rational inattention. *Journal of monetary Economics*, *50*(3), 665-690.

Simsek Z., Vaara E., Paruchuri S., Nadkarni S., & Shaw J.D. (2019). New Ways of Seeing Big Data. From the Editors. *Academy of Management Journal*, 62 (4).

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263-286.

Solberg, E., Traavik, L. E., & Wong, S. I. (2020). Digital Mindsets: Recognizing and Leveraging Individual Beliefs for Digital Transformation. *California Management Review*, 62(4), 105-124

Stock, G. N., & Tatikonda, M. V. (2008). The joint influence of technology uncertainty and interorganizational interaction on external technology integration success. *Journal of operations management*, *26*(1), 65-80.)

Strang, D., & Still, M. C. (2006). Does ambiguity promote imitation, or hinder it? An empirical study of benchmarking teams. *European Management Review*, *3*(2), 101-112.

Street, C. T., & Meister, D. B. (2004). Small business growth and internal transparency: The role of information systems. *MIS quarterly*, 473-506.

Stubbart, C. I. (1989). Managerial cognition: A missing link in strategic management research. *Journal of Management Studies*, *26*(4), 325-347.

Sukumar, S. R., & Ferrell, R. K. (2013). 'Big Data'collaboration: Exploring, recording and sharing enterprise knowledge. *Information Services & Use*, *33*(3-4), 257-270.

Sun, H. (2013). A longitudinal study of herd behavior in the adoption and continued use of technology. *Mis Quarterly*, 1013-1041.

Surbakti, F. P. S., Wang, W., Indulska, M., & Sadiq, S. (2020). Factors influencing effective use of big data: A research framework. *Information & Management*, *57*(1), 103146.

Sweeny, K., Melnyk, D., Miller, W., & Shepperd, J. A. (2010). Information avoidance: Who, what, when, and why. *Review of general psychology*, *14*(4), 340-353.

Teare, M. D., Dimairo, M., Shephard, N., Hayman, A., Whitehead, A., & Walters, S. J. (2014). Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. *Trials*, *15*(1), 264.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation*, *27*(2), 237-246.

Uras, M., Cossu, R., Ferrara, E., Liotta, A., & Atzori, L. (2020). PmA: A real-world system for people mobility monitoring and analysis based on Wi-Fi probes. *Journal of Cleaner Production*, 122084.

Urbinati, A., Bogers, M., Chiesa, V., & Frattini, F. (2019). Creating and capturing value from Big Data: A multiple-case study analysis of provider companies. *Technovation*, *84*, 21-36.

Vahn, G. Y. (2014). Business analytics in the age of Big Data. *Business Strategy Review*, *25*(3), 8-9.

Vaidya, S., Ambad, P., & Bhosle, S. (2018). Industry 4.0–a glimpse. *Procedia Manufacturing*, *20*, 233-238.

Van de Ven, A. H., & Poole, M. S. (1995). Explaining development and change in organizations. *Academy of management review*, *20*(3), 510-540.

Van Knippenberg, D., Dahlander, L., Haas, M. R., & George, G. (2015). Information, attention, and decision making. *Academy of Management Journal*, 58 (3), 649-657.

Van Maanen, J. (1979). The fact of fiction in organizational ethnography. *Administrative science quarterly*, *24*(4), 539-550.

Van Rijmenam, M., Erekhinskaya, T., Schweitzer, J., & Williams, M. A. (2019). Avoid being the Turkey: How big data analytics changes the game of strategy in times of ambiguity and uncertainty. *Long Range Planning*, *52*(5), 101841.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS quarterly*, 21-54.

Vidgen, R., Shaw, S., & Grant, D. B. (2017). Management challenges in creating value from business analytics. *European Journal of Operational Research*, *261*(2), 626-639.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, *34*(2), 77-84

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J. F., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356-365.

Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*, 3-13.

Watson, M. (2014). Three things you should know about big data and analytics. *CSCMP's Supply Chain Quarterly*.

Weatherall, J. O., & O'Connor, C. (2020). Conformity in scientific networks. *Synthese*, 1-22.

Wessel, M. (2016). You Don't Need Big Data—You Need the Right Data. *Harvard Business Review*, *3*.

Williams, B. D., Roh, J., Tokar, T., & Swink, M. (2013). Leveraging supply chain visibility for responsiveness: The moderating role of internal integration. *Journal of Operations Management*, *31*(7-8), 543-554.

Yadegaridehkordi, E., Hourmand, M., Nilashi, M., Shuib, L., Ahani, A., & Ibrahim, O. (2018). Influence of big data adoption on manufacturing companies' companies' performance: An integrated DEMATEL-ANFIS approach. *Technological Forecasting and Social Change*, 137, 199-210.

Yin, R.K. (2013), *Case Study Research: Design and Methods*. Sage Publications

Ylijoki, O., & Porras, J. A Recipe for Big Data Value Creation. *Business Process Management Journal*, *25*(5), 1085-1100.

Yoo, Y. (2012). The tables have turned: How can the information systems field contribute to technology and innovation management research?. *Journal of the association for information systems*, 14(5), 4.

Yoo, Y. (2015). It is not about size: a further thought on big data. *Journal of Information Technology*, *30*(1), 63-65.

Yoon, B., Jeong, Y., Lee, K., & Lee, S. (2020). A systematic approach to prioritizing R&D projects based on customer-perceived value using opinion mining. *Technovation*, *98*, 102164.

Zhang, H., van der Bij, H., & Song, M. (2020). Can cognitive biases be good for entrepreneurs?. *International Journal of Entrepreneurial Behavior & Research*.

Zhang, S. X., & Cueto, J. (2017). The study of bias in entrepreneurship. *Entrepreneurship theory and Practice*, *41*(3), 419-454.)

Zhang, Y., Huang, Y., Porter, A. L., Zhang, G., & Lu, J. (2019). Discovering and forecasting interactions in big data research: A learning-enhanced bibliometric study. *Technological Forecasting and Social Change*, *146*, 795-807.

APPENDIX A

Example of the Three Items Analyzed with the Use of the Protocol

Date/ Title	**The goal of the meeting/ document/message**	**The focus on Big Data dimension, quality, or related to Big Data aspects**	**Key issues to solve (if any)**	**Moments of disagreements/ tension between participants (if any)**	**Found solutions (if any)**	**People presented**
Online meeting, Manchester City, 30.05.2019	Planning locations of the sensors in Manchester	**Veracity:** Reliable sample to get representative data	Which locations to include, as it directly influences the output	Some locations that are representable could not be included due to the city policies (not possible to install in this public space)	Discussing during the live call the Creating a map with the locations in satellite mode on Google Maps, discussing the final solution with the city	WiFi sensors technology provider manager, coordination manager, project lead
The slack message, 05.06.2019	Set up of the sensors	**Veracity:** Visibility radius of the CCTV sensors	Lost data Not reliable data	NA	Create a Risk map Use an alternative power source follow the installs specifications	4 technology providers' managers, data scientist responsible,
Pilot Scope document	Importance of the locations	**Veracity:** geolocation	Strict suggestions to locations of the sensors to enable reliable and scalable results	N/A	Follow the guidance in the Scope	N/A