O. AK

METU

COMPUTATIONAL PARADIGM SHIFT: A HISTORICIZATION OF RECENT DEVELOPMENTS IN SOCIAL SCIENCES

ONURHAN AK

JULY 2019

COMPUTATIONAL PARADIGM SHIFT: A HISTORICIZATION OF RECENT DEVELOPMENTS IN SOCIAL SCIENCES

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF SOCIAL SCIENCES

OF

MIDDLE EAST TECHNICAL UNIVERSITY

BY

ONURHAN AK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR

THE DEGREE OF MASTER OF SCIENCE

IN

THE DEPARTMENT OF SOCIOLOGY

JULY 2019

Approval of the Graduate School of Social Sciences

Assoc. Prof. Dr. Bahar Öz
Director (Acting)

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Ayşe Saktanber
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assist. Prof. Çağatay Topal
Supervisor

Examining Committee Members

Assoc. Prof. Erdoğan Yıldırım (METU, SOC)

Assist. Prof. Çağatay Topal (METU, SOC)

Assist. Prof. Emek Barış Kepenek(Başkent Uni, SOC)

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Onurhan Ak

Signature :

ABSTRACT

COMPUTATIONAL PARADIGM SHIFT: A HISTORICIZATION OF RECENT DEVELOPMENTS IN SOCIAL SCIENCES

Ak, Onurhan
M.S., Department of Sociology
Supervisor: Assist. Prof. Çağatay Topal

July 2019, 117 pages

The use of big data and computational methodologies in social sciences have been getting attention and these changes in methodologies have been declared a "paradigm shift" in the contemporary literature. Motivated by such claims, this thesis discusses Computational Social Science by referring to Kuhnian theory of scientific revolutions and emphasizes the emergence of social big data in the former's rise. Later, a historicization and contextualization of social big data is provided by employing conceptual tools supplied by Jean-François Lyotard and Nigel Thrift to understand and discuss "scientific change" in general and the nature of this particular case of scientific change; the rise of computational methodologies.

Keywords: paradigm shift, Computational Social Science, big data, computerized societies, knowing capitalism.

ÖZ

BİLGİSAYIMSAL PARADİGMA DEĞİŞİMİ: SOSYAL BİLİMLERDEKİ SON GELİŞMELERİN BİR TARİHSELLEŞTİRİLMESİ

Ak, Onurhan
Yüksek Lisans, Sosyoloji Bölümü
Tez Yöneticisi: Dr. Öğr. Üyesi Çağatay Topal

Temmuz 2019, 117 sayfa

Sosyal bilimler literatüründe kompütasyonel metodolojilere ve büyük veri odaklı çalışmalara olan ilgi son yıllarda gittikçe artmakta ve bu tarz çalışmaların yaygınlaşması literatürde sosyal bilimlerde bir paradigma değişimi olarak nitelendirilmektedir. Bu savlardan yola çıkarak bu tez Kompütasyonel Sosyal Bilimi Kuhn'un bilimsel devrimler teorisine dayanarak tartışmakta ve sosyal büyük verinin sosyal bilimlerde kompütasyonel metodolojilerin yaygınlaşmasındaki rolünü vurgulamaktadır. Sonrasında Jean-François Lyotard ve Nigel Thrift'in analizlerine başvurularak sosyal büyük verinin ortaya çıkışının tarihsel ve bağlamsal bir analizi yapılarak sosyal bilimlerde kompütasyonel metodolojilerin yaygınlaşmasının dinamikleri tartışılmakta ve 'bilimlerde değişim' sorunu irdelenmektedir.

Anahtar Kelimeler: paradigma değişimi, bilgisayımsal sosyal bilim, büyük veri, bilgisayarlaşmış toplumlar, bilen kapitalizm.

To Çirkin and Hırçın

ACKNOWLEDGEMENTS

This thesis is a result of years of discussions with people that, over time, I came to hold very dear.

First of all, I want to thank my advisor and teacher Çağatay Topal for accepting me as his student, for his support and patience and without whom this thesis would be a collection of disorganized ideas and observations. I want to thank Erdoğan Yıldırım and Barıs Mücen for their guidance and for introducing me to their effort to make sense of the world, for teaching me how to approach and question it and without whom I would not have found my path in this effort. I am also thankful to Ulaş Murat Altay, an intellectually kindred spirit, for his friendship through the years.

I also want to thank to my family, I am grateful for their endless support and encouragement. Without them this study will not be possible. I am especially grateful for moments of kidding around I shared with my sister Bilgesu Ak during this time.

Finally, I would like to thank Seda Baykal for making my last 5 years memorable. It is her love, support and companionship that made difficult times bearable. I am glad that we got to share this beautiful, exhausting, valuable, and rewarding effort with each other.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

CSS	Computational Social Science
ICT	Information Communication Technologies
IaaS	Infrastructure as a Service
PaaS	Platform as a Service
SaaS	Software as a Service
SaaSS	Software as a Service Substitute

CHAPTER 1

INTRODUCTION

1.1 Research Focus

The aim of this thesis is to understand the factors that have contributed to changes in the methodologies in social sciences towards computational approaches. This is to be done by inquiring about the characteristics of Computational Social Science that make it a particular area of study and by showing the outside factors that have been acting as conditions of possibility of such changes.

The choice of this particular object of study depends on the fact that it is one of the most comprehensive changes as well as issues that have been talked about in a variety of disciplines in sciences and humanities. Taking this change as its primary problematic, this study will focus firstly on locating the particular features of computational methodologies in social sciences and then data as it is understood today, and finally will show the characteristics of it that makes the latter a particular phenomenal domain, an object of analysis. Then, this thesis will show the historical nature of the factors that have contributed to its emergence and thereby will provide the literature with a case where the relationship between the conditions of knowledge production and knowledge is evident.

To conceptualize the issue at hand, the thesis will rely on Thomas Kuhn's (1962) theoretical framework for understanding scientific change that he presented in "The Structure of Scientific Revolutions". By making use of his theory of scientific change, this thesis aims to show how the recent changes in the methodologies can be understood as a paradigm shift in Kuhn's terms while simultaneously showing the limits of the Kuhnian theory by providing an analysis that takes not only internal scientific dynamics but a wider set of relationships into account. The latter part of the analysis mainly relies on Jean-François Lyotard's theory of knowledge in contemporary societies and Nigel Thrift's analysis of contemporary capitalism. The aim here is to show the histories of factors that have contributed to the rise of computational methodologies in social sciences and make sense of it in Kuhnian terms to directly address the literature where the mentioned changes in social sciences are declared a "paradigm shift".

As such, this research aims to provide an explanation for scientific change by making use of a recent case but, contrary to Kuhnian approach, it does not intend to stay within the borders of science, disciplines or academia to do so. Instead of working through the binary separation of science and non-science which implies the purity of science from factors that supposedly do not reside in it, this thesis deals directly with the question of the influence of "non-scientific" factors upon the ways the science is and thought to be done. To understand these factors the notion of *computerized societies* presented by Lyotard and characteristics of contemporary capitalism presented by Nigel Thrift will be critical. These will be discussed in detail in Methodology and Literature Review sections.

1.2 Background

According to a report from 2017, around 90 percent of the data that is created in the world has been created in the last two years. The same report predicts that with the emerging technologies the growth in data production will increase even more (IBM, 2017). This increased production in data and the increase in capability to store in amounts unprecedented before and process them with specialized tools are referred in the literature as well as in public discourse as "the data revolution" (Kitchin, 2014b).

Many have predicted that this revolution will transform how we think about science, the methods that we use to produce knowledge. This transformation, sometimes referred to as the "Computational Turn", is tightly knit with the emergence of a new form of data, "big data". The term is often used to refer to emerging enormous datasets, but there is much more to it than its size. For example, a definition from Apache Hadoop in 2010 defined the term as "datasets which could not be captured, managed, and processed by general computers within an acceptable scope." (M. Chen, Mao, & Liu, 2014). Another definition, presented by Rob Kitchin, focuses specifically on three characteristics of the big data. Kitchin underlies that big data is specifically characterized by its *volume* meaning that it is huge, consisting of petabytes of data, by its *velocity,* meaning that it is not collected at a certain point in time but rather is being generated continuously and by its *variety,* meaning that it is rather messy when compared with the traditional datasets that have been collected at a particular time with pre-prepared variables and fields (Kitchin, 2013). Such data emerges in a state where

> …a wide, deep torrent of timely, varied, resolute and relational data that are relatively low in cost and, outside of business, increasingly open and accessible. (Kitchin, 2014b, p. XV).

The emergence of this state is attributed mainly to the recent developments in information and communication technologies, the proliferation of digital devices that connect to the World Wide Web that generate vast amounts of data as a result people's interactions with each other, with the devices themselves and from digital processes like transactions and so forth (Kitchin, 2014b). The resulting datasets are very much different than the traditional ones and, therefore, require specialized tools and skills to process them and make them usable in the production of knowledge.

This required transformation in tools and skills is sometimes referred as a paradigm shift. For example, Gray has argued that after experimental, theoretical and computational science paradigms[1], a new data driven scientific paradigm that is capable to affect multiple disciplines is underway (T. Hey, Tansley, & Tolle, 2009). Gray's paradigm change, being different from the conceptualization of the inventor of the term, Thomas Kuhn, depends on the changes in the forms of data and tools used to analyze them (Kitchin, 2014a). This line of argumentation is followed by many and resulted in claims such as

[1] What Gray means by "computational science" is a model-driven, simulation focused branch of computational science. This will be discussed in detail in Chapter 3.

> There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot. (Anderson, 2008)

As can be seen, the data revolution and the so-called Computational Turn is thought to be capable of transforming science, its ontological and epistemological principles, in a very much profound, fundamental way. According to Anderson's statement above, scientific research no longer has to know and conceptualize what it aims to know beforehand for every phenomenon can be made evident by statistical methods that trace patterns.

Recognizing this issue, Savage and Burrows, in their highly popular 2007 article, "The Coming Crisis of Empirical Sociology", identified the situation as a crisis. It is a crisis because sociology is no more equipped with the appropriate tools and methods to deal with the emerging forms of data which resulted in a situation where the location of most of the production of knowledge of the social shifted from universities and academia to industry and business (Savage & Burrows, 2007). Therefore, for them, the crisis is one of jurisdiction that requires sociology to respond adequately to the challenges proposed by the digital age.

The effect of big data appears to be not only limited to a change in the tools but also refer to a computational transformation in thought and research (Burkholder as cited in Boyd & Crawford, 2012). It has some sort of scientific use, and from the statements of Savage and Burrows and as well as the popularity of their argument and their article, it has a pressuring effect upon disciplines, forcing them to respond and change if necessary with respect to the changing conditions of the age. Therefore, the questions are being asked "How can we combine the depth of inquiry in the social sciences with the scale and robustness of statistics and computer science?" (Raghavan, 2014, p. 1).

In accordance with their call, there has been recent developments in Computational Social Sciences and in digital humanities, trying to fuse social sciences with computational tools to be able to leverage the massive amounts of data produced by the society. The effects of this has been identified by scholars such as David Berry who have stated that

> …computational technology has become the very condition of possibility required in order to think about many of the questions raised in humanities today. (Berry, 2011a, p. 2)

The problem of this thesis is this suggested transformation, whether it is a paradigm shift as conceptualized by Thomas Kuhn himself and if not, this thesis aims to determine what other factors may have been effective in this transformation. For this transformation to be able to be characterized as a paradigm shift, it has to fit in with criteria that are presented by Thomas Kuhn.

Thomas Kuhn, acting on the premise that scientific community cannot practice its craft without some sort of shared and received beliefs, conceptualized paradigm as something that is at the beginning an achievement that can attract a large portion of scientific community and that is open-ended enough so that there is still a good amount of work that has to be done by the scientists that follow the paradigm (Kuhn, 1970). What Kuhn calls "normal science" operates within a given paradigm that supplies it with questions that are guaranteed to be answerable with the fundamental assumptions about the nature of the world and the things, the corresponding methods that are also provided by the paradigm. However, Kuhn also states that it is common in normal science, that is organized around a certain paradigm, that there appear anomalies phenomena that are unaccountable. In fact, it is only with respect to a paradigm an anomaly can show itself. Scientific change is a result of the accumulation of these anomalies, preceded by a *scientific crisis* where the fundamentals of the established paradigm are called into question. This crisis is the condition of possibility of extraordinary science that can come up with a new paradigm, a paradigm shift (Kuhn, 1970).

In a more general level, Kuhn himself paves the way for an analysis that is to be attempted here. For example, one

> …must compare the community's paradigms with each other and with its current research reports. In doing so, his object is to discover what isolable elements, explicit or implicit, the members of that community may have

> *abstracted* from their more global paradigms and deployed as rules in their research. (Kuhn, 1970, p. 43)

An analysis of the research reports to come up with abstracted rules of conducting research is not an easy task. However, in Kuhn's theory establishment of a scientific paradigm is, in the end, a discursive phenomenon that is a matter of allegiances within the scientific community which are formed for various reasons on the side of the scientists, therefore, the primary material of this analysis will be the claims of the scientists themselves and their comparisons of ways of doing social science. By doing that it may still be possible to provide characteristics that make Computational Social Science distinct and to identify it in its particularity. This is justified for even if one searches for rules that guide science under a particular paradigm because "... the existence of a paradigm need not even imply that any full set of rules exists." (Kuhn, 1970, p. 44).

A paradigm is understood in this thesis as it is hinted by Kuhn, it is a collective phenomenon that enables a particular branch of science or discipline to be practiced and enables its practitioners to communicate on the basis of it, without feeling the need to provide justifications for their methods, conceptual tools and such (Kuhn, 1970).[2]

Kuhn's theory of scientific change highly depends on the internal dynamics of science and the scientific community. Based on the characteristics that Kuhn provides in his theory, this thesis will evaluate whether the crisis identified by Savage and Burrows above is one that prepares the ground of a paradigm shift which is referred by Berry as the Computational Turn above. In addition to Kuhn's theory, to be able to account for factors that stand outside the domains of science and scientific community this thesis will also refer to Jean-François Lyotard's famous book "Postmodern Condition: A Report on Knowledge" and Nigel Thrift's 2007 book "Knowing Capitalism".

[2] It is hinted or perhaps even stated explicitly in his work for he himself builds the resemblance between the concepts of "scientific paradigm" and Ludwig Wittgenstein's concept of "language games" which functions as a "... network of overlapping and crisscross resemblances." It is the existence of it that enables identification of an object or a notion all at once. In this sense, it is what enables communication between members of the scientific community (Kuhn, 1970, p. 45).

Lyotard's conceptualization of "computerized societies" will be vital to understand the fundamental change that is brought about by computational technologies. He uses the term "computerized societies" to refer to a state where knowledge is almost always in quantifiable form that enables it to be processed efficiently and does more and more resemblance a commodity (Lyotard, 1984). The resemblance between Lyotard's conceptualization of knowledge in computerized societies and the data or big data is fairly apparent. By the same token, Thrift's conceptualization of "knowing capitalism", capitalism in a state where information technologies are prominent and the nature of commodities and commodity relations are being transformed will also be critical (Thrift, 2005). Both of them allow this thesis to push the boundaries set by Thomas Kuhn's theory of scientific revolutions and to include historical and eco nomic factors that may have contributed to the case at hand.

1.3 Methodology

In this section, first of all, the theoretical framework that this thesis follows will be presented. Since the question at hand is one of scientific change and more specifically one of "paradigm shift", Kuhn's theory of scientific revolutions will be the core theoretical framework for a significant part of the thesis. After briefly elaborating on the theoretical line and my research question, I will present the relevant conceptual tools that will be of use in this study.

The conceptual framework supplied by Kuhn enables this thesis to take the recent discussions on the Computational Turn and the digital shift as an object of analysis and assess it in terms of a paradigm shift. This part of the study is a direct response to the literature for in the relevant literature recent rise of computational methods in social sciences has been considered as a paradigm shift and therefore is justified on the basis of this. The main problem this thesis deals with is whether the so-called "Computational Turn" or "digital shift" in social sciences constitute a paradigm shift in the Kuhnian sense. To assess the object of analysis at hand

Kuhnian concepts of scientific crisis, normal science and scientific paradigm will be employed.

A scientific crisis is one of the few preconditions of paradigm change in sciences. It is defined by Kuhn as the breakdown of normal puzzle-solving activity in a discipline mainly as a result of an accumulation of unresolvable and unexplainable anomalies that are capable of disturbing the established disciplinary matrix (Kuhn, 1970). In the examples given by Kuhn, all novel theories that trigger a new practice of normal science are a direct response to a crisis. Therefore, the notion of scientific crisis is critical to locate the dynamics of the Computational Turn in social sciences, the question of the practicalities of identification of a crisis will be discussed towards the end of this section.

The relationship between scientific crisis and normal science requires this analysis to conceptualize Computational Social Science as a normal science. Operationalizing the concept of normal science enables this analysis to identify the particular practices of doing science, in other words, it enables us to differentiate computational and data driven social science from the previous ways of doing social science. Normal science depends on paradigms as they are networks of beliefs, assumptions, commitments enable disciplines to shape the world, the empirical domain into a more or less structured entity where the object of analysis, the questions about it, the methods that can be used to answer these questions are for the most part are supplied (Kuhn, 1970). Through the notions of normal science and scientific paradigm, the particularity of Computational Social Science can be identified and compared. The critical part is to determine whether the computational social science is a scientific paradigm with its own phenomenal domain to the point that it is incommensurable with the previous one for

> …when the paradigms change, the world itself changes with them. Led by a new paradigm, scientists adopt new instruments and look in new places… It is rather as if the professional community had been suddenly transported to another planet where familiar objects are seen in a different light and are joined by unfamiliar ones as well. (Kuhn, 1970, p. 111)

What enables Kuhn to state such an argument besides his empirical analysis is the power of scientific paradigms and scientific revolutions to organize the world in a certain way and populate it with entities as well as to shape the sensibility of scientists. Therefore, to answer the question whether such a fundamental transformation has happened in the last few years with computational methodologies requires us to track these changes in the organization of the world and the sensibilities of the scientists. This will be done in this thesis through the analysis of the works of both sides of the supposed paradigm shift namely those scholars that provide a reflexive analysis of the changes that are happening and those scholars that champion the new methodologies to bring out the points of distinction. The justification for such a method can be found in Kuhn’s work as well. Kuhn clearly states many times that paradigm shift is not a result of the accumulation of anomalies alone. For there to be a paradigm shift, a competitor paradigm is needed. The success of the victorious paradigm does not depend on its ability to answer the questions the previous one fails to. Again, a paradigm “... must seem better than its competitors, but it need not, an in fact never does, explain all the facts with which it can be confronted.” (Kuhn, 1970, p. 17-18).

There are many qualities of the new paradigm that enables it to attract scientists and convince scientists to adhere to it besides its scientific vigor. Kuhn states there are always groups and individuals in the scientific community that resist the new paradigm and continue to adhere to the old one. Therefore, I think the reasons that are expressed by scholars and scientists that champion computational methods are critical to study to understand the failures of the supposedly old paradigm and the advantages of the new one. These advantages and identifications of failures of the old paradigm are the points of contact and confrontation between paradigms that function as self-demarcation points for the new paradigm. Building on these, the first part of the analysis will be a search for characteristics of Computational Social Science, its theoretical framework, assumptions that make it a particular scientific paradigm.

The second part of the thesis deals with a follow-up question. If we are to understand the rise of computational methods in social sciences as a paradigm shift, the Kuhnian framework can only explain this paradigm shift insofar as we limit ourselves to the scientific domain only. This question deals with whether normal science is a closed box as Kuhn presents it to be. Specifically, whether other factors contribute to the rise of computational methods and methodologies in social sciences. Doing so not only will expand our understanding of the particular case at hand but also will extend and push the limits of Kuhnian analysis and hopefully will bring us closer to the nature of scientific paradigm shifts and normal science. This is a matter of contextualizing the Computational Turn, to do that I will adhere to the works of Jean-François Lyotard where he deals with the position of knowledge in computerized capitalist societies and will also refer to characteristics of this "knowing capitalism" identified by Nigel Thrift.

Many of the scholars presented in the literature review part of this thesis admit that the first ones that employed computational methods to know the social were the companies especially those in the business of software development and similar, related sectors. The idea of the tendency of contemporary capitalism to know as claimed by Thrift and the specific relationship between technology, capitalism, and knowledge as identified by Lyotard are the building blocks of the contextualization of the Computational Turn as a paradigm shift in social sciences to understand the effects of this context on its emergence and origin. Doing so, the thesis aims to show the points of interaction between seemingly separate domains of science, which is even more present in the works of Thomas Kuhn, and the social and the economic influences.

This point interaction between different domains in our particular case will be identifying the emergence of big data. The claims of the paradigm shift in the literature refer to a specific Computational Social Science that is particularly data-driven which is enabled solely by the emergence of enormous datasets. In Kuhnian terms, big data appears to be the unaccountable that pushes the disciplines to retool, forcing disciplines to reconsider their fundamentals by initiating a crisis, in other words, to change the fundamental assumptions and beliefs of the disciplines for it cannot be accounted using the traditional tools of social sciences. As such, it is necessary to locate the features of big data in the context of its emergence to be able to understand it in its historicity. This is where Lyotard's and Thrift's account of contemporary capitalist societies come into play to explain the relationship between science and economy, specifically capitalism. The systematic history provided by Lyotard enables this thesis, along with the discussions on knowing capitalism, to historicize the Computational Turn. Lyotard, seeing science as a form of discourse that more or less operates like language game argued for the necessity of science to refer to a legitimizing ground. Stating that postmodern societies lack the previous forms of legitimization that referred to grand narratives, Lyotard argues that scientific knowledge production is legitimized by referring to the principle of efficiency, performativity in computerized capitalist societies (Lyotard, 1984, p. 44). This principle of performativity not only legitimizes science and knowledge per se but also the institutions of science and of knowledge production which are now judged, according to Lyotard, by their contribution to the optimization of system's performance. The increased reliance on technological tools in scientific knowledge production necessarily brings the questions who will be able to afford to do science and what are the determinants of scientific knowledge (Lyotard, 1984).

In Lyotard's account under capitalist circumstances knowledge more and more takes the form of a commodity whose worth is assessed by not referring to the truth value of the produced knowledge but by to its contribution to the accumulation of capital and institutions of education and knowledge production comes to be legitimized by their contribution to the efficiency of the system (Lyotard, 1984). The emergence of data as a valuable commodity and computational methods as a method of producing knowledge can be understood within this framework that Lyotard supplies. To do this, we will mainly focus on the emergence of new business strategies of companies that are in Information and Communication Technologies sector to what is called Software as a Service, Platform as a Service, Infrastructure as a Service and so forth. These new business strategies depend less and less on the price paid by consumers in exchange for the products, goods, and services but rather increasingly on the information collected from the users of the product or the platform. For such business strategies to be reliable, a particular set of conditions are necessary. They require a specific form of capitalism where information and its commodification can become a moment in accumulation of capital. While the mainstream analysis of the emergence of Computational Social Science mostly emphasizes and build upon the advancements in information and communication technologies, focusing on the capitalist context enables us to bring about the specific relationship between capitalism, technology, and knowledge. That specific relationship is that for massive data extraction and processing platforms and infrastructure to be actualized, they need to profitable. Following Lyotard, this thesis will place the economic conditions of knowledge production prior to technological advancements for the latter require investment to be made which in turn requires the promise of profit. Moreover, Thrift's analysis of the changing form of commodities and commodity relations in contemporary capitalism can be located in the center of our problematic for they particularly apply to digital commodities:

1. Commodities are now being produced together by both the producers and the consumers.
2. They are produced in the context of an "experience economy" that requires increasing effort on the side of consumers to invest on a particular commodity.
3. Commodities are produced or "developed" continuously and necessarily they are interactive (Thrift, 2005).

As I will show, these new developments in commodity relations can be located in the specific case of service-based business strategies mentioned above where the profit is made from the information provided by the users of the products. The emergence of these digital platforms and products were decisive factors in the materialization of social big data. It is in a feedback loop that takes place in a capitalist setting that makes information a part of the cycle of capital accumulation and where the information is continuously processed to optimize the apparatuses of data collection and is commodified to generate revenue. Big data, understood as a distinct form of information, one that exists in its exchange value, that arises out of the mentioned new business strategies is the bridge that ties together scientific domain and the economic conditions of knowledge production.

The context that big data emerges out of determines its content as well as its form, making it distinct from traditional datasets and thereby unaccountable using the conventional tools and methodologies. Locating the effects of the process of commodification that makes up the social big data will help us (1) understand why it pushes disciplines to undergo radical changes and (2) bring out the affectable nature of normal science which is mostly absent in Kuhnian framework. As such, in the third chapter, we will try to understand the Computational Turn as a paradigm shift and establish why big data is such a critical force in that process within the borders of the framework presented by Thomas Kuhn. By operationalizing Kuhnian concepts, we will discuss Computational Social Science as a scientific paradigm, its theoretical framework, how social world and the entities in it are thought to be and its advantages in comparison to the traditional social science. In the fourth chapter, we will recontextualize the Computational Turn by explicitly focusing on the emergence of big data with respect to its commodified nature, show the process out of which it emerges and what it means for the discussions revolving around the notion of paradigm shift.

Before we continue, a final issue must be addressed and that is the applicability of Kuhn's theory to social sciences. Kuhn rarely addresses social sciences but when he does it appears that social sciences are much more characterized by heterogeneity of accounts of their objects. For example, he states that student in disciplines of history, philosophy and social sciences

> …is constantly made aware of immense variety of problems that the members of his future group have, in the course of time, attempted to solve. Even more important, he has constantly before him a number of competing and incommensurable solutions to these problems, solutions that he must ultimately evaluate for himself. (Kuhn, 1970, p. 165)

It seems that social sciences are perhaps closer to, in Kuhnian framework, a preparadigmatic period where the dominance of one paradigm is not established meaning that science is not mature enough[3] (Peterson, 1981). This does not pose a problem for our analysis for a few reasons. First of all, even if one accepted that social sciences are not mature enough and are characterized by heterogeneity, the problem of paradigm shift in social sciences could be posed as the establishment of a paradigm for which case Kuhnian concepts and tools still apply. Secondly, the problem of this thesis is for the most part supplied by the contemporary literature where the rise of computational methodologies in social sciences already conceptualized as a paradigm shift. And finally, as will be seen in the analysis of the points of confrontation mentioned above, something rather interesting happens in the process of self-identification of Computational Social Science where it distinguishes itself from the traditional science. What happens is that, assuming social sciences are characterized by heterogeneity of accounts or perhaps paradigms as Kuhn argues, the past practices are unified, brought together when contrasted with the new one. The new paradigm does not posit itself as another possibility in the heterogeneity of ways the social can be accounted but unifies the past and situates itself as a next step in the evolution of social science. This point will be clearer once we elaborate on our case in Chapter 3.

1.4 Significance of the Thesis

The significance of this thesis stems from two sources.

First, this thesis is a direct response to the contemporary literature that considers the rise of computational methodologies and use of social big data is social science a Kuhnian paradigm shift. Showing the scientific framework upon which

[3] It must also be noted that, as Peterson also argues, Kuhn did not provide an examination of social sciences but took it in a rather taken for granted fashion that social sciences are characterized by disagreement and stands in opposition to the natural sciences which are characterized by agreement and unity (Peterson, 1981).

these methodologies and the its following promises are built upon, this thesis is an assessment of these claims.

Secondly, this thesis brings particular historical and economic conditions in the analysis of the rise of computational methodologies in social sciences and the emergence of social big data by relying on Lyotard's theory of knowledge in computerized societies and Thrift's identification of characteristics of knowing capitalism. Thereby, this thesis shows the lack of historical analysis both in the mainstream accounts of the issue at hand and in Kuhn's theory of scientific change.

Finally, by showing the particular practices and circumstances that act as conditions of possibility of a particular form of social science, this thesis both contributes to the literature of sociology of knowledge and science and technology studies for it provides a discussion on scientific progress through a particular case and provides a reflexive account of the recent changes that affect social sciences.

1.5 Thesis Plan

This thesis will begin by introducing the discussions on rise of computational methodologies in social sciences to provide a context for further analysis.

In Chapter 2, firstly, I will present the recent discussions about the position of social sciences in an age characterized by information technologies and digitalization. I will especially focus on the idea of 'crisis' as it is posed by Savage and Burrows where sociology is being challenged as a result of shifting locus of knowledge production (Savage & Burrows, 2007). Accordingly, I will also discuss the propositions in the contemporary literature to transform the methods of social sciences so that the latter can adapt to the requirements of the age. Later, I will show the critical position of big data in transformations of methodologies in social sciences. I will particularly focus on the discussions that revolve around the dynamics of emergence of big data and that present it as a force that forces social sciences to retool. Next, I will present the relevant arguments and ideas of Jean-François Lyotard and Nigel Thrift that this thesis will employ to further problematize its case. I will specifically discuss the issue of legitimization of knowledge and science in contemporary capitalist societies where knowledge becomes a key economic resource and the nature of commodities and commodity relations change accordingly (Lyotard, 1984, Thrift, 2005). In the final section of Chapter 2, I will introduce Thomas Kuhn's theory of scientific revolutions and discuss some key concepts that will be operationalized in the following chapters.

In Chapter 3, I will provide an assessment of Computational Social Science as a scientific paradigm in Kuhnian sense. I will show the world Computational Social Science constructs and operates in. To do so, the ontological and epistemological principles of Computational Social Science (CSS) will be discussed. I will first show the historical roots of CSS and how complexity theory acts as a foundation that defines the nature of the social and how the methods and methodologies of CSS are legitimized on the basis of this definition. Later, I will discuss the continuities and differences between two traditions of CSS, simulation-focused and data-focused Computational Social Science and how social big data is incorporated the scientific framework of CSS as a representation of social complexity so that CSS, as a scientific paradigm, can be understood as an 'inflexible box' that supplies the domain of knowable phenomena, questions, and methods (Kuhn, 1970). After establishing this, I will discuss the features of social big data that makes it unaccountable by traditional social scientific theories and tools, and will show how, ultimately, this has to do with the lack of scientific purpose in the data collection. Finally, I will present the advantages of CSS in comparison to traditional social science where the said advantages operate as a point of self-identification, 'promises' in the Kuhnian terminology, that find their place in the process of construction of allegiances in the scientific community.

After establishing the critical role of social big data in the recent transformations in social science in Chapter 3, in Chapter 4, I will provide a historicization of the emergence of social big data. To do so, I will rely on the analysis of the position of knowledge in contemporary computerized societies of Lyotard, how principle of performativity comes to be the main legitimizing mechanism of knowledge and knowledge production, and will show how the changing forms of commodities and commodity relations identified by Nigel Thrift fits into this picture (Lyotard, 1984, Thrift, 2005). The latter will be exemplified in the emergence of service-based digital products and platforms and how they operate as apparatuses of information extraction and commodification in contemporary capitalism. Later, I will conceptualize big data with respect to this context it emerged it and show how social big data gains its predicates from the said context. At this point, I will argue that the data collection, although lacks scientific purpose, is not devoid of purpose, rather, the motivation behind data collection is the potential of information to be commodified, transformed into capital and it is this motivation that determines the form and the content of social big data. After demonstrating this point, I will argue that it is not so much the advancements in technology that brought about a transformation in social scientific methodologies, but rather, the newly found exchange value of information and will discuss the problems in recognizing social big data as the correct representation of the social reality. Finally, I will touch upon the determinative and legitimizing power of the principle of performativity in higher education and will counter some arguments from the contemporary literature that are made in favor of understanding the recent change in methodologies as a ’paradigm shift’.

These two chapters together form a unified analysis of the particular of case of scientific change that this thesis takes as its main problematic and shows the role of non-scientific factors in transformation of sciences and the affectability of the enterprise of normal science. In the conclusion chapter a discussion on the problem of ’scientific change’, alongside an overview of the analysis, will be held.

CHAPTER 2

LITERATURE REVIEW

In this chapter, I will provide the relevant discussions in the literature such as the position of social sciences in a digital age, computerized societies and knowing capitalism.

2.1 Social Sciences in the Digital Age

For many, the data revolution has transformed many disciplines and sciences and continue to do so. However, many also argued that the transformation process in social sciences that can turn them into a computational discipline has been much slower compared to the disciplines such as biology and physics (Lazer et al., 2009). Lazer et al. also notes that it is not the case the Computational Social Science is not being done. What is at hand is that it is not the academic social scientists that do Computational Social Science but rather companies such as Google and Yahoo and governmental agencies are the leading producers of knowledge of the social in this digital age (Lazer et al., 2009).

This situation has been identified as a crisis by Savage and Burrows and has created a lot of discussion since its publication. In their account, the main issue that gives this situation the characteristics of a *crisis* is an issue of jurisdiction (Savage & Burrows, 2007). The realization of this crisis by one of the authors, Burrows, have happened during a series of interviews with the designers in the geodemographics industry who deal with massive amounts of social data. The realization was a result of the discourse held by these designers that have included sociology-specific terminology such as ideal types, habitus, weltanschauung and so forth (Savage & Burrows, 2007). Savage states he also realized during a research that

> There is plenty of research taking place in the cultural sector, but it does not depend very much on academic intervention. Cultural institutions have

> impressive databases, mailing lists, research projects and interventions. They have a range of 'rules of thumb', models and practices, which are informed by extensive research coordinated by consultants and partners as well as 'in-house'. For the most part, the kind of academic research carried out in the name of culture is largely irrelevant. The ideas of Bourdieu and Foucault, indeed all the glorious flourishes of the cultural turn, do not – with a few exceptions – speak to the workaday needs and interests of such institutions. (Savage & Burrows, 2007, p. 887-888)

Concluding that there is a field that they have named as *commercial sociology*, Savage and Burrows have argued in their article that academic sociologists have to rethink their methods to be able to sustain their claim over their object, the social, in this age. They have especially argued that between the years 1940 and 1990 the sociologists were in possession of methodological tools that were able to successfully grant sociology ways of access to their object thereby justifying sociology's claim over the social. However, since the beginning of the 21st century, where something such as the Data Revolution occurred and the data is being collected and analyzed continuously, sociologists' claim over their object is being challenged (Savage & Burrows, 2007). As a result of this challenge, they have invited a discussion on how to respond to this challenge as a discipline. Rather than ignoring the proliferation in the ways in which data is generated and analyzed, they call for recognizing the historicity of the methods of sociology and a mixture of methods and critical reflection.

Apart from identifying this situation as a crisis, some have celebrated the rise of computational methods and championed their usage in social sciences as well. For example, Conte et al. have argued in their article "Manifesto of Computational Social Science" that the recent advancements in information and communication technologies (ICT) improve the chances of uncovering the laws that govern the social, the laws of society. They have claimed that the massive amounts of data that ICT produces opens up many possibilities of scientific approaches in social analysis. Combined with this, the increasing computational processing power makes it easier to handle the data and come up with models that are on par with the complexity of society (Conte et al., 2012). In fact

> The traditional tools of social sciences would at most scratch the surface of these issues, whereas new tools can shed light onto social behavior from totally different angles. Possibilities ranging from supercomputers to distributed computing make the execution of large-scale, heterogeneous multi-agent programs possible, programs which prove particularly apt to model the complexity of social and behavioural systems. (Conte et al., 2012, p. 327)

As can be seen from the excerpt above, in their account the best research paradigm that fits into the digital age is a computational one that depends on the generation of huge amounts of data and the massive computational power available today.

Apart from this account, Ruppert has also argued that big data presents opportunities as much as challenges. It creates a situation where social science methods are more embedded in social worlds, in contrast to the previous situation where social worlds were objects of inquiry of distanced methods. This situation, according to Ruppert, creates an opportunity for social sciences to be more reflexive about the ontological and epistemological consequences of methods (Ruppert, 2013). This and similar propositions have sparked another discussion on the issue of "the social life of methods". Mostly brought into consideration by Mike Savage and Evelyn Ruppert, the discussion revolves around the critique of the neutral understanding of methods where it has been thought that methods are neutral tools that bridge the gap between representations and the reality. Instead, Savage and others have argued that (1) methods should be understood as social entities for they are not independently constituted by solely themselves and (2) methods are not only affected by the social but the relationship is rather reciprocal, methods help constitute the social worlds. However, Savage distinguishes these arguments from constructionism by arguing that constructionist account in general and of methods depends on the separation of the world and the action whereby the former is thought to be constituted. This separation, they argue, is the exact thing that they oppose (Ruppert, 2013). The importance of this recent discussion is that methods are being questioned in the sense that they are made objects of analysis themselves. In another article, Savage relates this discussion to the proliferation of the digital methods where he has argued that digital methods are transforming the ontology of the social. What is interesting in this discussion in terms of the problematic of this thesis is that the emergence of the digital, although not alone, has sparked a discussion on methods in social sciences. Savage and others are not simply pointing out the increasing quantification in social sciences or providing a critique of it as it is commonly being done in the literature. They point out the increasing involvement of methods everyday social life (Savage, 2013). For instance

> Social networking sites, audit processes, devices to secure "transparency", algorithms for financial transactions, surveys, maps, interviews, databases and classifications can be seen as modes of instantiating social relationships and identified as modes of "making up" society. (Savage, 2013, p. 5)

Although the discussions on the social life of methods are not only concerned with the digital and has a wider domain of inquiry, I will include only the parts that particularly relevant to the discussions on the digital. Furthermore, the emergence of "lively data", a data that is not standardized, challenges, along with qualitative methods, the positivist approaches in social sciences. This line of thought is important for us here in the sense that the claims of change are not simply that of increasing quantification but rather a digital and a computational one and therefore they require different steps that must be taken to make sociological methods suitable to produce the knowledge of the social in a digital era. In Law, Ruppert and Savage’s words

> Our objective is thus to pose questions about the consequences of digital devices for social scientific ways of knowing. If digital devices mediate and are in considerable measure the stuff of social, cultural, economic and governmental lives in contemporary northern societies, then what does this mean for our methods for knowing those lives? (Law, Ruppert, & Savage, 2011, p. 24)

In a collaborative article called "Reassembling Social Science Methods: The Challenge of Digital Devices" Savage, Law and Ruppert identify the steps mentioned above in accordance with the question that they have posed (Ruppert, Law, & Savage, 2013). Before going into these steps, it appears that digital devices are a particular issue of concern for them because through those devices the material of social lives are produced, and moreover those devices also create a big part of the apparatuses that are used to know these social lives. The approach that they take, combined with the framework of social lives of methods, is aimed at bringing out the specificity of digital devices and consequently the data that they produce. They are particular in the sense that, as Ruppert also argued separately, they are not only mediators of social but they take part in ontological as well as epistemological assumptions that we take in order to know the social (Ruppert et al., 2013). Their propositions provide to the point comparisons between conventional and digital methods as well as represent the necessary points of transformation for sociology and social sciences in general to study the social in the digital age. Their propositions to "reassemble the social methods" are as follows:

1. Transactional actors. This is to say that the data that are produced by digital devices are not similar to the ones purposefully collected by a researcher. Because they are generated during transactions the focus of the inquiry is must be the relationships between actors. Such a take on the issue, they argue, allows non-individualistic and non-humanist accounts of the social as well.

2. Heterogeneity. The purpose of inclusion of heterogeneity is to point out the non-human actors that act in a transaction. The networks in the digital domain are not only composed of people, therefore, are heterogeneous.

3. Visualization. Visualization, they argue, is vital to social analysis in this era for it allows the construction of something meaningful out of information or data. The point is that, in the face of massive databases conventional strategies of statistics or other quantitative tools are not as powerful as visualization tools.

4. Continuous, rather than bundled time. The conventional tools like surveys and interviews can detect change but to do that they must be, for example, repeated. Because, now, the new data is generated continuously and it is

not always possible to identify the holder of the data that is produced, such as an individual, these methods are not satisfying anymore for to do a comparison an identifiable entity is required.

5. Whole populations. In contrast with the old methods that made use of the sampling method, new data sources require an approach that is suitable to know the social at a population level.
6. Granularity. The subject in new datasets is identified in a unique way in different datasets.
7. Expertise. In contrast with the methods like surveys and interviews, the generation of the data no more requires an expert social scientist to correctly collect the data. The data is now created as a by-product and routinely generated and collected.
8. Mobile and mobilizing. This proposition is to point out the active nature of the public in the making up of the digital. As a result of its active nature, the digital is argued to be mobile and is capable of transcending the institutional boundaries.
9. Non-coherence. The data generated is, as a result of the proliferation of digital devices, distributed. Therefore, they argue, the incoherency of the knowledge of the social is made more visible in comparison to the past (Ruppert et al., 2013).

As can be seen, the conventional methods of analysis like interviews and surveys appear to be insufficient in many points in the face of the contemporary social. The consequence of such propositions is that, considering their arguments on the "social life of methods", the epistemological and the ontological assumptions of conventional research methods fail to account for the social, therefore, it can be argued that the object of knowledge of social sciences and particularly sociology has been in a process of transformation. The propositions that they present are important with respect to the problematic of this thesis for they are clues about the object of analysis of social sciences in the digital age, therefore, allows us to make a historical comparison between paradigms.

2.2 The Challenges of Big Data

Apart from big data presenting a crisis for social sciences in terms of epistemological jurisdiction, its employment in social sciences also presents challenges and obstacles preventing a smooth transition to a computational paradigm.

Working with big data and computational tools requires interdisciplinary skills that extend from computer science to statistics and social sciences. British Academy in their 2012 report stating that the information technology has revolutionized how data is collected and analyzed and that the UK has a rich and accessible data infrastructure that creates opportunities for research that did not exist before, argued that there is a skills deficit in social sciences. They have addressed this deficit by stating that most of the students in higher education are not equipped with desirable quantitative skills to be able to satisfy the needs of the workplace whether it is business or academy (BritishAcademy, 2012).

Similarly, Raghavan, Vice President of strategic technologies in Google and former head of Yahoo labs, in an interview have underlined the mismatch between supply and demand in skills. For instance, he has stated that

> In terms of the supply of social science researchers, my biggest obstacle in growing the social sciences group at Yahoo [where Prabhakar was previously head of Yahoo Labs] and here at Google is that we can't find enough people who are trained and interested in these issues. We need people who can straddle the disciplines. At a university, straddling disciplines and creating new disciplines is a matter of decades. In industry we run in quarters and years at the most, so we can't afford to wait for that. Getting people trained was my biggest issue – it's a supply chain problem. If I could find twenty more social scientists to hire, I would. (Mann, 2012)

So, it seems that one of the obstacles that have slowed down the transition of social sciences to computational methodologies, as also identified by Lazer above, is this skills deficit (Lazer et al., 2009). The students are not trained in the methods that are required to work with emerging forms of data and databases.

David Berry presents a comparison between the role of the university in this regard. In this historical comparison, Berry argues that that the subject produced by Humboldtian university composed of culture and a certain form of rationality should give way to a different subject, the computational subject. The requirements of the data-centric age require a subject that can unify different forms of information produced by a society, that knows where to recall culture in a *just in time* fashion, equipped with skills like computer code reading, data visualization and so forth (Berry, 2011a). This subject, Berry argues, is required to process and visualize the information generated by society quickly and effectively. The production of this subject necessarily requires a transformation in the established pedagogy employed in universities (Berry, 2011a).

In its early days, computing was brought into humanities and social sciences scholarship in order to supplement the knowledge production efforts. The main rationale was one of utilizing the services, mainly their efficiency, that the machines provide. As the tools and methods of computing have been incorporated more and more into the research, Berry argues that they have become a vital part of doing research (Berry, 2011a). In fact

> …computational technology has become the very condition of possibility required in order to think about many of the questions raised in the humanities today. (Berry, 2011a, p. 2)

Berry names this transition as *digital shift* meaning that it can be an indication of the beginning of *revolutionary science* in the Kuhnian sense of the term that will eventually lead to a new *normal science,* an epistemic change. Such a shift that affects many disciplines at the same time, Berry argues, would mean that there should be a common "hard core" that contains the ontological and epistemological principles among disciplines (Berry, 2011a). Such a "hard core", the shift being a digital one, must be one that bases itself on *computation.* This is not only valid for social sciences and humanities but positive sciences as well. Disciplines like physics and biology have already been transformed as a result of the increasing usage of computational methods (Lazer et al., 2009). Indeed

> As the advantages of the computational approach to research (and teaching) become persuasive to the positive sciences, whether history, biology, literature or any other discipline, the ontological notion of the entities they study begins to be transformed. These disciplines thus become focused on the computationality of the entities in their work. (Berry, 2011b, p. 27)

This, Berry argues, is capable of creating new ontological epoch that defines the intelligibility of the age (Berry, 2011a). Somewhat in support of this argument, Mike and Savage argued that

> However, 'data scientists' working with 'Big Data' offer a rather different challenge to the traditional sociological sensibility than the other professional actors enumerated above. They offer the possibility of describing the social world in a manner hitherto impossible. (Burrows & Savage, 2014)

That can be interpreted as pointing out to the incommensurability of the computational paradigm with its precedent. This means that the shift that Berry, Kitchin and many others identify is a fundamental one that is capable of changing, among many disciplines, the object of knowledge and necessarily the methodologies, the epistemological principles that are required to produce the knowledge of it. It is in this sense that this shift was identified as a paradigm shift. It is claimed that because (1) this shift capable of forcing a transition on the grounds of concepts and theories of many disciplines and (2) shows rapid growth in adoption in many disciplines as can be seen by recent proliferation in books, conferences, papers, recent funding and accelerating interest in digital humanities this shift can be named *computational turn.*

The whys and hows of this fundamental shift are attributed to a, perhaps, greater change in society in works of David M. Berry. The society itself, he argues, is a computed one (Berry, 2011a). The reason behind such a statement is the increasing involvement of software in our entertainment systems to communication and transportation mediums, a point similar to that of Ruppert et al.'s above. This is a society, Berry continues to argue, one that is characterized by software (Berry, 2014). As a result of this definition, it also appears that code and software is the condition of possibility of living in such a society where everyday life is increasingly embedded and mediated by computational structures which in turn leads to increasing reification of everyday life in Berry's account (Berry, 2014).

Berry, in my opinion, is right to point out the increasing role of computation in how people live in this particular society. Especially when one thinks about movements such as self-quantification or lifelogging, a form of self-analytics that rests on the premise that one can know more about himself/herself through the analysis of data that is already being collected about oneself, it seems only logical to conclude that the data, in this specific sense, is seen as a specific object of knowledge by itself. However, Berry's definition and its following consequences are not enough to make sense of this phenomena. Software, code or computationality, in the framework of this thesis, are not entities by themselves that are capable of creating such fundamental changes in domains of everyday and knowledge production. Following Berry's insight

> Computation is fundamentally changing the way in which knowledge is created, used, shared and understood, and in doing so it is changing the relationship between knowledge and freedom. We are starting to see changes in the way we understand knowledge, and therefore think about it. It encourages us to ask questions about philosophy in a computational age and its relationship to the mode of production that acts as a condition of possibility for it. (Berry, 2014, p. 4)

These changes must be thought within the larger domain of influence of economy and capitalism and the particular domain of influence of changes in software-making that is brought about by the former. To understand this problem in this manner, let us take a look at conceptualizations of capitalism in this era characterized mainly by computation and information technologies.

2.3 Knowledge Production Put in Context

The framework for contextualizing knowledge production to answer the research problem at hand, we need to refer to two conceptualizations that are provided by Nigel Thrift and Jean-François Lyotard. Although published almost 30 years apart, the two books "Knowing Capitalism" and "Postmodern Condition: A Report on Knowledge" share important similarities in their discussions. For both authors, one of the most determining effects of capitalism upon knowledge production can be found in the commodification of knowledge and the change in its forms. That discussion is necessary for us to be able to understand the paradigm shift or the Computational Turn and its conditions of possibility as a historical phenomenon.

The identification of the central role of knowledge in contemporary capitalism is made by Jean-François Lyotard in his "Postmodern Condition: A report on knowledge" by bringing out the specific relationship between capital, technology, and knowledge. The working hypothesis of the book is that

> …the status of knowledge is altered as societies enter what is known as the postindustrial age and cultures enter what is known as the postmodern age. (Lyotard, 1984, p. 3)

Amid the general transformations that Lyotard refers to as transition to the postindustrial and to the postmodern ages, the main argument is that the nature of knowledge cannot stay unaltered. It changes in ways that make it operational in the general mechanism of the capitalist mode of production which requires it to be translatable into quantities of information for

> Along with the hegemony of computers comes a certain logic, and therefore a certain set of prescriptions determining which statements are accepted as "knowledge" statements. (Lyotard, 1984, p. 4)

Moreover, similar to David Berry's argument about the transformations of university and subjectivity mentioned above, Lyotard states that

> The old principle that the acquisition of knowledge is indissociable from the training (Bildung) of minds, or even of individuals, is becoming obsolete and will become ever more so. (Lyotard, 1984, p. 4)

However, these changes are not to be attributed to the proliferation of computers or digital devices, communication and information technologies alone. The crux of the argument lies in Lyotard's conceptualization of science. Science, understood as a discourse, is always in need of legitimation. In scientific discourse established rules make assessments of truth claims in terms of their validity that stems from the correspondence between the rules of the game and the ways in which truth claims are made. However, the same rules cannot be employed to legitimize themselves (Lyotard, 1984). In other words, scientific proofs cannot be proved using by the same rules used in the production of the former. Science, as Lyotard argues in his historical analysis, relies on non-scientific narratives to legitimize itself. The postmodern condition is exactly the loss of the belief in narratives, leaving science without a ground upon which it can justify itself. The function of these narratives in contemporary capitalist societies is fulfilled by the principle of performativity. It does not only legitimize knowledge production but also higher education according to Lyotard for when knowledge ceases to be an end in itself, the legitimization of higher education depends on its contribution to the system's performance by creating skills necessary. So, the principle of performativity, in a basic sense, indicates the subjugation or having a role of knowledge and knowledge production in the optimization of system's efficiency in the process of accumulation of capital. Knowledge becomes a force production that is valuable not because of its use value but because of its exchange value (Lyotard, 1984).

The role of technology stems from two sources. First of all, when doing science, the production of proof is the point of contact between science and reality. Contemporary science depends more and more on technology to make things sensible or in other words collect and analyze data and thereby produce proofs such that technology becomes one of the determinants of truth (Lyotard, 1984). In Lyotard's words,

> By reinforcing technology, one "reinforces" reality, and one's chances of being just and right increase accordingly. Reciprocally, technology is reinforced all the more effectively if one has access to scientific knowledge and decision-making authority. (Lyotard, 1984, p. 47)

Secondly, the investment in technological apparatus, in a context where knowledge exists solely in its exchange value, requires it to be a part of the accumulation of capital. Technological apparatus works on the basis of the principle of performativity, efficiency for the less the input and the more the output, the better (Lyotard, 1984, p. 44). The reason behind arises from the fact that technological apparatus is judged on the basis of its contribution to the optimization of the system and that system is a capitalist one. It is in this setting that knowledge becomes subject to the mechanism of commodification; the investment in technological apparatus requires a return. Therefore, what organizes knowledge and assesses it in terms of its truth value is a capitalist motivation, the desire for wealth and power rather than the desire for knowledge per se (Lyotard, 1984). The content and the form that knowledge takes are dependent on the requirements of capitalism in contemporary societies as I will show how this operates in the analysis of a specific form of knowledge, big data. The novelty of Lyotard's theorizing in this thesis lies in its explanatory power in the analysis of big data that makes its commodified nature evident. This is a necessary analysis for big data is an indispensable part of Computational Social Science, our case at hand, and it is the point of contact between conditions of knowledge production and science that is mostly absent in Kuhnian theory of scientific change.

The contribution of Nigel Thrift's conceptualization of contemporary capitalism, "knowing capitalism", arises from somewhere else. The article we started our discussion with, "Coming Crisis of Empirical Sociology", begins with identifying the issue within the era of knowing capitalism (Savage & Burrows, 2007). This is for the reason that knowing capitalism is used to indicate a form of capitalism that has become knowledgeable in many ways. Thrift, in introduction of the book, claims that he wants to understand capitalism as

> …a vital intensity, continually harvesting ideas, renewing people, reworking commodities and recasting surfaces for the sake of profit, of course, but also because capitalism is now in the business of harnessing unruly creative energies for its own sake. (Thrift, 2005, p. 16-17)

There are three developments that are critical to understand the particularity of capitalism in this age according to Thrift. The first one is the notion of the "cultural circuit of capitalism". The concept is used to point to a loop of feedback that is critical for the capitalism of our age in order to survive despite its contradictions by modifying and renewing itself in a continuous fashion. This is materialized in business schools, management consultancies and gurus, etc. that are based on the will to absorb as much information as possible (Thrift, 2005).

This effectively allows capitalism to maintain itself in a continuous self-tuning state.

The second development is the changing forms of commodity and commodity relations. Perhaps as a result of increasing mediation in everyday life through digital devices commodities become more interactive in the sense that both producers and the consumers of the commodity actively take part in making-up of the commodity (Thrift, 2005). Now

> Consumers are expected to make more and more extravagant investments in the act of consumption itself, through collecting, subscribing, experiencing and in general, participating in all manner of collective acts of sensemaking. (Thrift, 2005, p. 7)

The third development that Thrift points out is the increasing ability of capitalism to manage space and time be it in the sense of developments of logistical means or in the sense of proliferation of spaces of consumption that specifically designed to increase productivity and profit on the one side and necessarily consumption on the other side (Thrift, 2005).

No doubt advances and changes in information technologies have to do with a lot of the problems we concern ourselves here. However, it is unjustifiably reductive to attribute the changing forms and uses of knowledge to them alone. Especially considering that the main employer of these new technologies are almost always companies or in particular IT companies as has been identified by Savage and Burrows, it is a necessity to consider the economic sides of the question at hand to understand how it may be the case that the characteristics of contemporary capitalism contribute to the changes in the methods and methodologies of knowledge production in sciences and disciplines. By the same token Thrift states

> …one could hardly argue against the view that the rise of information technology (and especially software) is an important development which is a necessary background to much of what is going on in contemporary capitalist economies ... I believe that it should be seen as having differential effects on numerous circuits of practice, rather than as having a uniquely determining effect of its own ... information technology is both more and less important than it is often depicted to be. (Thrift, 2005, p. 5)

Again, this unique insight is highly important to not to attribute whatever changes in the domain of knowledge production and everyday life to the inherent characteristics of information technologies, software, code or algorithms themselves. To be able to understand the condition of possibility of the present and to make a history of it we need to see it in the wider context, and in this particular case I believe it is the dynamics of the contemporary capitalist economy. In this sense, the most important contribution of Thrift's theory to the problematic of this thesis is the tendency of knowing capitalism to change the commodities and commodity relations. Digital products and platforms, as I will show, are the main tools through which contemporary capitalism relies on as apparatuses of information extraction. This takes place in a feedback loop, a loop of information used in the optimization of the system as Lyotard would put it, that is materialized most clearly in digital commodities. Software and digital platforms are the best examples of continuously developed commodities by making use of the interaction between the user and the product. This transformation in commodities and commodity relations will be presented in the analysis of the case of service-based digital products and platforms. Understood as distinct business strategies, they were decisive in the emergence of big data.

2.4 Thomas Kuhn's Theory of Scientific Change

Let us finally look at Thomas Kuhn's theory of scientific revolutions. In his 1962 book titled "Structure of Scientific Revolutions", Thomas Kuhn offers a theory of scientific change that was quite different than the conventional account of scientific progress that emphasized the cumulative development of science. One of the main premises of the book is that the scientific community cannot practice its craft without some sort of shared and received beliefs that have a critical role in the preparation of students of scientific practice (Kuhn, 1970). In this section, I will elaborate on the critical notions of Kuhn's theory such as scientific paradigms, normal and revolutionary science, scientific crisis and so forth.

Scientific paradigms can be understood in two senses. First of all, scientific paradigms are achievements that are unprecedented, they are exemplars. In this sense, they can be understood as instances of puzzle-solving activity in a discipline that paves the way, in a sense, for future research. They provide a context and a model for future scientific inquiry. In its second sense, scientific paradigms can be understood as providing a disciplinary matrix, a shared set of beliefs, terminology, methods, assumptions in a given discipline. In this second sense scientific paradigms provide the limits of scientific inquiry, they provide an object of knowledge and possible valid questions that can be asked about it which are guaranteed by the paradigm itself to be answerable (Bird, 2012). Without a paradigm there can be no scientific research. For instance

> In the absence of a paradigm or some candidate for paradigm, all the facts that could possibly pertain to the development of a given science are likely to seem equally relevant. (Kuhn, 1970, p. 15)

It is in this sense that a scientific paradigm provides a disciplinary matrix. It is sort of a common set of rules that configures the scientific apparatus in a given discipline by establishing limits and methods of inquiry as well as the object of knowledge. What Kuhn calls "normal science" is built on these paradigms. They operate within the borders that are established by the scientific paradigms. Before going into the innate characteristics of normal research activity let us take a look at how a paradigm comes to be.

Kuhn states that without a paradigm research begins with a collection of almost random facts. Without some sort of a shared paradigm, researchers confronting the supposedly same phenomena describe and account for it in different ways (Kuhn, 1970, p. 17). Later a preparadigmatic school that emphasizes a specific part of the collection facts appears. Note that there can be multiple preparadigmatic schools of research. It is out of their competition that a new paradigm emerges. To be able to emerge victorious out of this competition a theory

> …must seem better than its competitors, but it need not, and in fact never does, explain all the facts with which it can be confronted. (Kuhn, 1970, p. 17-18)

When the victor of this competition starts to attract more and more groups of scientists that other preparadigmatic schools gradually fade away and the new paradigm transforms the scientific group that adheres to it into a profession or a discipline (Kuhn, 1970). In this sense, the new paradigm's replacement of the old one seems to be a matter of construction allegiances within the scientific community as much as its ability to match the facts and provide solutions to the problems that led the old one into a crisis.[4] The establishment of a paradigm guides the whole group's scientific inquiry, in fact it is the single most important criterion that declares a field science.

Now normal science, built on an example, an achievement, is the development of a promise that is provided by the paradigm. This promise is, simply put, that with the given object of analysis, methods, and assumptions, the given questions are answerable and can be accounted for to a large part. Therefore, normal science activity is an effort to increase the explanatory power of a paradigm, to increase the correspondence between facts and a paradigm's predictions about them. Kuhn uses an interesting metaphor that gives us some clues about the relationship between the world as an object of knowledge and the paradigm as well.

> ... Closely examined, whether historically or in the contemporary laboratory, that enterprise (normal science) seems an attempt to force nature into the preformed and relatively inflexible box that the paradigm supplies. (Kuhn, 1970, p. 24)

How do scientific paradigms supply boxes, problems? What kind of problems normal science deals with? These questions are examined under two main categories, paradigms determinative effect in the fact-gathering and in the theoretical activity. For factual scientific investigation a scientific paradigm's determinative effects are as follows:

---

[4] However, Kuhn also states that since in its early days a paradigm cannot show as much evidence as the old one in its problem solving ability, the decision to follow a new paradigm is one that is mostly done on faith. (Kuhn, 1970, p. 158)

1. A paradigm reveals the effective facts that are vital for accounting for the nature of a particular discipline's objects. In other words, paradigm determines the domain of empirical facts that are worthwhile to try to study, measure, and know.
2. Some facts are studied not for their own sake but in order to compare them with the paradigm's predictions about them. This kind of work is aimed at demonstrating the agreement between the paradigm and nature.
3. The final type of factual scientific work is done in order to further articulate the paradigm and resolve certain ambiguities. This kind of work is more of an exploratory effort.

According to Kuhn, the factual part of normal science problems falls under these categories. The theoretical problems of normal science, on the other hand, is aimed at using the theory to predict some factual information. This, Kuhn explains, mostly done because a particular piece of factual information can be tested. Such occurrences are rare points of contact between the paradigm and nature (Kuhn, 1970). Through such work, a new domain of application for a paradigm can be shown or its precision can be increased. The activity of normal science is classified under three problems "… determination of significant fact, matching of facts with theory, and articulation of theory-exhaust … " (Kuhn, 1970, p. 34).

Normal science does not attempt to come up with major novelties or a new sort of phenomena rather it is mostly aimed at the further articulation of the paradigm and increasing its precision. This is why Kuhn considers normal science activity as "puzzle-solving". The core of the argument lies in the fact that puzzles are particular sets of problems that test skill and dexterity that are *solvable.* A paradigm acts as a criterion for normal science activity to determine and supply such kinds of problems. So if normal science is a totality of practices that are aimed at further development of a paradigm how does scientific change occur? According to Kuhn's explanation, in the course of normal science, the expansion of the established paradigm more and more requires specialized equipment and language to penetrate into not so easily accessible phenomena and in this process, the science becomes more and more exact and inflexible so that a scientist knows exactly what to expect in an experiment. The appearance of a problem that can disrupt the established paradigm depends heavily on this solidification of the science. An anomaly can only be characterized as one with respect to an established paradigm that defines normal phenomena. Therefore, the more a science matures, the more sensitive becomes its pointer for an anomaly (Kuhn, 1970, p. 64-65).

Sometimes an anomaly in a limited domain can cause a small-scale paradigm shift as Kuhn explains with examples of discoveries of Leyden jar and X-rays. In the cases of larger scale paradigm shifts, they are often preceded by a period of crisis. Such larger scale shifts are caused by the emergence of new theories. As normal research activity goes on and anomalies accumulate, the normal science as puzzle-solving activity can be disrupted. An anomaly does not necessarily have to cause a crisis in a given science, if they are minor issues that can be set aside they can be ignored by the scientists for some time or if they are acute and urgent issues but can be accounted for through ad hoc manipulations of the existing theory, the normal problem-solving activity can go on. One of the most important indications of a science in crisis is the loosening up of the paradigm's rules. If an anomaly is no more thought as another puzzle but something that demands recognition and more research to account for, the efforts to attack and explain the anomaly requires more and more changes, ad hoc adjustments in the original paradigm, therefore, results in proliferation of different articulations of the paradigm. In such cases, the homogeneous hold of the paradigm over scientists of the discipline gets loosened up (Kuhn, 1970). This, Kuhn argues, is the beginning of all crises. However, all crises do not necessarily end up in a paradigm shift. Kuhn states that scientists are always reluctant to denounce the paradigm that has led to a crisis. Sometimes normal science can prove itself to be capable of handling the crisis evoking anomaly and the research can return to normal. Sometimes the anomaly can be attributed to the lack of tools that are necessary to account for it so the crisis can be postponed until a later time with further developed tools and technology. The last possibility is the emergence of a new paradigm or multiple ones. The crisis-induced paradigm can be declared invalid only if there is a new paradigm that can take its place for there can be no scientific research without a paradigm, science's declaring its only paradigm invalid would be its declaring itself invalid (Kuhn, 1970). As a result of this paradigm shift, the discipline itself is reconstructed and reconfigured from the fundamentals necessarily changing the methods, the toolsets and the methodologies of the discipline.

Kuhn goes as far as to say that such paradigm shifts cause changes in the world view. Paradigm shifts reconfigure the sensibility of the scientists of a discipline. In fact

> ... during revolutions scientists see new and different things when looking with familiar instruments in places they have looked before. It is rather as if the professional community had been suddenly transported to another planet where familiar objects are seen in a different light and are joined by unfamiliar ones as well. (Kuhn, 1970, p. 111)

A change in paradigm organizes the world in a different manner to the point that it is incommensurable with the conception of the previous paradigm. Kuhn argues that this change is not a matter of interpreting the same phenomena differently after a paradigm shift. The object of a science and the research problems about it, in Kuhn's account, is not given by the virtue of themselves but by the paradigm. The phenomena, the object of a science is perceived differently after a paradigm shift occurs and, therefore, the questions that can be asked about it, the methods to know its properties all get transformed. Sometimes in literature, because of this argument, Kuhn is labeled as an idealist and/or a relativist[5] and has been the target of a lot of criticism.[6]

---

[5] See Hoyningen-Huene, 1989 for a discussion of this point.

[6] He later revised this notion of incommensurability into a sort of untranslatability between paradigms (X. Chen, 1997). In this version, a rational comparison between paradigms are made possible by reducing the previous notion of incommensurability into a change in taxonomy and lexical structures.

The important piece of his last argument for this thesis is the emergence of a new object of knowledge as a result of the reconfiguration of the world by a paradigm shift. That is the first clue for us to begin to try to understand the digital shift or the Computational Turn as a paradigm shift is to locate its object of knowledge and trace its emergence.

One final consideration must be given before concluding this section and that is about Kuhn’s account of scientific progress. Two levels of progress must be mentioned here, (1) progress in periods of normal science and (2) progress through extraordinary science. In periods of normal science, as discussed, science moves the fastest for the fundamentals of the discipline is already established and the practitioners are equipped with necessary tools to answer the questions posited by a particular paradigm, making normal science a puzzle solving activity. However, Kuhn’s question is aimed at the second level of progress. The questions are simply that why scientific revolutions have to progress in a path that leads to a fixed aim, in other words, why scientific progress has to be teleological. This question arises because once Kuhn establishes how paradigms are communicated in training of new scientists, students, the paradigm shifts, revolutions seem to be invisible. In other words, the continuity is more of an attribute of the narrative built in the scientific community rather than of science and its progress. This, we must bear in mind, is more of a question than a claim. It arises as a question because within the concerns of Kuhn’s analysis scientific progress and continuity appear to be a function of the scientific enterprise, perhaps working as a mechanism of legitimization of scientific activity.

This concludes the literature review part of this thesis. In this chapter, I tried to provide the general literature on the problem at hand as well as the main conceptual framework that will be used in this thesis to construct and locate the object of analysis and the research question along with the perspectives that will be employed to provide an analysis of it. I hope I was able to give a sense of the severity of the discussion and its consequences as it is discussed in contemporary literature along with the theoretical and conceptual frameworks that will be operationalized to take this issue as an object of analysis and make sense of it in its historicity.

CHAPTER 3

ACCOUNTING FOR THE PARADIGM SHIFT

> The move to computational social science in the presence of big data involves a Kuhnian scientific paradigm shift. (Chang, Kauffman, & Kwon, 2014, p. 68)

In this section, an assessment of the so-called Computational Turn or the digital shift will be provided with the aid of the conceptual tools supplied by Thomas Kuhn. The particular purpose of this chapter is to lay out the particularities that make Computational Social Science a distinct area of inquiry. This should be understood as a two-fold process. Following Kuhn's conceptualization, firstly I will show how Computational Social Science as a scientific paradigm supplies a certain world-view and secondly how it distinguishes itself from traditional social science and posits itself as a distinct discipline or area of inquiry. Accordingly, I will first present the historical, theoretical, and methodological roots of Computational Social Science to show how they are justified and how they work together to construct a particular world that lends itself to Computational Social Science as an object of analysis. With respect to the second point, again following Kuhn's cue, I will present the points of confrontation between Computational Social Science and, unified in this confrontation, the traditional social science. These points of confrontation refer to the problems that are not solved or for which the answers are deemed unsatisfying and upon which Computational Social Science can build its promises and distinguish itself. Together these two points form an analysis that directly addresses the contemporary literature and assesses the case at hand with respect to the concepts scientific paradigm and paradigm shift to try to understand CSS as "normal science".

It is worth noting here that Computational Social Science is by no means a homogeneous discipline. Perhaps, it can be divided into two crude and very general subgroups, simulation focused CSS and data focused CSS. This division lends itself to historical analysis as well for Computational Social Science, at first, was a simulation focused discipline and was not linked to large-scale data (Törnberg & Törnberg, 2018). It is only after the emergence of "big data" that CSS started to concern itself with such large-scale data and to a certain degree its focus shifted from simulation and modeling to data analysis. It is the data focused CSS that is receiving immense amounts of attention in the literature and is the referent when the claims of the paradigm shift are made. Therefore, the arrangement of this chapter will also reflect this shift. Accordingly, the layout of the chapter is as follows. Firstly, I will briefly elaborate on theories of social complexity that forms the basis of Computational Social Science and will show how it was utilized in scientific inquiries carried out with Agent-based Modeling. Then, I will show the continuities between the simulation and modeling focused CSS and data focused CSS in terms of their fundamental assumptions about the world. Later, I will show how big data comes into this picture, its critical role on the matters of paradigm shift and how it is understood as the embodiment of the ontology and epistemology supplied by the CSS as a scientific framework. Finally, I will elaborate on the promises of CSS, what kind of problems it is argued to be better and more advantageous. Together, these all make up a "network of commitments", composed of conceptual, theoretical, instrumental and methodological commitments, that make up the field into a science and enables it to function as a "normal science" with a certain configuration of the world with its corresponding ontology, epistemology, theories and tools which make the world knowable (Kuhn, 1970).

3.1 The Roots of Computational Social Science

There are continuities between data focused CSS and simulation focused CSS which is most apparent in their conceptualization of the world which supplies them with statements on what the world is like and the valid ways to know it. It is this conceptualization that enables a field to act as a "normal science". As Kuhn argues, this kind of conceptualization is prior in the determination of sciences in comparison to shared rules in scientific research (Kuhn, 1970, p. 43). We will take a look at the history; the roots of Computational Social Science first and then will try to locate its reflections in today's Computational Social Science practices. Now let us take a look at the fundamental notions and ideas that make Computational Social Science possible.

3.1.1 Social Complexity

Perhaps the most important fundamental idea behind CSS is the notion of "social complexity". Apart from its methodological merits, it forms the basis of a certain view of the social world and acts as the condition of possibility of emergence of social phenomena as an object of analysis. The term refers to society that is viewed as a "complex adaptive system" that changes itself with respect to and reacts to changing conditions (Cioffi-Revilla, 2017b). Let me elaborate on the concept by showing different conceptualizations of the idea.

Cioffi-Revilla defines five fundamental notions to understand social complexity namely bounded rationality, emergence, near-decomposability, modularity and hierarchy (Cioffi-Revilla, 2017d, p. 206-207). In his account, these refer to certain principles and definitions of social phenomena. The notion of "bounded rationality" acts as a definition of the individual in the framework of CSS; individuals are always goal seeking and they seldom or never act on purely rational choices. This principle acts, in my opinion, as a negative principle that limits possible forms of interaction and as an assumption that gives possible forms of interaction content. In simple terms, it tells us that any interaction between individuals should be understood in its intentionality to a certain goal, but still, individuals' reasoning to achieve that is not a completely rational and calculated process. Secondly, emergence refers to the process of aggregation of micro level phenomena making up macro level ones. This notion is critical to understand the particularity of CSS. Every social phenomenon in this framework must be understood as "emergent" before everything else. What this means is that every social phenomenon is studied in order to understand the process whereby microscopic phenomena such as individual interactions make up the macroscopic, aggregate phenomena. Social complexity itself, according to Cioffi-Revilla, is an emergent phenomenon for it is the result of the aggregation of goal seeking decisions and determined by the notion of bounded rationality. Thirdly, the principle of near-decomposability refers to the structure of social systems. Social systems must be understood as being composed of smaller components or modules such that social systems are also understood as modular. Together they allow the complexity of social systems to be seen and broken down for analytical and scientific purposes. (Cioffi-Revilla, 2017d, p. 209). These are all principles and definitions that make up a social world and act as a condition of possibility of social scientific phenomena that can be studied. In very simple terms a computationally studiable social phenomena must be understood as an emergent one that is an aggregation of actions of different parts, as a modular system.

Another conceptualization of the idea can be found in Castellani and Hafferty's work where social complexity as a theory is understood as a scientific framework not aimed at explanation of the world per se but rather as its organization in an effective manner for research. Its two assumptions according to Castellani and Hafferty are that

1. "... a social system is a type of social practice." (Castellani & Hafferty, 2009, p. 44)
2. "... social practices are the building blocks of a social system." (Castellani & Hafferty, 2009, p. 45)

Social practice, in turn, is defined as

> ... any pattern of social organization that emerges out of, and allows for, the intersection of symbolic interaction and social agency. (Castellani & Hafferty, 2009, p. 38)

Social practice is comprised of a few elements, namely, interaction, social agents, communication, social knowing, and coupling. The notion of interaction refers to the interdependent actions, behaviors of social agents which are not necessarily individuals but specifically an agent which can take the form of an institution or a group. It also consists of communication because interaction does not always take place between individuals, groups but between discourses and social codes and so on. Social knowing, according to Castellani and Hafferty, is the human element of social practice, it enables the notion of social practice to align with the world, giving it a purpose with respect to, for examples, desires, concerns and wants of individuals. Finally, the characteristic "coupling" refers to the ability of social practices to connect, attach, unite with other social practices (Castellani & Hafferty, 2009). These characteristics are important for us to review because they represent the necessary attempt to reframe social in a dynamic manner that enables it to be studied with the perspective of Social Complexity Theory. The similarities between Castellani and Hafferty's conceptualization and Cioffi-Revilla's are apparent. Both of them use the notion of social complexity as a fundamental principle that is a result of the aggregation of social practices that are determined by the principles that we have elaborated above. The result of this process, the emerging social system, coupled with the assumptions presented above is one that is understood as

> …emergent, self-organizing, bounded, functionally autonomous, thematically centered, differentiated, agent-based, rule-following and complex (that is, they are comprised of a dense number of connections and interactions and often a large number of variables). They are also dynamic, evolve across time-space, and are situated within and impacted by a variety of environmental systems and forces. (Castellani & Hafferty, 2009, p. 44)

The social system must be understood like this because it is thought of being composed of micro level interactions and is complex to the point that it is nearly impossible to keep track and understand a phenomenon in its emergence. This problem is offset by the assumption that the social system is rule-following. For example, the notions of "bounded rationality" presented by Cioffi-Revilla and "social knowing" presented by Castellani and Hafferty are negative statements that act as limitations of possible actions in their form and as well as content so that social phenomena understood as emerging out of micro level interactions is not chaotic and can be made known and understandable if the rules that determine the social system are known. Agent-based modeling is a great example of this idea in practice so let us turn to it.

3.1.2 Agent-based Modeling

Agent-based modeling is the approach to model a system from bottom-up that is mostly used in simulations where the purpose is the observation of social complexity in a virtual environment (Cioffi-Revilla, 2017c). Rather than focusing on defining the system in the first place, the focus is on configuring the relations between agents in a particular manner so that the end result is the expected system. This configuration of the relations depends on the rules that govern the possible forms and contents any relation in a given system can take.

Craig Reynold's study on simulating a flock of birds is particularly illuminating. A flock of birds, which, in the first instance, seems as if the movements of the flock was determined by a purpose that is shared by all of the components of the flock. Reynolds, in 1987, was trying to simulate the movements of the flock and the approach that enabled a successful simulation of the flock was based on a bottom-up approach in contrast to a top-down one that would focus on the purpose of the totality of the flock. He established three basic rules for the agents in the simulation that are

1. Collision Avoidance: avoid collisions with nearby flockmates
2. Velocity Matching: attempt to match velocity with nearby flockmates
3. Flock Centering: attempt to stay close to nearby flockmates (Reynolds, 1987)

Working together, they resulted in a successful simulation of a flock of birds without any knowledge of the nature of the flock in its totality.

This approach to computation is what is called Agent-based Modeling (ABS) (Macy & Willer, 2002, p. 144). Computational approaches in social science and in sociology differentiate themselves from the earlier accounts through this approach. While the earlier accounts of society were understood through a hierarchical top-down system in which individuals are shaped by institutions, norms, computational approaches to society claim to start with bottom-up processes. The society, understood as a system, is the particular form of aggregation of relations between parts of it which are assumed to be, as was in the framework presented by Castellani, Hafferty and Cioffi-Revilla, limited by certain rules. As a result, it follows that they can be simulated in a computational environment, a model, which then can be analyzed to produce knowledge about real world societies. In this sense, to a large part, the object of analysis appears to be the model itself in simulation focused approaches (Squazzoni, 2010).

The consequences of the theoretical claims elaborated in the previous section are visible. The view of the social world as a complex system where every phenomenon must be understood as emergent does not allow an analysis to start from the result of the aggregation, the social fact. In a very fundamental level, this view only allows a certain type of questions to be raised about only a certain type of phenomena, questions must be that of emergence for the phenomena is emergent. If anything, that is how CSS as a scientific paradigm supplies the "inflexible box" that acts as a precondition for scientifically studiable phenomena to emerge (Kuhn, 1970, p. 24). A small set of assumptions and beliefs about the nature of the world and social phenomena limits, to a great part, scientifically studiable social phenomena and the possible questions that can be asked about them. It reveals facts that are critical to account for the nature of the discipline’s objects of analysis.

The ideas and frameworks presented in the last sections are more or less two or three decades old. However, it is only in the last decade that there have been talks of the paradigm shift in social sciences. Computational Social Science today is in many points different, but the ideas presented above forms the theoretical and methodological basis of the contemporary CSS. Let us now carry our discussion to Computational Social Science as it is understood today.

3.2 Computational Social Science Today

Computational Social Science today, as argued above, carries certain continuities as well as differences that set it apart. A few articles in the literature especially stand out among others in their attempt to declare CSS as a discipline such as "Manifesto of Computational Social Science" by Conte et al. (2012) and "Understanding the paradigm shift to computational social science in the presence of big data" by Chang et al. (2014). They are both suitable starting points to present the nature of Computational Social Science (CSS) as it is understood and referred to today as well as its goals, promises, and drawbacks.

Computational Social Science bases itself on the idea that the advancements in information and communication technologies (ICT) enables a particular and a promising form of social science. This idea arises from two points. First one is that the increased digitalization of everyday life results in floods of data that was not available before, therefore, it opens up a domain of empirical research that was not available before. Secondly, the advancements in ICT results in the increased ability in being able to store and process that data to come up with computational models that reflect the complexity of the social and therefore can help uncover the laws of society (Conte et al., 2012, p. 327).

This kind of study of the social bases itself on the premises that can be followed to the theory of social complexity or Complexity Science in general, which to a large part makes up the epistemology and ontology of Computational Social Science. This is most obvious in Computational Social Science's promise and potential for uncovering the laws of society. The laws of society are that which determine the forms of emergence of social phenomena that I have considered in their negative effect in making the world knowable in the previous section. The ontology of complexity that acts as a foundation for empirical computational social research which, simply put, is the approach, within the limits of social science, to the social by not through social facts understood as aggregations but through the emergence of the aggregate patterns (Törnberg & Törnberg, 2018). The ultimate purpose of the research is not accounting for the aggregations themselves which is claimed to be the purpose of traditional social science but rather the processes that result in aggregate patterns so that the problem revolves around the rules or the laws that make them possible. This is stated very clearly in the "Manifesto of Computational Social Science"

> The computational study of social phenomena has been focused on the *emergence* of all sorts of collective phenomena and behaviors from among individual systems in interaction… (Conte et al., 2012, p. 328)

The scientific promise of Computational Social Science depends on this potential ability that through making use of the immense floods of data and the substantial computational power the emergence of social phenomena can be observed, explained and predicted. It is important to point out that this argument was possible in pre-big data CSS through its emphasis on models that are assumed to be able to correctly represent the real world. Therefore, the computational models could legitimately be the object of analysis. After big data the same promise is actualized in a different manner which will be discussed in the next section. In any case, the possibility of such social science requires *agentification* which is "... the process of formalizing a social theory as an agent-based model." (Conte et al., 2012, p. 333). This process perhaps is the most defining characteristic of CSS. Agentification is required for CSS to be able to be applied to societies, social entities must be transformed into computational ones. This is a valid method, considering the framework that I have elaborated in the previous section, because the complexity of social systems is thought to be the result of simple actions of agents understood in the process of *adaptation.* In Conte et al.'s words "Social complexity as an emergent phenomenon is caused by successful adaptation." (Conte et al., 2012, p. 333).

These are, above all, epistemological and ontological statements. The claim that massive amounts of data and computational power can be used to study social phenomena and their emergence necessarily requires one to assume the objective validity of data as a source of knowledge.[7] Moreover, for CSS to be possible it must also be assumed that the social world and the entities in it can be translated to computational objects with all their attributes. Computational Social Science, in contrast to the point about the models being objects of analysis themselves above, is about the complex real-world societies, not about computational variables and equations. Every social entity, people, ideas, artefacts, and their relations can be modeled and encapsulated as computational objects[8].

These very general two assumptions are necessary to make up the world, the real as CSS can know it. A very clear definition of the real follows these assumptions in Chang et al.'s 2014 article.

> The real world is a *complex adaptive system* with a collection of entities and their interactions. In such a system, the entities are continuously adapting to an ever-changing environment. Within this dynamics system, a series of events that arise based on the actions and reactions of the entities occur in different contextual, spatial and temporal settings. These typically can be observed and quantified, though some aspects may be unobservable due to the limitations of the data capture methods. (Chang et al., 2014, p. 71)

The complexity of the real world is an ontological condition for its existence for CSS. As stated above the complexity of the real world is a matter of successful adaptation for which "... a set of critical functions is necessary ... (for it to) operate and endure." (Conte et al., 2012, p. 333). It is an ontological condition because real world societies "... must be complex or could not exist" (Conte et al., 2012, p. 333). Considering this problem in terms of the necessity of certain functions which are in turn understood as uncoverable laws form the basis of CSS. This, however, also appears as a narrative of the increasing complexity of the real world that forces sciences to adopt computational methods. For example, Castellani and Hafferty provide a narrative that legitimizes and posits the necessity of a shift in

[7] This will be elaborated on in the next section.

[8] This finds its correspondence in different programming paradigms such as "Object Oriented Programming" which allows computational entities to be created with attributes that can mimic their counterparts in the real world.

sciences to computational methods not only on the basis of scientific rigor but on the basis of historical inevitability. Their point is that in the last few decades Western societies have reached point that changed the organization of it along with globalization, post-industrialization and computer revolution that resulted in, in fact, the increasing complexity of real-world societies for which normal tools of science cannot account for (Castellani & Hafferty, 2009, p. 21). The world as the object of knowledge, in this case, is not reorganized according to the principles of complexity science or Computational Social Science but it itself changed.[9] Apart from its legitimizing function, this narrative reminds us of the Kuhnian point that the paradigm change results in the transformation of the world. The point here is not that the world itself has changed and was a different entity before but rather that it is simply naturalized. The real world as a complex adaptive system that could only be studied by computational methods is self-evident for Castellani and Hafferty. As Kuhn states, scientists do not see a different entity in a different manner after a revolution, "... they simply see it." (Kuhn, 1970, p. 85).

These assumptions, or commitments in Kuhn's terms, are necessary for a paradigm to be able to provide to scientists (1) "... what sort of entities the universe did and did not contain ... " and (2) "... what ultimate laws and fundamental explanations must be like ... " (Kuhn, 1970, p. 41). Combining these, the conceptual framework that defines the society in a certain manner that builds upon the notion of *complexity* thereby providing ontological and epistemological grounds to understand social phenomena as *emergent*. As a consequence, the questions that can be asked and the valid methods to answer them are determined. To put it very simply, research under such a paradigm is concerned with the detection of the forms of aggregation of interactions of elements that make up wider phenomena. Accordingly, CSS is defined as

> ... a new field of science in which new type of data, largely made available by new ICT applications, can be used to produce large-scale computational models of social phenomena. (Conte et al., 2012, p. 333)

[9] This, of course, is from the point of the scientists themselves. In Kuhnian framework what is at stake is always reorganization of the world according to the principles of the paradigm.

This definition is significant for us because it shows how the link between scientific research and the empirical is created on the ground that what is happening, since it is more and more happening on digital platforms, can be registered, stored and analyzed. Since every phenomenon can be registered, what is now unobservable potentially will be observable as more efficient and comprehensive methods of collection and storage being developed. This is a certain way of viewing the world that is crucial for the validity and legitimacy of CSS. So, what sets the contemporary CSS apart is the increasing focus on data. The critical point is how that data is conceptualized in a suitable fashion that allows it to be incorporated in the framework that I have elaborated above. Let us now turn to that discussion.

3.2.1 Big Data

Computational Social Science literature has changed since the emergence of big data which has played a big part in its popularization as well. What occurred was a shift in emphasis from agent-based modeling to data based approaches (S.-h. Chen & Venkatachalam, 2017). [10] In this section, I want to show certain continuities in these approaches, even if there was a shift in focus, following Törnberg & Törnberg (2018) in that regard, and the possible position of big data in the framework CSS supplies.

Let us start by reiterating over big data's characteristics. What makes Big Data is not so much its volume or extent. Although it must be admitted that they make up a big part of it, in our discussion and in the context of understanding the shift to computational methods in social sciences as a Kuhnian paradigm shift what makes big data distinct is the its quality that makes it unworkable with traditional approaches (Törnberg & Törnberg, 2018). It is in this sense that big data can be understood as a particular phenomenon, or a tool depending on the standpoint, that requires and brings with itself epistemological changes. These changes must be understood with respect to the framework that Complexity Science or its

[10] There are scholars that consider such an approach that revolves around big data to be incompatible with CSS's bottom-up, generative understanding of social systems. See O'Sullivan, 2018.

adaptation in social sciences and sociology supplies. As we have covered in Chapter 2, big data is already being talked about as a sort of force that is changing the organization and objects of knowledge (Kitchin, 2014a; Boyd & Crawford, 2012; Berry, 2011a).

How is big data qualitatively different from the previous datasets? Kitchin and McArdle provide several categories or characteristics that are identified about big data in the literature namely volume, velocity, variety, exhaustivity, resolution and indexicality, relationality, extensionality and scalability, veracity, value and variability which are by no means exhaust the different characteristics identified in the literature (Kitchin & McArdle, 2016). These categories refer to

1. In terms of volume, the huge quantities of information present in big data in terms of measurable size in bytes. This volume doubles approximately every two years (Kitchin, 2014b).
2. In terms of velocity, the real time production of information in big data. It is this characteristic that separates big data from small data that is gathered in at a fixed point in time. At the same time, it is because of its velocity that Chang et al. are able to claim that big data inherently aids longitudinal research (Chang et al., 2014).
3. In terms of variety, the fact that big data can be structured, semi-structured or not structured at all. This does not mean that small data are always structured but rather that big data is more likely to be unstructured for data does not consist solely of numbers but of texts, images, video, audio and so on (Kitchin, 2014b).
4. In terms of exhaustivity, the fact that big data provides information about the whole population making sampling in research mostly unnecessary (Kitchin, 2014b; Kitchin & McArdle, 2016).

5. In terms of resolution and indexicality, big data is fine-grained and indexical in nature for most of the entities in digital form are coupled with unique identifiers (Kitchin, 2014b)
6. In terms of relationality, the fact that big data can be combined with or be a combination of numerous datasets which means that they all contain common denominators that enable conjunction. According to Kitchin, this is partly the result of its indexicality because with unique identifiers same entity can be tracked in different datasets (2014).
7. In terms of extensionality and scalability, the fact that big data can be modified and can be increased in size easily and quickly. Kitchin here notes that the design for data collection can be changed at any time because vast amounts of data are continuously produced and therefore there is no risk of losing representativeness (Kitchin, 2014b).
8. In terms of veracity, the fact that big data can contain errors, uncertainties and may need working on it to make it into something usable in research.
9. In terms of value, the fact that big data can be repurposed and used under other contexts (Kitchin & McArdle, 2016).
10. In terms of variability, the fact that the meaning of the information changes with respect to the context it is produced (Kitchin & McArdle, 2016)

As the reader can see, there is no shortage of characteristics attributed to big data. What is important within the problematic of this thesis is to understand and relate the characteristics at hand with the framework of CSS. Why big data cannot be processed and worked with using traditional methods, theories and tools? One of the critical characteristics of big data making it different from the previous forms of data is the fact that it is not collected with a scientific purpose in mind. Big data is not scientific, the data is not about the answers to carefully crafted questions. There is no research question that determines a particular method of collecting data. As a result, it is not structured and it is not about a particular phenomenon per se. Therefore, it becomes impossible to work with traditional scientific tools for they were designed to be working with data that is collected under a certain framework or a theory. This, I think, did not receive the attention and elaboration it deserves in the literature. Starting from the unscientific nature of the big data many characteristics that are collected by Kitchin and McArdle can be deduced particularly, its variety, relationality, veracity, and value. Collecting data about a particular phenomenon imposes structure and purpose to the data collected. That makes it workable with the tools at hand because the data collection is done with consideration of the available tools and methods at hand. Big data, however, because it is not scientific in the traditional sense of the word, is unstructured for there is no scientific purpose that can impose a certain structure on it. By the same token, social big data requires additional effort e.g. data wrangling or data cleaning to make it usable in social science research. The fact that it is relational can also be deduced from the fact that it is not about a particular phenomenon per se. This indicates the supposed neutral nature of the domain data is collected from. It can be combined with numerous other datasets because it is untainted, so to say, by a (1) a research problem, (2) a researcher. Every dataset that emerges is from a neutral domain, therefore, creates little problem when combined. That domain posits itself as an objective source of information in all its rawness. For example, this will be further elaborated in the next section, Chang et al. states that big data diminishes the distinction between field experiments and laboratory experiments (Chang et al., 2014). A big assumption lying beneath such a statement is that the data available to the researchers, now, is about the things as they are in themselves. Because there is no question that must be asked to the subjects of the research or that there is no research problem that simultaneously creates and limits the related facts about the phenomena, the data at hand is "... directly from the real world as *digital traces* of human behavior." (Chang et al., 2014, p. 7).

This bold argument must be understood in its relation to its respective ontology and epistemology. As Törnberg also argues, big data in this context is conceived as a direct and correct representation of the workings of the social in all its rawness and naturality (Törnberg & Törnberg, 2018). It is complex, dynamic, and observable; it the claim to observe the *emergence* of social phenomena bottom-up possible because it is conceptualized as such. In fact, it is argued that it

> ... give us the chance to view society in all its complexity, through the millions of networks of person-to-person exchanges. (Manovich as cited in Törnberg & Törnberg, 2018, p. 3)

Two assumptions work hand in hand here to create a very solid but perhaps tautological stance. The data generated on digital platforms is thought to be being generated as a by-product of people's interactions and naturalized as a source of knowledge and because they are generated on digital platforms, they are structured in a particular manner that enables them to be studied through computational methods, giving them the ability to be translated to computational objects.

Another important point that distinguishes CSS from traditional quantitative social analysis lies here as well. Again, because data to be analyzed is not structured with respect to a certain research problem, the variables in data are not decided, again, with respect to a certain research problem conventional statistical methods cannot work with it. In Conte et al.'s words

> One of the most interesting aspects of social life is out-of-equilibrium. Most social distributions are not Gaussian, or bell-shaped. They are often heavy-tailed, power-law (Pareto), Weibull (exponential, Rayleigh, others), log-normal. Natural raw data reflects the typical disequilibrium of social complexity – "normalizing" data using a-theoretical transformations (for regression analysis) may destroy valuable information about generative processes. (Conte et al., 2012, p. 334)

This quotation is important for us for two reasons. First, it shows us that CSS should not be understood as an extension of quantitative approaches in social sciences. According to Savage, the lively data at hand exceeds, as Conte et al. also hints at, the capabilities of standard quantitative tools like regression and the general linear model, therefore, breaks the "straitjacket imposed by positivist statistical procedures" (Savage, 2013, p. 6-7). The second reason is that it shows us how "natural raw data" about the social does not require a research problem to be collected. It is established as a neutral phenomenal domain that only after it is collected a researcher approaches it with a certain research problem in mind. Working in this manner, the phenomenal domain of CSS is distinguished and determines the possible objects and as well as the methods of analysis. With its incorporation into discipline on the grounds of the theory of social complexity, together they effectively "... restrict the phenomenological field accessible for scientific investigation at any given time." (Kuhn, 1970, p. 60). Further, it is, for the reasons mentioned above, is what pressures scientists to question and rethink the fundamentals of the discipline. In Kuhnian terms such a force that can push scientists to question the fundamentals of discipline can be conceptualized as the *anomaly*, the unaccountable, triggering a moment of crisis, creating an occasion where retooling becomes necessary (Kuhn, 1970). Considering that the discourse around computational paradigm shift coincides with the incorporation of big data into social science, it seems that although the theoretical grounds and roots of contemporary CSS predate big data, the most effective element in this transformation was, still, the coming into being of social big data.

Scholars like Kitchin, citing Anderson, 2008, A. J. G. Hey & Hey, 2009 and many others, considered big data enabled Computational Social Science as a new form of empiricism. Such formulation is supported by four points and assumptions about big data and this new form of empiricism that are (1) big data captures the data of whole populations, (2) big data makes completely data-driven science possible without a priori theories or hypotheses, (3) through data analytics data itself can make meaningful patterns evident without human interpretation, (4) meaning is independent of the context it is generated in therefore can be decoded by anyone with required skills (Kitchin, 2014a, p. 4). Together these four points make up what Kitchin calls the new form of empiricism based on big data. However, such a stance is unacceptable for us, within the problematic of this thesis, for data-driven CSS to be a scientific paradigm and for there to be a paradigm shift in social sciences as claimed by Chang et al. (2014) and others, CSS should be able to supply the building blocks of the phenomena it deals with, that is, the social, in the form of assumptions, beliefs and/or commitments. Kitchin's interpretation of the change in social sciences falls short of such structured or semi-structured explanation of the epistemological problem at hand. Therefore, big data should be considered, as argued above, within the wider framework of CSS that builds upon the ideas from social complexity theory since the paradigm shift declared in the literature is not solely about big data but about computational methodologies. However, relying on the accounts provided by Chang et al. and Conte et al., it seems that it is the emergence of big data that supplied CSS with a phenomenological domain that popularized it in the scientific community and ultimately lead to declarations of the paradigm shift in social sciences. It must be noted that although there are continuities in terms of ontology and epistemology of CSS between model or simulation-driven CSS and data-driven CSS, big data itself changed CSS tremendously by changing the focus from theory-driven studies that take models as their object of analysis to data-driven studies where data is thought to be representing the all configurations of the phenomena, its complexity and therefore reducing the need to introduce abstract elements from traditional modeling (Pietsch, 2013). It is this form of CSS where big data is a representation of social complexity, the new science of complexity as Pietsch calls it, that is considered as the new paradigm of social sciences, therefore, requires an analysis of the history of big data.

Now, finally, let us look at the advantages that CSS provides for that discussion is a point of confrontation with the traditional social science and acts as a mechanism of self-demarcation.

3.3 Advantages of Computational Social Science

The promises of this conceptualization of the world and the related necessary epistemological and methodological moves and arguments provide, apart from setting Computational Social Science distinct from other disciplines, certain promises and advantages to CSS which are critical to understand CSS as a scientific paradigm. Note that for there to be a paradigm shift there is no need for the new paradigm to be able to solve all problems identified in the old one. Paradigm shift, in the end, is a discursive phenomenon that mostly relies on the allegiances in the scientific community. For that to happen, the new paradigm has to look more promising than the old one and should be solid enough to allow further articulation of itself (Kuhn, 1970).

In this final section, I will present advantages of CSS as a new scientific paradigm that I extracted from the relevant literature and I will discuss them under three subsections. In the subsection "Practical Advantages" I will mostly focus on the matters that allow researchers to conduct more rigorous research easier. In the subsection "Objectivity", I will present some arguments from the literature that favors CSS because of its promise on the matters of increasing objectivity in social science research. In the third subsection "The Micro-Macro Link" I will discuss one of the most prominent arguments in the literature that CSS enables social science to move beyond the dilemma of micro and macro, consequently structure and agency. In the final subsection "Unification of Sciences", I will present the view that computational methodologies bring natural and social sciences closer under a shared paradigm and therefore allow more comprehensive and interdisciplinary research.

3.3.1 Practical Advantages

One of the most prevalent arguments about and in favor of Computational Social Science is on the transformation of research methods in the presence of vast amounts of data and the Internet.

For example, Chang et al. elaborate on Runkel and McGrath's three-horned dilemma in research which refers to the mutually exclusive nature of generality, control, and realism in research. It means that they cannot all be maximized at the same time, focusing on or maximizing one necessitates giving up others. For instance, in Runkel and McGrath's formulation, focusing on realism requires one to conduct research on the field which requires the researcher to give up control and generality, focusing on control requires a researcher to employ laboratory experiments or computer simulations which simultaneously mean letting go of realism and generality and so on (McGrath, 1981, p. 183-184, Chang et al., 2014). By attributing these issues to traditional research methodologies and differentiating CSS from them, Chang et al. argue that the research designs leveraging computational methods and the available data successfully evades such problems (Chang et al., 2014, p. 70).

This is a direct consequence of epistemological and ontological assumptions elaborated in the previous section. Because data comes from real world settings without any intervention from the researcher's part, it is thought to be the correct representation of the reality in all its intricacies and complexity. The data is not understood to be produced or collected but rather it is viewed as *traces* of interactions that the researcher can directly work with. Moreover, the research methodology that revolves around the naturalization of data solves, according to Chang et al., some other concerns about research too. For instance, the ready to hand nature of the data makes allows researchers to easily collect and analyze data multiple times, gives the researchers more control over the timing of their observations, reduces costs of research because researchers do not have to fund a survey, makes researches less interfering with the subjects of research for the researcher no longer has to ask questions to get answers.

In this way, Computational Social Science becomes a viable alternative to traditional research methodologies (1) by reducing the costs of the research, (2) by making research less obtrusive for the subjects of the research, (3) by limiting the possible effects of researchers to data to minimum.

3.3.2 Objectivity

The promise of increased objectivity is another widespread argument in favor of Computational Social Science. This line of argument is mainly built on the idea that Computational Social Science makes social scientific research verifiable and reproducible in the scientific community which was lacking in traditional social science. Thus, it enables CSS to move beyond the discussions about objectivity in social sciences.

For example, in Benthall's conceptualization, "scientific algorithms" are the main tools that make this argument possible. The concept refers to "... special algorithms that perform the logical operations that correspond to an ideally rational observer." (Benthall, 2016, p. 14). There are two characteristics of scientific algorithms that make them relevant for the discussion of objectivity. The first one is that scientific algorithms are seen as an extension of human rationality in an ideal manner such that they do not contain any biases that may arise from any, especially human, partiality. The critical part of this argument is that because they are extensions of human rationality and human rationality is universalized to include all rational subjects, they are communicable in the scientific community. So, the first characteristic of scientific algorithms is that they are intersubjective. The second characteristic of them is about the form of this intersubjectivity or communicability that is the formal proof. Together scientific algorithms are conceptualized as intersubjective formal proofs that can be communicated among practitioners of the discipline and in this way operates as a force that establishes formal rigor, guarantees reproducible results and by the same token guarantees objectivity (Benthall, 2016).

By making use of that Benthall is able to argue that CSS can overcome some of the most prominent conceptual discussions present in the field of social sciences. These are namely (i) postmodern interpretivism which, according to him, argues for the fact that every phenomenon require interpretation, (ii) situated epistemology which emphasizes the contextuality of knowledge. Both of them are points that can be used to develop a critique of CSS. The operation of "scientific algorithms" is the central argument that enables CSS to resist these critiques. Scientific algorithms, being subject to the expectation of rigor and scrutiny from scientific community works as an intersubjective and verified criterion of Computational Social Science proper. Scientific algorithms work in an *algorithmic situation* that is "collectively and intersubjectively validated in the process of its construction and continues to be validated afterward through the practice of computational social science." (Benthall, 2016, p. 23). It is through this continuous collective validation of the methods used and its capability of producing reproducible research that human partiality, as in the argument from situated epistemology, can be overcome and objectivity in social sciences can be secured.

3.3.3 The Micro-Macro Link

Another direct result of CSS's construction of the world as an object of inquiry, epistemology, and ontology can be found in the argument that CSS makes the age-old conceptual problem of micro-macro link obsolete. This problem is conceived as a conceptual dilemma in research characterized by the question "Is it the macro entities that determine the micro level ones or the vice versa?".

The argument directly follows from ontological and epistemological claims of CSS. Because social phenomena are understood as arising from interactions of different parts of the system and it is thought to be capable of being modeled or being observed thanks to vast amounts of dynamic, lively data. For instance, in the case of simulation focused CSS Squazzoni claims

> Social scientists can ... study the micro mechanisms and local processes that are responsible for the macro outcome under scrutiny, as well as the diachronic impact of the latter on the former, so that the self-organized nature of social patterns can be subject to modeling, observation, replication and understanding. (Squazzoni, 2010, p. 199)

That argument is possible because social phenomena is conceived as enclosable in computational entities and the resulting models from their interaction can be studied to make the aggregative process that gives rise to macro level social entities evident.

In the case of data-focused CSS, Chang et al. state that the problem of the relationship between micro and macro in social scientific research was a result of the shortage of data about the social so far. The proliferation of data creates a new spectrum for social scientists to work on that enables tracking and identifying patterns in social interaction so that the relation of micro to meso to macro level processes can be observed. Therefore, Computational Social Science enables a simultaneous study of multidimensions and makes it possible to bridge the gap that was present in traditional social science between micro and macro levels of analysis (Chang et al., 2014).

In the end, the CSS's contribution to the solution of this problem is not that CSS can completely explain the link between micro and macro levels. But it is that CSS makes it possible for researches to be conducted in multiple levels simultaneously so that the researcher does not have to pick up a stance before the analysis about whether the object of analysis will be studied in a structural level or in a micro level.

3.3.4 Unification of Sciences

A final advantage that can be found in the related literature is about computational methodologies' ability to bring natural and social sciences closer. This point can be deduced from all others considering that natural sciences are mostly characterized by their formal nature in the works that we have covered and CSS is considered as a means to formalize the complexities of the social which could not be expressed in any other way so far (Cioffi-Revilla, 2017a). It is in this manner that social and natural sciences are brought closer for CSS enables social sciences to adopt the methods of the latter.

An example can, again, be found in Benthall's 2016 article where scientific algorithms establish the ground that allows communication between rational subjects and as well as disciplines. That process operates as the extension of rigor and objectivity that can be found in natural sciences and computational sciences to the domain of social phenomena (Benthall, 2016). A similar logic can be found in Christakis (2012). According to him, the immense amounts of data and advances in ICT allows social scientists to adopt the method of experimentation and other tools of natural sciences. This creates a domain that intersects natural and social sciences and enables convergence of disciplines or sciences (Christakis, 2012). Similarly, it has been argued that digital allows sciences to study the "physics of culture" now that there is enough data produced and computational power is high enough to handle it (Manovich as cited in Törnberg & Törnberg, 2018).

3.4 Conclusion

The purpose of this section was to address the literature directly by framing Computational Social Science as a unified paradigm. Following Kuhn's theory of scientific paradigms, the matters that needed discussion to understand the particularity of contemporary CSS were (1) the ontological and epistemological grounds of CSS as a paradigm that is composed of certain assumptions, beliefs, and commitments. I have argued that, in contrast to Big Data's empiricist reading, data-driven CSS must be understood within its historicity and it is in this historicity that big data can be a viable domain to conduct scientific work. In the framework that starts with the complexity of the social, big data is understood as a direct representation of this complexity, allowing empirical work. The fundamental propositions of this framework, understood as the conceptual ground of CSS as a paradigm, both create a domain of analysis by distinguishing it from others and enables certain forms of questions to be asked about the phenomena existing in that domain. (2) Second matter that required discussion was the advantages of CSS that make it a viable alternative to the traditional social science. These advantages, I argued, mainly follow from the basic and following intermediate propositions of the discipline. They, understood as total or partial solutions to the problems of the old paradigm, are points of confrontation with the latter that particularize the discipline and set it apart. However, there is one point that requires further analysis, that is, the emergence of big data. In Kuhnian explanation, normal science continuously encounters phenomena that it cannot explain for various reasons, however in this case, if we are to understand the shift to computational methods as a paradigm shift, the phenomena that cannot be accounted for by normal social science, big data, cannot be understood in this manner. The phenomenon was not discovered by normal science; it emerged somewhere else that stands outside the domain covered by normal social science and incorporated into social science on the basis of particular networks of commitments, beliefs, assumptions that allow big data to be a viable domain of inquiry. It requires further conceptualization of the transformation at hand to understand the operation of normal science and the nature of the paradigm shift. Therefore, the purpose of the next chapter is to contextualize the emergence of big data in the particular historical conditions that lead to its emergence which also make it a feasible field of analysis in social science.

CHAPTER 4

CONTEXTUALIZING THE PARADIGM SHIFT

In this chapter, I will provide a different account for the emergence and rise of computational methods and big data. The mainstream explanations of the issue emphasize the advancements in Information and Communication Technologies, the becoming widespread of online platforms, the increase in computational processing power and data storage ability and on the basis of these the propositions to transform social science are made as we have covered in Chapter 2. In contrast, this chapter is dedicated to bringing out the conditions of possibility of the Computational Turn in a context that goes beyond science as such. This is to be done mainly through the conceptual tools and the historical background provided by Jean-François Lyotard in his "Postmodern Condition: A Report on Knowledge". Accordingly, first I will elaborate on the characteristics of contemporary capitalism as they are discussed by Lyotard and some others. Then, I will show how changes in business strategies of software companies fit into the picture provided by Lyotard and how they were determinant in causing the emergence of a particular form of knowledge, big data, in our case social big data. These changes mainly refer to the transformations in business strategies to generate profit that arise from the changing position of knowledge. What is at stake is, in simple terms, the moving away of software companies from selling their software to make a profit to providing their software for free in exchange for information from the users. As I will show, this is the moment where knowledge becomes a moment in the circulation of capital. Contextualizing knowledge as such enables us to understand and historicize the form and the characteristics of social big data. This is critical to understand the problem of this particular scientific change and how the operation of contemporary capitalism was and still is determinant in its emergence.

4.1 Knowledge in Computerized Capitalist Societies

The novelty of Lyotard's argument and historical account of the transformation of knowledge lies, in contrast to Kuhn's approach where science is on its own in its transformations, in his inclusion of economic and historical factors in the analysis. This is the stance I intend to take in this chapter while recontextualizing the Computational Turn. In simple terms, I will not consider science as a black box denying all affectability and maintains its purity but as a part of wider economic structure i.e. capitalism in post-industrial societies.

Since Lyotard's problematization is based on the notion of legitimation, I will follow the same line of argumentation. Let me briefly remind some of Lyotard's arguments and discuss them in relation to the problem at hand. One of the basic claims of Lyotard is that science is always in need of legitimation which is a critical as well as a suitable point for our discussion (Lyotard, 1984). The justification behind such statement stems from the fact that scientific claims get their legitimacy from their compliance with rules of, so to say, the game. What this means is that thinking this together with Kuhn's ideas, the rules of the game are provided by the dominant scientific paradigm in any given field of inquiry. The same mechanism that grants legitimacy to scientific claims falls short of granting legitimacy to science or paradigm itself. The reason behind this is that since the rules of the game, being thought as a certain unity of ontology, epistemology, methodology, methods and so forth, cannot readdress themselves and therefore science cannot prove the legitimacy of its basic assumptions like it can with claims made within it. In Kuhnian terms this operates as the conditions of truth within an established paradigm supplied beforehand to scientists, making them able to conduct normal science as puzzle-solving activity (Kuhn, 1970). However, the same criteria cannot be applied to the paradigm itself. For Lyotard science needs to appeal to other mechanisms of legitimation. In Lyotard's words

> Scientific knowledge cannot know and make known that it is the true knowledge without resorting to the other, narrative, kind of knowledge, which from its point of view is no knowledge at all. (Lyotard, 1984, p. 29)

A similar point can be found in Kuhn's theory, although by no means the same because in Lyotard's account legitimation seems to be a structural matter while in Kuhn's it is closer to a personal matter. Kuhnian explanation emphasizes, in particular cases of paradigm shift, personal belief or faith as the mechanism that grants legitimacy to the new paradigm. It would not be implausible to think of this as applying to the problem put forward by Lyotard as well. Because paradigms are discursive phenomena and rely on allegiances in the scientific community to become dominant, it can be argued that the legitimating function is provided by the same network of commitments as well. Again, along these lines, Lyotard claims

> The conditions of truth ... the rules of the game of science ... are immanent in that game ... there is no other proof that the rules are good than the consensus extended to them by the experts. (Lyotard, 1984, p. 29)

The particularity of the position of knowledge in contemporary capitalist, postindustrial societies arises from the fact that, according to Lyotard, science no longer has any metanarratives that it resorted before to legitimize itself. This loss pushes Lyotard to search for the mechanism of legitimation of science in capitalist societies that establishes the conditions of truth. What is at stake in capitalist societies that can legitimize research is, Lyotard argues, the principle of performativity. The principle of performativity passes judgments on the basis of the questions of economic value and efficiency (Gane, 2003). This is the case because research activity relies heavily on the production of proof, especially in normal research if we combine this idea with Kuhn's where the production of proof coincides with the further articulation of scientific paradigms. The proof is where the link between research activity, knowing is linked with its referent, that is the "reality" (Lyotard, 1984, p. 44). Performativity principle finds its place in this production of proofs for scientific observation increasingly relies on technological apparatus to gather data and gain access to things as objects of knowledge. What is expected from the technological apparatus is to expend as little energy or input as possible to produce as much output as possible. It is important to keep in mind that this principle of performativity is not something intrinsic to technology, it should not be attributed to it. Rather, the link can be found in the relationship between technological apparatus and the investment that is necessary to actualize, develop and maintain it (Lyotard, 1984). This is perhaps the most critical argument in Lyotard's analysis for it is this point that establishes the link between capitalism and nature of knowledge and is the basis to understand the emergence of big data as a condition of possibility of computational social science today in its contextuality. It is the motivation of profit behind the investment in technological apparatus that results in the performativity principle which essentially refers to the maximization of surplus-value. The investment in technology, according to Lyotard, means reinforcing reality for it enables more efficient production of proof, the point of contact between reality and research, and enables the owner of the apparatus to increase their chances of being right, effectively functioning as the condition of truth in capitalist societies (Lyotard, 1984, p. 47). As the most determinant factor in knowledge production stems from the logic of capital, the ultimate goal of this production is always exchange. Not only that knowledge becomes a commodity under these circumstances, because it enters into the process of circulation of capital, but also it becomes a force in structure of commodity production. It becomes "... the principle force of production ... " in contemporary capitalist societies (Lyotard, 1984, p. 5).

In a similar vein, Thrift's conceptualization of "knowing capitalism" is also important to understand the role of knowledge takes in contemporary circumstances. Thrift understands capitalism as performative, as subject to change in a continuous manner. The position of knowledge is determined by this characteristic of capitalism that perpetually transforms itself and adapts by centralizing the role of "knowing" (Thrift, 2005) and one of chief tools of "knowing" is materialized in digital platforms and commodities. The reason that Savage and Burrows situate the empirical crisis of sociology in the age of knowing capitalism is this effective position of knowledge in capital accumulation which results in private sector having and investing in sophisticated research infrastructure which results in the shifting of locus of knowledge production from academe to business (Stanley, 2008). Along with Thrift's conceptualization capitalism, others like data-driven capitalism, platform capitalism, informational capitalism are all attempts at conceptualizing a particular form of capitalist economy where knowledge has a central role in its operation (Sadowski, 2019, Fuchs, 2010, Srnicek, 2017). Knowledge, however, can only exist in a certain form that allows efficient accumulation of capital which emerges as computationally processable, easily transmittable quantities of information, data for it is determined by the operations of capital.

The context that we have is built on these few, well-known propositions about the nature of contemporary capitalism where information has become a key resource. It is in this context that we will reconsider the emergence of social big data. Recall that the "crisis" identified by Savage and Burrows was not a scientific one in the Kuhnian sense of the concept but still carried resemblances in its effects for it, for Savage and Burrows and for many other accounts we have covered in the literature review section, requires social sciences, particularly sociology, to undergo changes in its fundamentals, to search for new building blocks that grant, once again, legitimacy to sociology in terms of its ability to produce the knowledge of the social (Lyotard, 1984, Kuhn, 1970). The crisis is one of domain and jurisdiction where private companies have been gaining more authority and competence over the object of knowledge of sociology. The nature of this crisis, when thought within the context provided by Lyotard, makes sense for what drives research, in the case of private companies, is not so much the desire for knowledge but the desire for increased power and profit. The relationship between science and technology becomes reversed because it's the efficiency provided by the principle of performativity that contributes to the optimization of the system and the increase in profits not knowledge itself as such (Lyotard, 1984). Our task, then, is to inquire about this is to uncover whether such transformation of logic can be identified in the business practices of private companies and in the nature of big data.

4.2 Software Development Put in Context

In this section, I will present a particular shift in the business strategies of software companies which is a concrete case that illustrates the pivotal role of knowledge in capital accumulation and is directly related to the emergence of social big data. This shift will be located in the process of transformation of commodity and commodity relations as identified by Thrift (Thrift, 2005). The analysis of this shift serves two purposes here: (1) it makes the relationship between technology, capital, and knowledge concrete and (2) shows how economic factors and decisions of businesses were decisive conditions for big data to emerge. The latter is highly important to address the affectability of normal science, particularly social science in our case, which is rather absent in Kuhn's conceptualization of science and scientific change. Let me first explain what the mentioned shift consists of.

4.2.1 Profiting from Digital Commodities

We will continue with two distinct mechanisms of generating profit from software or from digital commodities and services in this section. The first one basically consists of selling licenses to users along with the copy of the product while the latter is built upon providing services is thought to be a part of cloud computing ecosystem which includes Software as a Service (SaaS), Platform as a Service (PaaS), Infrastructure as a Service (IaaS) and so forth. I will consider the latter as an example in a more encompassing mentality that provides products as services on demand. These strategies are new ways of generating profit that only make sense in a capitalist economy where knowledge becomes a force of production. I will first elaborate on what these are, then will establish their relation to our discussion. Let us start by discussing the former.

Until recently the strategy for generating profit from software was similar to any other commodity meaning that a software company produces a commodity that satisfies a certain need or want and sells the copies of it along with rights to use it to the consumers. The main mechanism of generating revenue here is this exchange between the producers and the consumers of the software. The implications of this business strategy are that the computation mostly happens in the devices owned by the consumer of the software and in most cases, the data generated within the interaction between the device, the software, and the consumer stays in the consumer's device the software is being run on. This is essentially a method of licensing the product which now gives its place to another type of licensing and a new business strategy.

SaaS and other service-based products and platforms in general have been becoming more and more prevalent in the last decade. Information technology and software companies, especially the most popular ones, increasingly prefer to generate revenue using the second approach.

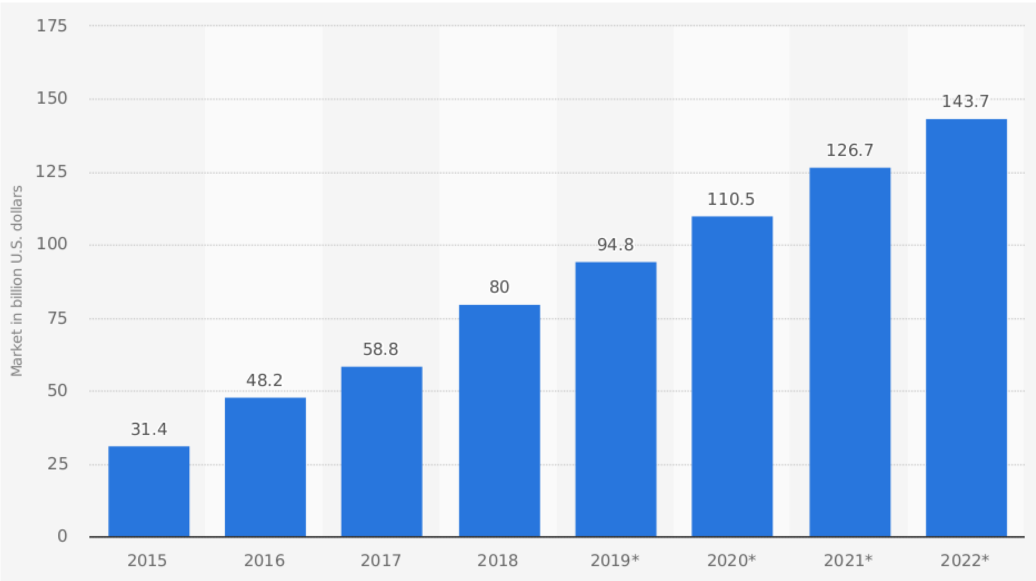

Figure 4.1: SaaS market revenues worldwide from 2015 to 2022 (Gartner, n.d.)

As can be seen in the figure 4.1 above, the market revenues for SaaS products have been showing a steady increase in the last 4 years and the same trend is expected to continue until 2022. Moreover, the most popular companies with the biggest share in the market continue to increase their investments. Top 10 companies in the SaaS market account for more than half of the total market share as you can see in figure 4.2 below.

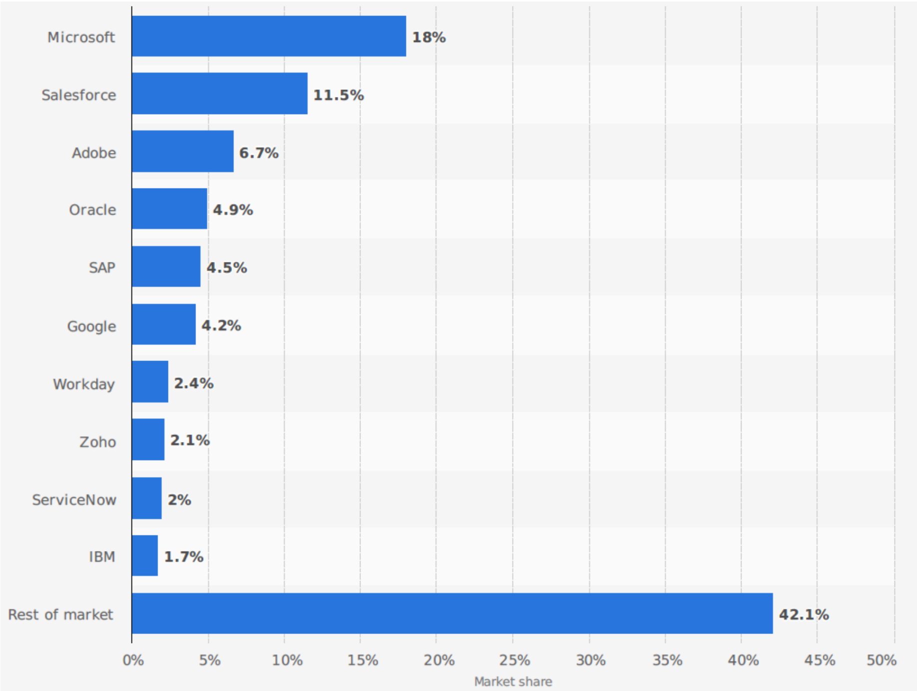

Figure 4.2: Market share of public cloud SaaS vendors worldwide in 2017 (Website(fourquadrant.com), n.d.)

The statistics above are mainly are that of companies which provide environments for others to host their software or provide their own on-demand service-based applications. What is meant by "cloud computing" is essentially this idea of using the infrastructure provided by a third party to provide services to end users. Such platforms that provide infrastructure to service-based applications are basically the infrastructure of contemporary capitalism where information extraction, storage, and processing is a key aspect of the business. Not only these companies that provide infrastructure generate revenue from the rental hardware but also these enable them to collect information for their own purposes (Srnicek, 2017).

In our analysis, we will also consider online platforms such as Facebook, Twitter and so on in this category as well for what they provide are services in a platform that are hosted by the companies themselves. The importance of this increase in service-based platforms and software is their particular profit generating mechanism. There are basically two ways to generate profit from a SaaS platform or product, either the company charges a subscription fee periodically for its customers to access the software or the platform or the software is distributed free of charge and the profit is made through data mining and its use in advertising. This business strategy has been mainly successful as we can see in the figures above and moreover, The Economist states that SaaS market is growing 50 percent every year (The Economist as cited in Schütz, Kude, & Popp, 2013).

Most popular companies like Amazon, Facebook, Twitter, Google, Microsoft, etc. all provide free access to their services but still are able to generate large amounts of profit. Companies like Facebook, Google, Twitter provide platforms as service to users in exchange for user information which proved itself to be highly profitable. For instance, Facebook had a profit rate of 51.2% in 2011 even though the times were of a global economic crisis (Fuchs, 2015). Another example is Google which was able to generate 8.5 billion US dollars in profit in 2010 (Fuchs, 2012). Other software companies such as Microsoft follow the same strategy. Microsoft recently published Windows 10 for free for a period instead of selling licenses to users and generating revenue from them. Now, the company has moved its strategy to charge users for a periodical fee as well as collect telemetry data from the devices and displaying ads in parts of the operating system.

It cannot be denied that the technological advancements in the industry contributed to this service paradigm to become a reliable infrastructure for businesses. SaaS applications essentially require high-speed access to a server that is capable of computing on demand by the customer base, capable of holding the data of the customers and stable enough to provide a reliable service. Apart from its profitability, there are some technical and business advantages to provide a platform or software as a service for the producer of the software. For example, because the software runs on a server controlled by the producer, the producer no longer has to support different devices, architectures, operating systems. Any device that is capable of connecting to the World Wide Web and running a web browser can access the software. Also, the producer can modify the software on-the-fly and does not have to publish different releases periodically for the software can be continuously updated on demand easily (Kaplan, 2007). Ultimately this increases the productivity of the company and allows for different distributions of available labor power. That is because a company that releases a piece of software does not have to maintain the necessary infrastructure and only has to maintain one central copy of the product. Based on these, SaaS is claimed to be allowing companies to focus on their core business rather than spending time and money on the infrastructure (Godse & Mulik, 2009). On the other side, these can be considered advantages for customers as well for the customer does not have to maintain her own copy of the software on her personal device. Moreover, it is generally cheaper for the customer because the heavy computing is done on machines owned by the company and, therefore, the customer can access the software on relatively cheaper low-end machines as well.

There is more to SaaS platforms and the like that enables them to be a core component of knowledge-based capitalism. So far, we have defined SaaS as pieces of software that run on the servers owned by the producer company itself and generates revenue by either charging a subscription fee or by data mining to display ads or using both of these strategies. The critical point is that the nature of SaaS is intrinsically enabling for practices such as data mining because all data is generated and stored not in the consumer’s device but in the companies’ servers. This is considered as a potential disadvantage in the literature for the consumer of the end product and the producer company in terms of data security. Because if the company that provides the service does not own the infrastructure, the producer of the product has to put their trust in the technical competency of the third company that provides the infrastructure and has to have faith in the latter’s good intentions (Godse & Mulik, 2009).

This type of computing makes it easier to collect and store large amounts of data because (1) it shifts location of computation, (2) it allows the companies to rent cutting edge infrastructure rather than investing on it. This shift, understood as a business decision or strategy, is, to a large extent, responsible for datasets that are categorized under the notion of "big data". And this shift must be considered in the context provided by Lyotard and Thrift to go beyond the common explanations in the literature that mostly attributes the emergence of big data to the proliferation of digital devices and advancements in information and communication technologies. These explanations are not enough for the sole reason that the advancements in these sectors could not have happened if they have not found their place in the circulation of capital. Therefore, these advancements as well must be put in the specific context of contemporary capitalism that is characterized mostly by its "computerized" nature (Lyotard, 1984) and the central location of "knowing" in it (Thrift, 2005).

4.2.2 SaaS Put in Context

To go beyond technological explanations of this shift in SaaS, we will consider a few implications of Lyotard's claims. It must be noted that the necessity to go beyond technological explanations is not important in itself but in relation to our core problematic in this thesis. Showing that there is more to the transformations at hand will also reflect on our analysis of the specific scientific change we have gone over in the previous section. Otherwise, the argument would be limited to the mainstream understanding of scientific change with a particular emphasis on the determinative power of technology. The task for us here, then, is to build on the relationship between technology and capital, and, ultimately, knowledge that is supplied in Lyotard's as well as Thrift's conceptualizations of capitalism.

The peculiarity of contemporary capitalism arises from its distinct mechanisms of capital accumulation where (1) information is both a resource and does act as a currency of exchange, (2) the commodities are, to a large extent, produced jointly by producers and consumers. In this sense, contemporary capitalism ultimately is a specific configuration of conditions that enable these two points to operate.

In contemporary capitalist societies, knowledge takes the form of value for it is submitted to the specific relationship of exchange. The particular business model that profits off of information produced by the users of the product follows this logic of commodification (Cohen, 2008). The process whereby information becomes valorized in its exchange value is mostly its use in personalized advertisements that, above all, requires constant surveillance of users. As discussed above, SaaS indicates a shift in the location of computation and therefore in the location of data generation. This supplies the providing platforms with the control of the digital space where users interact with each other, with the software, with other digital entities to the owning company. Such levels of control make these platforms perfect apparatuses in data extraction (Srnicek, 2017). The payoff of the investments in technological infrastructure stems from the location that knowledge holds in contemporary capitalism that determines its exchange value. The concrete mechanism that makes information valuable and makes this business model profitable comes mainly from two mechanisms that correspond to the two points mentioned above.

The resource and currency like nature of data come from its employment in online advertising. The data extraction platforms generate revenue by collecting and processing the information which is produced in the space that is owned by these platforms. The user is tracked not only on specific platforms but constantly as they surf in the Web which makes the collected information more extensive and more valuable. The collected and processed information is a raw resource for advertising companies to be refined and repurposed for a variety of reasons but ultimately to generate revenue (Srnicek, 2017). This value takes part in the exchange both between businesses, the platform that collects data and the company that buys it for whatever reason it may have and as well as in the exchange between the platform and its users. The services are provided for free in monetary sense like in the cases of Twitter, Facebook, Google and so forth however the users pay for them in the currency of information, data about their consumption preferences, location information, demographic information and so forth.

The second point that the commodities are produced both by the producers and the consumers is another factor that makes this business model more efficient and profitable and is a key characteristic of contemporary capitalism according to Thrift (Thrift, 2005). Some part of the productivity of the system stems from the fact that these services and platforms are made what they are by the users themselves. As Fuchs explains, the secret behind the profitability of this business model is the exploitation of unpaid user activity (Fuchs, 2015). The content in the platforms is generated by the users who share opinions, media, links, etc. with other users on the platforms which in turn generate traffic to the website, making it a suitable platform for advertising. This is a core feature of Web 2.0 where software and computing become a "platform" that involves no economic cost for the users but expects users to contribute to these platforms, making them more useful, attractive, and profitable (Hardey & Burrows, 2008). This information is organized around the sole motivation of profit-making data operate as a form of capital in a cycle that is self-reinforcing. The key commodity that holds value and generates revenue is the data itself and its generation and valorization both depend on the efficiency and capability of the platform to store and analyze it as well as on the generation of content by the users. It is in this sense that both consumers and producers take part in the production of commodities. Both as a form of commodity and a form of capital data collection forms its own cycle. The companies are not interested in direct use-value of information but in its potential exchange value. For example, Andrew Ng, a researcher who worked at Google, Coursera and Baidu, states that

> At large companies, sometimes we launch products not for the revenue, but for the data. We actually do that quite often ... and we monetize the data through a different product. (Stanford Graduate School of Business as cited in Sadowski, 2019, p. 5)

There is another sense in which the data collection is self-reinforcing. Data extraction platforms depend on gathering information and processing it to optimize the performance of the apparatus which in turn enables these platforms to generate, extract and process more information (Sadowski and Pasquale as cited in Sadowski, 2019). It is in this relationship knowledge is positioned in contemporary capitalist societies where capital accumulation depends on the performativity of the technical apparatus. In Lyotard’s words

> A technical apparatus requires an investment; but since it optimizes the efficiency of the task to which it is applied, it also optimizes the surplus-value derived from this improved performance. (Lyotard, 1984, p. 45)

Because the optimization of the system, the relationship between input and output, is directly related to the production of surplus-value, it becomes imperative to construct a feedback loop that is continually used in improving the performance of the technical apparatus. Lyotard continues

> All that is needed for the surplus-value to be realized, in other words, for the product of the task to be sold. (Lyotard, 1984, p. 45)

As such, the relationship between knowledge, profit, and technology becomes materialized. The imperative to collect more information to optimize the system operates on the basis of the desire for profit. SaaS and similar other business strategies are perhaps the perfect materializations of the logic of knowing capitalism. As discussed above, SaaS and similar service-based platforms, because they shift the location of computation, are technically enabling for this kind of feedback loop to emerge and become self-reinforcing. However, the promise of profit precedes the technological advancements for the simple reason that technological infrastructure requires an investment which necessitates additional value to be gained from it. It was not until recently that companies started to keep the user-generated data; in the past investing in the infrastructure to store and process the data was not deemed profitable and companies simply got rid of the data (Oracle and MIT Technology Review Custom as cited in Sadowski, 2019). That is why emergence of the big data should not be solely attributed to technological advancements or proliferation of digital devices and so forth but capitalism's need to expand the available markets and sources of value. Srnicek locates this in the particular example of business strategies of Google. After the burst of the dot-com bubble[11] around 1999, Google needed to find a way to generate revenue and that was to create a system of commodification of user data which was previously only used for improving the search mechanism. In the latter

[11] Dot-com bubble refers to an era of increased speculation and hype around Internet based companies which caused extreme growth in the respective time period. Eventually, at the peak of their values, the company shares owned by venture capital were sold which caused the "burst" or the crash of the dot-com bubble.

use, there was no surplus value left for Google to generate value from data. By creating advertising spaces and increased surveillance of users, Google was able to successfully build an independent mechanism of capital accumulation. In fact, in the first quarter of 2016 89% of Google's profits was from advertisement (Srnicek, 2017).

Now that we have established the context of its emergence, we are able to situate social big data as a particular phenomenon that is enabled by the logic of contemporary capitalism. Let us now discuss big data once again to trace the implications of this context in which it emerged.

4.3 Understanding Big Data in Context

Now that we have covered the terrain in which big data can be contextualized and understood properly in its relation to the economic factors that contribute to its emergence, we can inquire in detail about the features of big data to, again, see them in relation to wider structure we have established with the help of Lyotard and others. The purpose of this section is to show (1) how big data emerges out of an unscientific domain, (2) the effects of this domain that it emerges out of on the nature of big data. Ultimately, these two points together form the basis of our interpretation of the alleged crisis in empirical sociology and paradigm shift in social sciences as well as the reasons behind them.

The most defining feature of contemporary capitalism is the emergence of information as a form of currency used in the exchange between producers and consumers as well as businesses. Big data must be understood in its exchange value for it is now evident that the origin of big data is the desire for wealth, not for knowledge per se. Emerging in a capitalist setting and out of practices of capital accumulation as exemplified in service-based business strategies affects the nature of big data. In a sentence, the most important precondition for big data's emergence was an economic setting that enables data to be a form of capital (Sadowski, 2019). As such, we will go over the characteristics of big data we have covered in the last chapter once again to show these effects.

When discussing the nature of big data in the previous chapter, it is argued that the chief feature that set it apart stems from the fact that social big data was collected without a scientific purpose in mind. That is still the core feature for big data to be a force, acting as an anomaly in Kuhnian framework that pushes social sciences to reconsider their fundamentals. However, now that we have established the context big data emerges out of, it is possible to carry on our analysis in a deeper level to get a sense of this alien nature of big data that makes it unaccountable for by traditional methodologies. Let us now go over some of the features of big data that were identified in the literature to see how they connect with the context we have established.

Following the point that information acts as a currency under knowing capitalism, it follows that information must be valuable for various actors that may not always share the same particular purposes. As it was identified by Kitchin and McArdle one of the key features of big data is its *value*, meaning that it can be repurposed in different settings (Kitchin & McArdle, 2016). This point is directly related to big data's capability to become a part in the circulation of capital. For big data to be used in exchange it must be valuable in various contexts meaning that it should be free from the baggage of particular research questions. The data collection, considered in this manner, can only have one purpose, that is, profit. Profit as a shared motivation for capitalist actors determines the content of big data so that it can be used for a multitude of purposes in contrast to scientifically collected data that is tailored to answer a specific research question. This meta-purpose of profit allows data to be taken out of specific context it is generated in that grants a piece of information a particular meaning. This sort of collection and analysis that is aimed at generating profit do not have a purpose to understand for the logic seems to be "Who cares why consumers are choosing what they choose? They just are, so let us make money." (Wyly, 2014, p. 680-681). This kind of logic may work in business for "... it is possible to know how to produce effects without knowing how those effects are produced." (Nightingale as quoted in Wyly, 2014) but its compatibility with social science's ideal purpose, "to understand", is doubtful. As we go over the remaining features of social big data, we shall see how this specific feature of "value" that stems from the meta-purpose, profit, functions as a ground for the rest of the characteristics.

The next feature of big data is that it is *fine-grained in resolution* and *indexical in identification* (Kitchin & McArdle, 2016). In the business model that we have covered in the last section, the source of the surplus value is the commodification of information provided by users that are later used to track the users in the Web and generate advertisements accordingly (Fuchs, 2015). This requires not only the identification of individual users but also tracking of other entities to be able to record and monitor all of the transactions taking place. The users are identified with digital cookies, media in digital forms with DOI numbers. The personal tailoring of advertisements is critical for this business model and as such requires the collected information to be fine-grained and indexical. In contrast to the traditional scientific collection, the point is not to generalize out of a sample but to get access to information about every unique component. In other words, personalized advertisement requires the identification of every single agent. This brings us to the next features of big data, its *volume* and *exhaustivity* (Kitchin, 2014b). The surveillance of every unique agent in the Web and the business model that depends on their identification logically requires the collected data to grow in size, the more information about more agents and ideally about the whole population results in huge volumes. Such practice is impractical in traditional social science research for the costs of collecting and processing information of that size is simply too high. For data collection of this extent to be possible, it was a necessity for information to have its place in accumulation of capital so that the required investment in infrastructure would be made. It is this relationship between capital, knowledge, and technology that Lyotard points out that made possible the emergence of social big data.

Furthermore, the features *relationality* and *scalability* referring to resulting data's ability to enable conjunction of different datasets and to grow in size rapidly respectively also follow the reasoning above. They are implications of the meta-purpose of profit or of the location of data in contemporary capitalism. The commercial nature of data collection determines what kind of information is relevant. In scientific research, that role is fulfilled by the research question that decides on what fields or variables are relevant for the research. In this case, they are determined by the collected information's contribution in capital accumulation, formulated in the questions "... what use is it? ... Is it saleable?" (Lyotard, 1984, p. 51). It is in this sense the motivation of profit functions as a ground for the rest of the features of social big data. Since the questions are the same for all actors collecting information about users on the World Wide Web, the resulting datasets, logically, contain similar fields which in turn allows these datasets to be conjoined. Also, since the personal identifiers are critical for advertising method of generating profit, it is safe to assume that the user, if nothing else, is the common denominator for these datasets. Their scalability, the ability to grow in size rapidly, follows from here. Datasets collected by different actors that are driven by the same economic motive results in datasets that have common denominators and share similar structures which enables them to be conjoined and ultimately in their scalable nature.

It is constructive to consider the scalable nature of big data in relation to its *velocity* and to the nature of commodity and commodity production. The commodities and necessarily the accumulation of capital rely heavily on the feedback loop the capitalism operates with (Thrift, 2005). It is this feedback loop in which

> … consumers and producers now increasingly interact jointly to produce commodities, and, increasingly, commodities become objects that are being continuously developed … (Thrift, 2005, p. 7)

This claim also finds its counterpart in the statements of companies that legitimize data collection on the Web. The personalization of online experience offered to the users justifies data collection under the same notion of "feedback" working in Thrift's conceptualization of "experience economy" (Thrift, 2005). The continuous development of the software or the platform necessitates continuous, real-time data collection as we have discussed in the previous section. Google, Facebook and other social platforms constantly change the design of their services and capture the data about the reaction of their users which in turn is used to modify the design for a variety of purposes such as encouraging certain actions like increasing the rate of clicks for a certain link or advertisement (Kitchin, 2014b). Tailoring the Web and the advertisements to realize certain ends depends on this persistent stream of data to be analyzed perpetually both to assess the reactions of the subjects of the advertising and to increase the efficiency of data extraction and analysis system. The velocity of social big data can be read as a result of this feedback loop in this context and scalability of it an outcome of the latter.

One more point is necessary before concluding this section. We have already established the relationship between capital and technology on the basis of the principle of performativity and efficiency. The form social big data takes is not unaffected by its location in capital accumulation. As Lyotard reminds us, computerization or technology finds its place in contemporary capitalism with respect to the principle of performativity. Technology as a means for capitalist ends has a role in structuring the production of knowledge in the sense that the digitalization of knowledge goes hand in hand with technology's capacity to increase efficiency. Since the criterion of truth works on the basis of value which is proportional to the efficiency of the system in which it is produced, knowledge takes a particular form, it becomes digitalized, transformed to quantities of information that can be efficiently sent, received, processed and exchanged (Gane, 2003). As the relationship between knowledge and technology becomes reversed, it is the technological apparatus that determines the form of knowledge. It is worth noting once again, technology or efficiency are not ends in themselves but always find their places in accumulation of capital.

To a considerable extent, the characteristics of big data can be deduced, as I have tried to carry out, from its role in the accumulation of capital. It is a particular form of knowledge that exists, perhaps solely, in its exchange value. Its production costs are minimized for the consumers themselves produce it.

Theoretically what this means is the infinite exploitation of users because the labor that creates value is essentially unpaid (Fuchs, 2015). The surplus value emerges from this increase in performance and efficiency of the system, creating maximum output with minimum input. It is within this logic that social big data emerges, gains its predicates and should be understood.

4.4 Computational Social Science as a Scientific Paradigm

So far, we have discussed the theoretical framework of CSS in a timeline where the most effective factor in its rise and development as a distinct area of inquiry was the emergence of big data. That is why locating the origin and the features of big data were necessary. Its importance will become even more apparent when we locate it in the Kuhnian theory of scientific change.

What causes a scientific revolution in the picture Kuhn presents to us is a combination of factors but nevertheless self-contained ones, effective only in the scientific domain. Accumulation of anomalies that force the scientific community to make more and more ad hoc modifications to the paradigm that loosens the latter over time. However, no paradigm shift can occur if there is only one paradigm without a rival (Kuhn, 1970). But again, there is no need for a paradigm shift if the established paradigm can more or less successfully account for the entities in the phenomenal world it opens up. However, there is something unaccounted for in this explanation that forced us to bring forth the context of scientific change and lay emphasis on it. The determinative power of a scientific paradigm comes from a combination of its ontology, epistemology, fundamental assumptions as well as the technical apparatus it dictates. This determinative power not only organizes the phenomena in a certain manner but also creates a set of expectations which when not met, allows us to refer to some phenomenon as an anomaly.

> ... the decision to employ a particular piece of apparatus and to use it in a particular way carries an assumption that only certain sorts of circumstances will arise. There are instrumental as well as theoretical expectations, and they have often played a decisive role in scientific development. (Kuhn, 1970, p. 59)

The problem is this. It appears that scientific paradigms' determinative power is limited for if it were not so, no anomaly could ever arise. In Kuhnian framework, the emergence of anomalies and the unaccountable is limited to the domain of science and to a large extent only becomes uncovered in the process of a scientific investigation. In our case, the situation is radically different and this is where Kuhnian theory loses its explanatory power.

The unaccountable in our case that demands a different paradigm is big data. Considering its origins, it is not a result of the failure of the tools of established paradigm to account for the problems that it defines. The occasion that demands retooling is brought about by the specific relationship between capitalism, knowledge, and technology in our time. The historicization of this demand for retooling shows us two possibilities: (1) it is either that the process of scientific change has become altered in our times or (2) that the nature of normal science and scientific change was never pure of social and economic motivations as Kuhn discusses it to be. We are not in a position to argue for either of them but both of them points to the same direction. the Computational Turn in social sciences, to a large extent, is a result of the phenomenal domain opened up not as a consequence of the desire for knowledge but for profit. It is a domain that came to be in the process of expansion of the capital accumulation. To remind, surveillance of users of the World Wide Web has started simultaneously when IT and ICT sectors were in a crisis where the solution was to create new sources of surplus value. It is the mechanism that creates this source of surplus value that depends on the commodification of personal information and its use in personal advertising that determines the form and the content of the phenomenal domain opened up by big data.

As Kuhn himself states "As in manufacture so in science retooling is an extravagance to be reserved for the occasion that demands it." (Kuhn, 1970, p. 76). Capital's need for expansion, to find another domain that can be commodified and be made part of capital accumulation is what demanded retooling in the private sector. The condition of possibility of the popularization of cloud computing that made SaaS and the similar other business strategies possible and profitable and which shifted the location of computing and along with the place where data is generated and stored and therefore effectively working as a data extraction apparatus was this demand. What ties together the science and the economy or the contemporary capitalism is this relationship. Technological advancements or revolutions in measurement cannot precede this relation. As we have discussed big data gains its predicates, content, and form from exactly this relation. Nevertheless, social big data comes across as an objective reality in all its rawness that demands a sensibility enabled by technology to make sense of. As Mazzocchi states

> ... Big Data will allow us to lessen our yearning for exactitude. Rather than seeking accurate results under controlled and simplified conditions, scientists are driven to see in the messiness of data a reflection of the complexity of nature. (Mazzocchi, 2015, p. 1252)

Within the context of our discussion, it seems that an aggregation of information where every variable one way or another is of a consumer and is organized around the motivation to generate profit appears as the neutral representation of social reality that awaits scientific inquiry. That scientific inquiry has to be computational since the raw material cannot be accounted with traditional tools. The problem with this, as Boyd and Crawford also argue, is that data is something that needs to be extracted; it has to be regarded as data prior to its analysis (Boyd & Crawford, 2012). The commodified nature of social big data and the capitalist context it emerges out of is important exactly for this reason. Social big data, prior to its consideration as an objective source of information for scientific inquiry, is considered as a source of value. Accordingly, it is generated and analyzed with maximum efficiency in mind. As I have tried to show with the help of Lyotard, the particular relationship between technology and knowledge is not an accident, it is determined by contemporary capitalism. Therefore, it can be argued that it is not so much the complexity of the real social world that demanded computational methodologies but rather it is the social big data that demands it. The interactions between people, digital devices, digital entities, digital transactions are made sources of value, meaning that they are commodified, with a certain technical apparatus and with certain tools in mind to process it. Therefore, the correspondence is not between the real world and its representation in social big data but rather between social big data and the technical apparatus that generates it.[12]

This, however, does not mean that CSS is not a legitimate scientific domain of inquiry or that it does not produce knowledge. The theoretical framework that CSS presents to us that we have covered in the previous section allows CSS to be a discipline on its own. The fundamental assumptions, premises, and purposes of Computational Social Science can be traced before the emergence of social big data and that contributes to our point. As Kuhn states paradigms are often developed, albeit not fully, in the absence of a scientific crisis or of its recognition (Kuhn, 1970). In these terms, the crisis is understood as a moment where the power relationship between paradigms change. In our case, that crisis was induced by the emergence of social big data that transformed CSS from a simulation and modeling focused discipline to a data-driven one.

Within the framework that CSS supplies that data has come to be a source of scientific knowledge but for that to happen it has to be considered as an objective, raw source in the first place. While whether data can ever be raw is another discussion[13], as stated above, the particular data we have been discussing is made value, is commodified prior to its recognition as a scientific source of knowledge. It is organized around the purpose of profit which allows its generation, collection, and analysis without context that grants them any meaning that can be uncovered. This loss of meaning and the reorganization around another context works for commercial purposes but it does not necessarily produce knowledge because the data is not meant to go beyond reflecting relations of correlation.

[12] This claim should be read not as a one that is about the nature of reality but as a logical conclusion based on the simple idea that the technical apparatus that is used to generate information determines the nature of that information.

[13] See Boyd & Crawford, 2012 and Gitelman, 2013 for such a discussion.

O’Sullivan argues for another Computational Social Science that is more organized around Complexity Science and tries to explain phenomena, and that would expand our understanding of the world in contrast to one that merely correlates them (O’Sullivan, 2018). But still, the paradigm shift in social sciences is attributed not to that kind of Computational Social Science that would be more focused on producing knowledge through theoretically grounded models but to one that bases itself on the empirical domain opened up by social big data. As Wyly puts it "The capitalist correlation imperative is clear: spurious correlation is fine, so long as it is profitable spurious correlation." (Wyly, 2014, p. 681) and that is justified in business for it offers profit and efficiency.

Let us now briefly consider some of the points discussed in the previous chapter once again in the light of these arguments. As the reader will remember one subsection of advantages of data-driven CSS was dedicated to its practical advantages like the low-cost research, reducing interference of researcher in data collection and so on as Chang et al. argued. As we discussed data is something that needs to be extracted, made visible, and this process, in the case of social big data, is the commodification process and social big data’s exchange value. One of the direct consequences of our argument is that it is not so much the social big data enables researchers to get access to an objective, raw reality that almost resemblances a laboratory environment thereby resolving the three-horned dilemma of research but it is that the researchers get to work with a bundle of data that is formed not through their research questions but one that is formed through extraction of information from users whose relevance is determined by its potential to be transformed to capital. The relationship between knowledge, capitalism, and technology we have been discussing makes it impossible to consider social big data an objective source of information. Moreover, in terms of data-driven Computational Social Science’s ability to go beyond the dichotomy between micro and macro levels of analysis, again, claimed by Chang et al., it follows that the content of social big data determined by profitability and principle of performativity contains only interactions that have the potential to be useful for capital accumulation, that can be sold to advertisers. Since the content gains its relevancy through its potential to become a part of capital accumulation it cannot be argued to contain every aspect of social whether in micro or macro levels. Increased objectivity was another advantage of computational methodologies in social sciences that we have discussed. Benthall's argument that scientific algorithms make intersubjective validation in science still stands but does it overcome human partiality that is identified as a problem of traditional social science by Benthall? Considering that social big data, if nothing else, is partial and therefore its usage in the verification of models loses its ground once its determination by commercial purposes is established. The last advantage of computational methodologies was its ability to bring natural and social sciences closer, allowing social sciences to conduct a physics of culture (Christakis, 2012) which is open to discussion now that the epistemology that makes culture visible is clearly entangled with capitalist priorities and interests (Wyly, 2014) which clearly contradicts with idealization of natural sciences as the pinnacle of scientific objectivity.

In social sciences, to consider this data a correct representation of the world is an epistemological stance that is not grounded in the scientific framework of Computational Social Science. In Kuhnian sense, scientific paradigms are what make the world knowable for they supply a certain sensibility that populates that world (Kuhn, 1970). Because social big data does not emerge out of a scientific paradigm or with respect to it, it brings the question of purity of normal science. Social big data's logic is a capitalist one and it is organized with respect to it. It can be treated as a source of knowledge only if one accepts that it is populated only by the information that relates to capitalist ends. In the case of individuals, for example, Curry states "When individual consumers act, they create digital individuals." (Curry as cited in Wyly, 2014, p. 681). The aforementioned stance, in this sense, essentially declaring the world in its totality a capitalist one and therefore leaves out what is not. It is in this way that it acts as a negative principle, limiting sensibility to technologically enabled empirical manifestations for consumer acts and interactions. As a result, the problem with understanding data-driven CSS as a scientific paradigm stems not from its assumptions or theoretical foundations or methods but from the particular historical determination of the phenomenal domain it operates on.

4.5 Scientific Paradigms in Higher Education

One final issue must be touched upon before concluding this chapter and that is the effect of economic conditions on the organization of universities and higher education. A paradigm change can also be traced to the operation of institutions where knowledge production occurs. A scientific paradigm's ability to solve the problems that it poses is what makes it a consistent entity and enables it to continue without modifications or revolutions (Kuhn, 1970). This continuity is supplied not only through the paradigm's ability to successfully create a coherent domain of problem-solutions but also through the training of new scientists in higher education. Although the central problematic of this thesis is organized around the production of scientific knowledge, Lyotard's historical analysis opens up some points of discussion that we can relate to Kuhn's theory of scientific change.

Both Lyotard and David Berry make a similar point in this regard. The problem-solutions used in training of the students does not match to the requirements of the age, so to say. To remind, Berry claimed that the universities have to give rise to another subjectivity, it needs to train a student whose ability rests not on the content that she accumulates in her education but on technical ability as well as on her ability to come up with a solution that is "here and now" in a just in time fashion (Berry, 2011a). Similarly, the principle of performativity demands a different education system than the previous ones where metanarratives were in a determining position. Higher education, determined by the principle of performativity, has to be one that focuses not on ideals or cultivation of an elite group of people but on creating skills (Lyotard, 1984). What is at stake is that while in Kuhnian framework scientific problems or those that are studied in universities are posited by the established scientific paradigm, in our case, along with the studiable phenomenal domain, the problems are brought about not necessarily by scientific concerns. The principle of performativity does not only legitimize knowledge as such but also the education system. We have covered a few accounts from people in industry positions about the skills mismatch between higher education and the needs of the industry in Chapter 2, however, what is important for us is not that mismatch alone but the demand to change the content of the higher education. This essentially means that the problem-solutions in the education system must be replaced. Because the knowledge production shifts from universities to private sector or industry, its use value gives way to its exchange value, the role of education system transforms into one that enables students to work with the technical apparatus instead of training them in content. Lazer et al. in their article "Computational Social Science" argue for the need to "... develop a paradigm for training new scholars." (Lazer et al., 2009, p. 722). What is at stake is the demand for trained people that can work on the phenomenal domain opened up by commodification.

As a final consideration, the growth of interest, journals, publications and funding in Computational Social Science has been regarded as an indicator of the Computational Turn in social sciences and humanities. Following Lyotard, once again we are faced with the fragility of normal science. When science is dependent on funding which is essentially an investment, a return is expected in the form of a contribution to optimization of the system. Where the principle of performativity is in a determinant position

> Research sectors that are unable to argue that they contribute even indirectly to the optimization of the system's performance are abandoned by the flow of capital and doomed to senescence. (Lyotard, 1984, p. 47)

The relationship between capital, technology, and knowledge makes it evident that the increase in funding for Computational Social Science projects, research centers, and departments as a fact cannot be used in support of the argument for the paradigm shift in social sciences in Kuhnian sense. What it instead shows is not the development of normal science in itself but the fact that the development of scientific enterprise is subject to the operation of a larger system that works on the basis of the principle of performativity.

4.6 Conclusion

What this chapter argued for, very simply, is that knowledge cannot stay unaltered amongst other changes. Looking through the lens provided by Lyotard makes the relationship between technology, capital, and knowledge in contemporary capitalism evident. It is through this relationship, through the extension of commodification that social big data came to be, and it is again on the basis of this relationship that it is not a surprise it emerged out of the industry.

Best exemplified in digital commodities and platforms, commodities and commodity relations change accordingly to become a part of the activity of "knowing" under contemporary capitalism. The particular shift in business strategies of companies that provide digital commodity and services particularly manifests the importance and value of information in contemporary knowing capitalism. Moreover, because it is situated and emerged under these conditions, social big data is determined by its exchange value. Rather than a research question, social big data as a source of knowledge is determined by its potential to generate profit on the basis of a particular strategy of personal advertisement. This, as I have tried to show, has its consequences on the content and on the form of social big data.

The recent rise of computational methodologies in social sciences could not have happened without social big data emerging as a source of information because it is a particular form of Computational Social Science that is celebrated today. And if we are to understand this change as a paradigm shift as Lazer and many others have claimed, it is necessary to locate the alien nature of big data that makes it unaccountable by traditional social science methods. I argued that, complementing our point that the core feature of social big data in contrast to traditional datasets is its purposeless collection in the scientific sense, what defines social big data is its location in contemporary capitalism, in capital accumulation and necessarily the capitalist purpose of profit. The predicates social big data acquires from the context it is generated in disturb the traditional understanding and operation of social science. However, the source of this disturbance is to be found in the historical conditions that determined social big data and not in its intrinsic characteristics. As discussed, data is something that needs to be extracted and that occurs through a particular purpose and mechanism that make it sensible. In our case this process more of a commodifying one that belongs to a capitalist sensibility than a scientific one in Kuhnian sense. What is at hand, as I have tried to show, is opening up and organization of a particular world as an object of knowledge not through a scientific paradigm but through contemporary capitalism where knowledge is a force of production and is organized around commodity relations (Thrift, 2005).

As a final point, I have discussed that, following Kuhn, a scientific paradigm is not something explicitly decided on but shows itself in the training of students of a certain discipline through which they internalize a certain organization of the world. However, as a logical implication of Lyotard's points and of our historical analysis, in our case, it is more likely that the demand for changes in higher education stems not from the desire for knowledge as such, but the higher education is expected to adapt to encompassing capitalist structure. Therefore, to use the increase in funding for computational social science projects, research centers, etc. as a fact that indicates an ongoing Kuhnian paradigm shift in social sciences becomes problematic.

CHAPTER 5

CONCLUSION

The aim of this thesis is to understand the dynamics of the recent transformation in scientific methodologies in social sciences that is argued to be so fundamental that it constitutes a paradigm shift. By bringing out these dynamics, the conditions of possibility of the rise of Computational Social Science are discussed. The analysis of the said dynamics is made mainly through analytical and conceptual tools provided by Thomas Kuhn’s theory of scientific revolutions, Jean-François Lyotard’s report on the position of knowledge in computerized capitalist societies and Nigel Thrift’s theory of knowing capitalism. Let me first give an overview of the core concepts of the analysis and then make some concluding remarks on the particular case of the thesis and the problem of scientific change in general.

The main theoretical framework that enables the problematization of the subject matter in Thomas Kuhn’s "The Structure of Scientific Revolutions" published in 1962. Conducting a sociological analysis of a recent case of scientific change, the problem of this thesis is to discuss whether the development of science is a result of self-contained dynamics or not, and if not, what other dynamics contribute to that process. Although Kuhn’s theory of scientific revolutions is considered a radical account of scientific change, it is still based on the idea of self-contained science. Therefore, since the rise of computational methodologies in social sciences is discussed as a case of paradigm shift in the contemporary literature, the first part of this thesis is dedicated to understand this recent change in methodologies in the Kuhnian framework.

In Chapter 3, by relying on Kuhn’s conceptualization, this thesis looks at the theoretical framework of Computational Social Science and searches for the consistency between ontology, epistemology, and methods that construct a certain world and enable Computational Social Science to operate on it and know it. It is this consistency that enables a discipline to be established as a "normal science" and make it a puzzle solving activity (Kuhn, 1970). Without the said consistency no problem posited by a scientific paradigm can be guaranteed to be answerable. Therefore, in the relevant chapter, this thesis discusses the scientific framework of Computational Social Science and presents the world constructed by the former. It is argued that Computational Social Science works on a world that is understood as a result of aggregations at different levels and aims to observe and uncover the manners in which these aggregations come to be. Put in this way, these manners in which aggregations occur are considered laws that make up and limit the domain of possible interactions in the given world. Likewise, the society is thought as a complex, emergent, and an adaptive system that consists of various elements and is made up of the interactions of the latter. This assumption, as discussed, is both present in simulation and model focused Computational Social Science and in data focused one, for latter social big data is the representation of the complexity of the social.

The most common point raised in order to argue for the validity of Computational Social Science is that the advancements in Information and Communication Technologies, and in storage and processing abilities are considered to be capable of bringing out the complexity of the social. Argued not only by the proponents of Computational Social Science, the technological developments are considered to be the main mechanism that brings scientific change in our particular case. In this view, the argument perhaps can be summarized in Aral’s words "Revolutions in science have often been preceded by revolutions in measurement." (Aral as quoted in Kitchin, 2014b, p. 128). In this view, it seems that the technological advancements and the increased digitalization of social life results in an increase in the scientific ability to gain access and record the dynamics of the social and to analyze it in a rigorous manner. Operating on this idea, Computational Social Science is claimed to be more advantageous than traditional social science because it is capable of granting direct access to both micro and macro levels of the social and enabling intersubjective verification of methods in social sciences, and thereby, increasing objectivity. It is argued that these advantages are points of confrontation with the old paradigm that act as points of self-identification and it is here that the separation between different paradigms are most apparent. In Kuhnian terminology, these advantages are the promises of the new paradigm that are in effect in the process of construction of allegiances within the scientific community. They are promises because the shift to another paradigm does not rest solely on the ability of the new paradigm to overlap with natural phenomena, but it is actualized through a process of competition between paradigms (Kuhn, 1970). So, Chapter 3 serves to show the importance of social big data with respect to the rise of computational methodologies, to establish the world conceptualized by Computational Social Science, and finally, the manner in which social big data fits in that world conception.

The second part of the thesis is a questioning of Kuhnian conceptualization of scientific paradigms and scientific change. Accordingly, Chapter 4 is dedicated to the historicization and contextualization of the claimed paradigm shift. This thesis aims to go beyond technological explanations of its problematic for said explanations do not account for the effects of the purpose of data collection and its apparatus. For contemporary Computational Social Science, characterized by its increasing focus on data analysis, to become a puzzle-solving activity, the phenomenal domain it operates on must match with the fundamental assumptions of the discipline. On the basis of this, Chapter 4 begins with contextualizing the emergence of social big data to understand its characteristics and features. Following Thrift's arguments on knowing capitalism, by locating the changing forms of commodities and commodity relations in the emergence of service-based digital products and platforms, exemplified in SaaS and the similar business strategies, this thesis discusses the particular conditions that enabled social big data to emerge. After showing particular practices that allowed social big data to come to be, the said practices are put into context through Lyotard's ideas on computerized societies and the critical relationship between capital, technology, and knowledge is shown. It is in this relationship that knowledge becomes a part of capital accumulation and principle of performativity becomes a criterion of truth, expressing itself through questions that are concerned with the value of knowledge. The particular characteristic of our case is that the use value of knowledge is intertwined with its exchange value, if not subordinate to it. Its determining power shows itself in the emergence of service-based platforms whose existence depends on their ability to successfully and efficiently commodify user information. It is argued that social big data, which is argued to be the unaccountable through traditional social science's method and tools, is a result of this commodification process. Therefore, its characteristics and features should be understood with respect to this process as well, not as intrinsic to it or to the social as such; its features are shaped and required by its role in capital accumulation. As argued, social big data's emergence is dependent on a particular form of capitalism and a business model that are organized around the extraction information through user surveillance that has to be fine-grained, collected continuously because profit mechanisms depend on personalization of user experience and advertising.

The significance of the above contextualization is made evident in the discussions on the recognition of social big data as the correct or the exact representation of the social world. The questions raised in this thesis are to indicate the taken for granted nature of accepting social big data as neutral, objective, raw information about whole populations that include every aspect, interaction, and transaction between agents in the social. Once the emergence and the features of social big data is put into context, especially when the importance of research question in the collection, organization, and analysis of data in social science is recognized, the danger in the said supposition becomes obvious. It is argued here that the determinative effects of the principle of performativity and metapurpose of profit are not to be ignored in studies of social big data and its potential as a source of knowledge for social sciences.

Finally, again following Lyotard's cue, this thesis touches upon the possible effects of the same principle of performativity as a legitimizing principle that affects higher education. Once we establish the commodified nature of the social big data as a phenomenal domain that is to be worked with and that demands technical expertise, it is argued that it goes beyond mere speculation to argue that the demands for changes in training of students have more to do with than just desire for knowledge.

Let us bring all of these arguments and claims under the Kuhnian theory of scientific change once again for it was the starting point of this thesis. Kuhn decidedly separates scientific paradigms from the surrounding structures, especially in the case of natural sciences. It is this insulation that enables scientists to focus on the problems posited by the scientific paradigm. Once he establishes that purity of science what brings change in science is necessarily limited to itself. What we see in our case, however, the problem being one of paradigm shift, sciences or perhaps social sciences are in constant interaction between other parts of the structure they are practiced in. The domain of data-driven Computational Social Science is opened up not by the previous paradigm or by its failures but by contemporary capitalism's need to expand the areas of commodification. Technology, as Lyotard puts it, working as an extension of sensibility, a point of contact with reality, is not neutral, in contrast, is entangled, in our particular case, with capitalist motives and expectations. What this indicates is nothing other than the partiality of the knowledge produced which is a serious problem considering Computational Social Science aims at producing a universal knowledge on laws of societies and such. However, this point, in the literature, is often ignored.

But still, the fundamental nature and the strength of the changes brought about by social big data and the data-driven Computational Social Science cannot be disregarded. Within the limits of the Kuhnian framework, the move to computational methodologies can be conceptualized as a paradigm shift. Chapter 3 was dedicated to showing how the knowable world established by CSS is distinguished from that of traditional social science and how it requires a complete change in theories, methodologies, and toolkits. This much, however, can only establish that Computational Social Science can be a scientific paradigm itself.

The question of paradigm shift is a more complicated one to answer in Kuhnian framework because what induces a crisis in a discipline and how extraordinary science comes to be to establish a new practice of normal science are not clear if we assume that science works in isolation as Kuhn does. Perhaps with the insights from Savage and Burrows in their article where they declare a jurisdictional crisis, this thesis shows (1) how a scientific domain is opened up by what is not science and without scientific concerns and (2) how the force behind a paradigm shift does not have to emerge in the scientific enterprise or research. In our particular case, this thesis argues that it is a capitalist sensibility that allows objects of knowledge to come to be by making them commodities before they are made objects of knowledge.

In conclusion, what is argued in this thesis is, in a more general level, normal social science is open to the effects of what is not science as such and, in a more particular level, the case of Computational Turn is an example of this. The claims concerning paradigm shift in social sciences require us to take a more reflexive attitude before establishing and justifying the transformation of methodologies by referring to their capability to represent real as it is. This thesis is an attempt at such practice and hopefully will foster further discussion on the issue.

REFERENCES

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. Retrieved from http://www.wired.com/2008/06/pb-theory/

Benthall, S. (2016). Philosophy of Computational Social Science. *Cosmos and History*, *12*(2), 13–30.

Berry, D. M. (2011a). The Computational Turn: Thinking About the Digital Humanities. *Culture Machine*, *12*, 1–22.

Berry, D. M. (2011b). *The Philosophy of Software: Code and Mediation in the Digital Age*. Palgrave Macmillan.

Berry, D. M. (2014). *Critical Theory and the Digital*. Bloomsbury.

Bird, A. (2012). The Structure of Scientific Revolutions and its Significance: An Essay Review of the Fiftieth Anniversary Edition. *The British journal for the Philosophy of Science*, *63*(4), 859–883. Retrieved from https://academic.oup.com/bjps/article-lookup/doi/10.1093/bjps/axs031 doi: 10.1093/bjps/axs031

Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, *15*(5), 662–679.

BritishAcademy. (2012). *Society Counts: Quantitative Skills in the Social Sciences and Humanities*. London, UK: British Academy.

Burrows, R., & Savage, M. (2014). After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data & Society*, *1*(1).

Castellani, B., & Hafferty, F. (2009). *Sociology and Complexity Science: A New Field of Inquiry*. doi: 10.1007/978-3-540-88462-0

Chang, R., Kauffman, R., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, *63,* 67–80. doi: 10.1016/j.dss.2013.08.008

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mob. Netw. Appl.*, *19*(2), 171–209.

Chen, S.-h., & Venkatachalam, R. (2017). Agent-based modelling as a foundation for big data. *Journal of Economic Methodology*, 1–22. doi: 10.1080/1350178X .2017.1388964

Chen, X. (1997). Thomas Kuhn's Latest Notion of Incommensurability. *Journal for General Philosophy of Science / Zeitschrift Für Allgemeine Wissenschaftstheorie*, *28*(2), 257-273. Retrieved from http://www.jstor.org/stable/25171092

Christakis, N. (2012). *A new kind of social science for the 21st century*. Retrieved from https://www.edge.org/conversation/nicholas_a_christakis-a-new-kind-of-social-science-for-the-21st-century

Cioffi-Revilla, C. (2017a). Computation and Social Science. In *Introduction to computational social science: Principles and applications* (pp. 35–102). Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-50131-4_2 doi: 10.1007/978-3-319-50131-4_2

Cioffi-Revilla, C. (2017b). Introduction. In *Introduction to computational social science: Principles and applications* (pp. 1–33). Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-50131-4_1 doi: 10.1007/978-3-319-50131-4_1

Cioffi-Revilla, C. (2017c). Simulations I: Methodology. In *Introduction to computational social science: Principles and applications* (pp. 375–414). Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-50131-4_8 doi: 10.1007/978-3-319-50131-4_8

Cioffi-Revilla, C. (2017d). Social complexity I: Origins and Measurement. In *Introduction to computational social science: Principles and applications* (pp. 193–246). Springer International Publishing. Retrieved from https:// doi.org/10.1007/978-3-319-50131-4_5 doi: 10.1007/978-3-319-50131-4_5

Cohen, N. (2008). The Valorization of Surveillance: Towards a Political Economy of Facebook. *Democratic Communiqué*, *22*(1), 5. Retrieved from http:// journals.fcla.edu/demcom/article/view/76495

Conte, R., Gilbert, N.G., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertész, J., Loreto, V., Moat, S.J., Nadal, J., Sánchez, A., Nowak, A., Flache, A., Miguel, M.S., & Helbing, D. (2012). Manifesto of computational social science. *The European Physical Journal Special Topics, 214(1)*, 325-346.

Fuchs, C. (2010). Labor in Informational Capitalism and on the Internet. *The Information Society*, *26*(3), 179–196. Retrieved from https://doi.org/ 10.1080/01972241003712215 doi: 10.1080/01972241003712215

Fuchs, C. (2012). Google Capitalism. *tripleC: Communication, Capitalism Critique. Open Access journal for a Global Sustainable Information Society*, *10*(1), 42– 48. Doi: 10.31269/ ripleC.v10i1.304

Fuchs, C. (2015). *Digital Labour and Karl Marx*. Routledge.

Gane, N. (2003). Computerized capitalism: The Media Theory of Jean-François Lyotard. *Information, Communication & Society*, *6*(3), 430–450. Retrieved from https://doi.org/10.1080/1369118032000155294 doi: 10.1080/1369118032000155294

Gartner. (n.d.). Public cloud application services/software as a service (SaaS) market revenues worldwide from 2015 to 2022 (in billion U.S. dollars). In *Statista – the statistics portal.* Retrieved March 10, 2019, from https://www.statista.com/statistics/505243/ worldwide-software-as-a-service-revenue/

Gitelman, L. (2013). *"Raw Data" Is an Oxymoron*. MIT Press.

Godse, M., & Mulik, S. (2009). An Approach for Selecting Software-as-a-Service (SaaS) Product. *2009 IEEE International Conference on Cloud Computing,* 155-158.

Hardey, M., & Burrows, R. (2008). Cartographies of Knowing Capitalism and the Changing Jurisdiction of Empirical Sociology. In *The handbook of online research methods* (pp. 507–518). SAGE Publications.

Hey, A. J. G., & Hey, A. J. G. (2009). Jim Gray on eScience: A Transformed Scientific Method. In *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Microsoft Research.

Hey, T., Tansley, S., & Tolle, K. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research.

Hoyningen-Huene, P. (1989). Idealist Elements in Thomas Kuhn's Philosophy of Science. *History of Philosophy Quarterly, 6*(4), 393-401. Retrieved from http://www.jstor.org/stable/27743912

IBM. (2017). *10 Key Marketing trends for 2017.* Retrieved from https://www-01.ibm.com/common/ssi/cgi-bin/ssialias ?htmlfid=WRL12345USEN

Kaplan, J. M. (2007). SaaS: Friend or foe?. *Business Communications Review*, 48–53.

Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, *3*(3), 262–267.

Kitchin, R. (2014a). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1).

Kitchin, R. (2014b). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Los Angeles, CA: Sage.

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1).

Kuhn, T. (1970). *The Structure of Scientific Revolutions* (Second ed.). University of Chicago Press.

Law, J., Ruppert, E., & Savage, M. (2011). The Double Social Life of Methods. (CRESC Working Paper No.95).

Lazer, D., Pentland, A., Adamic, L., Aral, S., Fowler, J., Gutmann, M., ... Contractor, (2009). Computational Social Science. *Science*, *323*(5915), 721–723.

Lyotard, J.-F. (1984). *The Postmodern Condition: A Report on Knowledge. Translation from the French by Geoff Bennington and Brian Massumi. Foreword by Frederic Jameson.* (Vol. 10).

Macy, M. W., & Willer, R. (2002). From Factors to Actors: Computational Sociology and Agent-Based Modeling. *Annual Review of Sociology*, *28*, 143–166. Retrieved from http://www.jstor.org/stable/3069238

Mann, R. (2012). Five minutes with Prabhakar Raghavan: Big data and social science at Google. (2012, September 21). Retrieved April 20, 2019, from https://blogs.lse.ac.uk/impactofsocialsciences/2012/09/19/five-minutes-with-prabhakar-raghavan/

Mazzocchi, F. (2015). Could big data be the end of theory in science?: A few remarks on the epistemology of data-driven science. *EMBO reports*, *16*(10), 1250–1255. doi: 10.15252/embr.201541001

McGrath, J. E. (1981). Dilemmatics: The Study of Research Choices and Dilemmas. *American Behavioral Scientist*, *25*(2), 179–210. Retrieved from https://doi.org/10.1177/000276428102500205 doi: 10.1177/000276428102500205

O'Sullivan, D. (2018). Big data: why (oh why?) *this* computational social science? In *Thinking big data in geography: New regimes, new research* (pp. 21–38). University of Nebraska Press. Retrieved from http://www.jstor.org/stable/j.ctt21h4z6m.7

Peterson, G. L. (1981). Historical Self-understanding in the Social Sciences: The Use of Thomas Kuhn in Psychology. *Journal for the Theory of Social Behaviour*,*11*(1), 1-30. doi:10.1111/j.1468-5914.1981.tb00020.x

Pietsch, W. (2013). Big Data – The New Science of Complexity. Retrieved from http://philsci-archive.pitt.edu/9944/

Raghavan, P. (2014). It's time to scale the science in the social sciences. *Big Data & Society*,*1*(1), 205395171453224. doi:10.1177/2053951714532240

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH 87*. doi:10.1145/37401.37406

Ruppert, E. (2013). Rethinking empirical social sciences. *Dialogues in Human Geography*, *3*(3), 268–273. doi: 10.1177/2043820613514321

Ruppert, E., Law, J., & Savage, M. (2013). Reassembling Social Science Methods: The Challenge of Digital Devices. *Theory, Culture & Society*, *30*(4), 22–46.

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*,*6*(1), 205395171882054. doi:10.1177/2053951718820549

Savage, M. (2013). The 'Social Life of Methods': A Critical Introduction. *Theory, Culture & Society*, *30*(4), 3–21. doi: 10.1177/0263276413486160

Savage, M., & Burrows, R. (2007). The Coming Crisis of Empirical Sociology. *Sociology*, *41*(5), 885–899.

Schütz, S. W., Kude, T., & Popp, K. M. (2013). The Impact of Software-as-a-Service on Software Ecosystems. In G. Herzwurm & T. Margaria (Eds.), *Software Business. From Physical Products to Software Services and Solutions* (pp. 130–140). Springer Berlin Heidelberg.

Squazzoni, F. (2010). The Impact of Agent-based Models in the Social Sciences After 15 Years of Incursions. *History of Economic Ideas*, *LVIII*.

Srnicek, N. (2017). *Platform Capitalism*. Polity.

Stanley, L. (2008). It has Always Known, and we have Always been ‘other’: Knowing Capitalism and the ‘Coming Crisis’ of Sociology Confront the Concentration System and Mass-Observation. *The Sociological Review*, 56(4), 535–551. https://doi.org/10.1111/j.1467-954X.2008.00804.x

Thrift, N. J. (2005). *Knowing capitalism*. SAGE Publications.

Törnberg, P., & Törnberg, A. (2018). The limits of computation: A philosophical critique of contemporary Big Data research. *Big Data & Society*,*5*(2), 205395171881184. doi:10.1177/2053951718811843

Website (fourquadrant.com). (n.d.). Public cloud software as a service (saas) market share worldwide in 2017, by vendor. In *Statista - the statistics portal*. Retrieved March 11, 2019, from https://www.statista.com/statistics/478112/saas-vendor-market-share-ranking-worldwide/

Wyly, E. (2014). Automated (post)positivism. *Urban Geography*,*35*(5), 669-690. doi:10.1080/02723638.2014.923143

APPENDICES

A. TURKISH SUMMARY/TÜRKÇE ÖZET

Bu tezin amacı son yıllarda sosyal bilimlerde gittikçe popülerleşen bilgisayımsal metodolojilere ve büyük veri kullanımına dayanan yaklaşımların yaygınlaşmasına katkıda bulunan faktörleri anlamak ve tartışmaktır. Bu faktörleri açık etmek ve tartışmak bilgisayımsal metodolojilerin sosyal bilimlerdeki yaygınlaşmasının ve genel anlamda bilimlerde değişimin dinamiklerini anlamak için gerekli ve önemlidir. Bu özette çalışmanın kullandığı genel kavramsal çerçeve kısaca tartışılacak ve bilimlerde değişim problemi hakkında belli başlı sonuçlar sunulacaktır.

Bu tez literatür taraması bölümünde öncelikle bilgisayımsal sosyal bilimin yaygınlaşmasını bir vaka olarak ele almış ve literatürdeki bu konudaki tartışmaları sunmuştur. Bilgisayımsal metodolojilerin yaygınlaşması, büyük verinin bir fenomenal alan olarak sosyal bilimlere açılması ve bilgisayımsal sosyal bilimin bir disiplin olarak popülerleşmesi literatürde Kuhn'cu anlamda bir paradigma değişimi olarak tartışıldığından önce bu konudaki görüşler sunulmaktadır. David Berry'nin de ifade ettiği gibi "…beşeri bilimlerde sorulan soruların çoğunun düşünülebilirliğinin temel koşulu artık bilgisayımsal teknolojiler olmuştur." [14] (Berry, 2011a, p.2)

Böylesine derin bir değişim ise Chang, Kauffman ve Kwon tarafından Kuhn'cu bir paradigma değişimi olarak sınırlandırılmıştır (Chang et al., 2014). Savage ve Burrows ise 2007 yılında yazdıkları "The Coming Crisis of Empirical Sociology" makalelerinde ise bu durumu bir kriz olarak ifade etmişlerdir. Krizin temel nedeni, Savage ve Burrows'a göre, sosyolojinin artık gelinen çağda objesinin bilgisini üretmek için yeterli ve gerekli metod ve araçlara sahip olmamasıdır. Bu

[14] Aksi belirtilmedikçe bütün çeviriler tarafıma aittir

durum ise sosyalin bilgisinin üretiminin akademiden çıkıp şirketler ve kurumların hükmü altına girmesiyle sonuçlanmaktadır (Savage & Burrows, 2007).

Bu tartışmalar ilk kısımda dijital çağda sosyal bilimlerin pozisyonu ve büyük verinin ortaya çıkışının sosyal bilimlere etkisi anabaşlıkları altında izlenmiş ve bu kısım büyük verinin bir fenomenal alan ve sosyal bir fenomen olarak etkileri sunularak bitirilmiştir. Literatür taramasının ikinci kısmında ise tezin izleyeceği genel analitik çerçeve sunulmuş ve bilgi üretimi bir tartışma kounsu olarak ele alınmıştır. Öncelikle Jean-François Lyotard'ın "Postmodern Condition: A report on knowledge" adlı kitabında ortaya sunduğu teorik çerçeve tartışılmış ve kapitalizm, bilgi üretimi ve teknoloji arasındaki ilişkinin taslağı çizilmiştir. Daha sonra ise Nigel Thrift'in "Knowing Capitalism" kitabında metalar ve meta ilişkilerinin bilen kapitalizm altındaki değişimleri konusunda yaptığı saptamalar sunulmuş ve bu değişimlerin Lyotard'ın sunduğu çerçeve ile ilişkisi gösterilmiştir. Literatür taramasının son kısmında ise Thomas Kuhn'un "The Structure of Scientific Revolutions" adlı çalışmasında sunduğu teorik çerçeve tartışılmış ve bu tartışmanın bilgisayımsal metodolojilerin yaygınlaşması vakası ile olan ilişkisi kurulmuştur.

Bu tartışmanın vakasının analizinde kullandığı kavramlar genel olarak şunlardır. Tezin ilk kısmı bilgisayımsal sosyal bilimi bir disiplin ve bilimsel bir paradigma olarak kurma çabasında olduğundan, bu kısımda kullanılan ana kavramlar Thomas Kuhn'un sunduğu çerçeveden alınmıştır. Thomas Kuhn bilimsel çalışmaların bir paradigma olmaksızın işleyemeceğini şöyle ifade etmektedir. "Bir paradigmanın yokluğunda bir bilimsel alanı ve onun gelişimini alakadar edebilecek bütün unsurlar aynı oranda alakalı gözükür." (Kuhn, 1970, p. 15). Bir bilimsel paradigmanın yokluğunda fenomenal alanı düzenleyecek ve şeylerin bilimsel objeler olarak varlığa gelmesine olanak verecek herhangi bir düzenleyici prensip bulunmaz ve dolayısıyla bilimlerin ve disiplinlerin karşılaştığı her unsur neredeyse aynı derecede bağıntılı ve alakalı gözükür.

Thomas Kuhn bilimsel paradigmaları iki anlamda tanımlar. İlk tanımında bilimsel paradigmalar karşımıza daha önce yapılmamış, bilimsel anlamda devrimci ve orijinal örnek çalışmalar olarak karşımıza çıkar. İkinci anlamında ise bilimsel paradigmalar hakim oldukları disiplini ontolojik, epistemolojik ve metodolojik anlamda organize eden yapılar olarak tartışılır (Kuhn, 1970). Kuhn bunu şöyle ifade eder:

> …yakından bakıldığında normal bilim doğayı paradigma tarafından verilmiş, önceden belirlenmiş ve çoğunlukla esnek olmayan bir kutuya sokma çabası olarak karşımıza çıkar. (Kuhn, 1970, p. 24)

Bilimsel paradigmaların bu düzenleyici etkisi bir bilimin objesini, obje hakkında sorulabilecek soruları ve kullanılabilecek metodları belirlediği gibi bilim insanlarının kendi camialarında iletişim kurabilmesini mümkün kılar. Bu tanımdan yola çıkarak analizin ilk kısmında bilgisayımsal sosyal bilim bir bilimsel paradigma olarak anlaşılmaya çalışılmış ve ontolojisi, epistemolojisi ve metodolojisi kurulmuş ve sunulmuştur. Öncelikle bilgisayımsal sosyal bilimin tarihsel gelişimi ve kompleksite teorisi ile olan ilişkisi tartışılmış ve bu teorinin belli bir anlamda bilgisayımsal sosyal bilimin ontolojik ve epistemolojik çerçevesini kurduğu ileri sürülmüştür. Bu anlayışın bir örneği olarak ise *ajan temelli modelleme* metodu verilmiştir. Bu bölümün ikinci kısmında ise bilgisayımsal sosyal bilimlerin tarihsel gelişimi sunulmuş ve öncelikle modelleme ve simülasyon odaklı bir disiplin olan bilgisayımsal sosyal bilimin büyük verinin ortaya çıkışından sonra daha veri ve veri analizi odaklı bir disiplin olmaya başladığı ve aynı zamanda son yıllarda popülerleşen bilgisayımsal metodolojilerin bu kanada ait olduğu belirtilmiştir (Törnberg & Törnberg, 2018). Disiplinin bu tarihsel değişimi büyük verinin ortaya çıkışının bilgisayımsal metodolojilerin yaygınlaşmasındaki kritik etkisini göstermektedir. Dolayısı ile analizin bu kısmında büyük verinin onu daha önce sosyal bilimlerde kullanılan verilerden farklı kılan, sosyal bilimlerde metodolojik bir değişimi gerektirmesine sebep olan özellik Kitchin ve McArdle tarafından sunulan hacim, hız, çeşitlilik, kapsamlılık, çözünürlük ve dizinsellik, ilişkisellik, genişleyebilirlik ve ölçeklenebilirlik, değerlilik, değişkenlik özellikleri tartışılmıştır (Kitchin & McArdle, 2016). Özetle bu özellikler büyük verinin hacminin çok büyük olduğunu, hızla genişleyebilirliğini, birçok kaynaktan beslenebildiğini, hakkında bilgi bulundurduğu entitelerin takip edilebilirliğini ve dolayısıyla farklı veritabanları arasındaki ilişkiselliğini, çok farklı sorulara cevap vermek veya farklı amaçlar için kullanılabilirliğini ifade etmektedir. Büyük verinin literatürde tartışılan bu özellikleri geleneksel sosyal bilimlerin metodolojilerinin ve metodlarının değişmesini zorunlu kılan en önemli faktörlerden ve dolayısı ile bilgisayımsal metodolojilerin kullanımının gereksinilmesinin en büyük sebeplerinden biri olarak ortaya çıkmıştır. Analizin son kısmında ise, yine Thomas Kuhn'u takip ederek, bilgisayımsal metodolojilerin kullanımının geleneksel metodolojilere ve teorilere kıyasla avantajları tartışılmıştır. Bu tartışmanın önemi, bu avantajların yeni bilimsel paradigmanın eski paradigma ile karşılaşma noktaları olmasından gelmektedir. Kuhn'cu anlamda bilimsel devrim veya paradigma değişimi bilimsel komünite içerisinde kurulan birliktelikler veya ittifaklar sonucu ortaya çıkar (Kuhn, 1970). Bu ittifakların kurulması ise bilim insanlarının yeni paradigmaya olan inançlarına dayalıdır. Yeni bir paradigmanın eskisinin yerine geçmesi için yeni bilimsel paradigmanın karşılaştığı bütün fenomenleri açıklayabilmesi gerekmez. Bu değişim tamamen bilimsel topluluklar içerisinde kurulan ittifaklar sonucu gerçekleşmektedir (Kuhn, 1970). İnanç ve yeni bilimsel paradigma hakkındaki umutların kurulması ise yeni paradigmanın avantajlarına tabidir. Yeni bilgisayımsal paradigmanın literatürde en çok konuşulan belli başlı avantajları ise şunlardır:

1. Pratik avantajlar: Veri sürekli olarak toplandığından sosyal bilimci açısından araştırma yapmanın maliyeti düşmüştür. Toplanılan veri ise dijital ortamdaki interaksiyonların "doğal" bir sonucu olarak görüldüğünden sosyal bilimcinin veya araştırmacının veri toplama aşamasındaki dahiliyeti minimuma inmiştir (Chang et al., 2014)

2. Sosyal bilimlerde nesnelliğin sağlanması: Bilgisayımsal metodolojiler sosyal bilimlerde tekrarlanabilir çalışmaları mümkün kılmaktadır. Bunun

yanında kullanılan algoritmalar bilimsel komünite tarafından sürekli incelemeye tabi olduğundan sosyal bilimlerin objektifliği artmaktadır.

3. Mikro ve makro ikililiğinin aşılması: Büyük verinin en küçük interaksiyonlar hakkında da en büyük makro yapılar hakkında da bilgi bulundurmasından dolayı, sosyal bilimci veya araştırmacı artık mikro veya makro seviyeler arasında seçim yapmak durumunda değildir. Büyük veri mikro ve makro seviyeler arasındaki ilişkiyi görünür ve gözlemlenebilir kılmaktadır.
4. Doğa bilimlerinin ve sosyal bilimlerin yakınlaşması: Bilgisayımsal metodolojilerin sosyal bilimlerde kullanılması sosyal bilimleri daha formalize hale getirdiğinden doğa bilimleri ve sosyal bilimler arasındaki arayı kapatmaktadır.

Analizin bu bölümünün sonuçları şunlardır. Bilgisayımsal sosyal bilim bir disiplin olarak ontoloji ve epistemolojisini çoğunlukla kompleksite teoirisnde bulur. Büyük verinin ortaya çıkışından sonra daha veri analizi odaklı bir disipline dönüşmeye başlamıştır. Büyük verinin bu disipline bir fenomenal alan olarak açılması ise ancak büyük verinin sosyal kompleksitenin objektif, doğru, kesin ve ham bir temsili olarak anlaşılması ile mümkün olmuştur. Büyük verinin geleneksel toplanan verilere kıyasla en büyük farkı büyük verinin herhangi bir bilimsel amaçla toplanmıyor oluşudur. Bilimsel bir amacın yokluğu büyük verinin sayılan özelliklerinin çoğunun ana sebebini oluşturur. Analizin bu kısmı sosyal bilimlerdeki son gelişmelerin, paradigma değişimi ve bilimsel devrim iddialarının büyük verinin ortaya çıkışı ile olan kritik ilişkisini açık etmektedir. Büyük verinin yüksek derecede metalaşmış bir bilgi kaynağı olması ise analizin ikinci kısmında Lyotard ve Thrift’in sunduğu çerçeveler üzerinden yürütülen tartışmayı gerektirmiştir.

Tezin dördüncü, analizin ikinci kısmı direk olarak kapitalizm, teknoloji ve bilgi üretimi arasındaki ilişkiye odaklanmakta ve büyük verinin metalaşmış doğasını ve dolayısı ile bilgisayımsal sosyal bilimin popüleşmesini tarihselleştirmektedir.

Jean-François Lyotard "Postmodern Condition: A Report on Knowledge" adlı çalışmasında kapitalist ve bilgisayarlaşmış toplumlarda bilgi üretiminin değişmeden kalamayacağını ifade eder (Lyotard, 1984). Kapitalist ve bilgisayarlaşmış toplumlarda bilgi üretimi gittikçe daha çok teknolojik aygıtlar sistemine dayandıkça bilgi üretimi ve kapitalizm arasındaki ilişki daha da güçlenmektedir. Teknolojik aygıtlar öncelikle kendilerine yüklü miktarda yatırım yapılmasını ve her yatırım bir kar beklentisini gerektirir. Bu ilişki kapitalist toplumlarda bilgi üretimini anlamak açısından kritiktir. Lyotard'ın anlatısında bilim sürekli kendini meşru kılmaya ihtiyaç duyar. Fakat bilim kendi meşruiyetini kendisi kurabilecek bir sistem değildir. Bilim, Wittgenstein'cı anlamda bir dil oyunu(language game) olarak anlaşıldığından, bilimin içerisinde kullanılan kurallar kendi alanlarındaki iddiaları meşru kılabilir. Fakat bilimin kendi meşruiyetini kurması kendi alanının dışına çıkmasını gerektirdiğinden ve bu alanda bilimin kendi kuralları işleyemeyeceğinden bilim kendi kendini meşru kılamaz (Lyotard, 1984). Lyotard bunu şu şekilde ifade eder:

> Bilimsel bilgi kendisinin tek ve doğru bilgi olduğunu kendisinin bilgi olarak bile görmediği diğer anlatı tarzında bilgiye başvurmadan bilemez ve iddia edemez. (Lyotard, 1984, p. 29)

Bu meşruiyet kaynakları tarihin belli dönemlerinde ve farklı bilimsel geleneklerde büyük anlatılar(metanarratives) olmuştur. Kapitalist ve bilgisayarlaşmış, postmodern toplumlarda ise bu meşruiyet kaynakları, büyük anlatılar, ise kaybolmuştur. Bu sebeptendir ki bilgiyi ve bilimsel bilgi üretimini meşru kılan şey artık verimlilik prensibidir (principle of efficiency) (Lyotard, 1984). Bu şu demektir; üretilen bilginin değeri artık kurulan sistemin verimliliğine olan katkısı ve kapitalin akümülasyon sürecinde bir yer edinip edinememesi ile alakalıdır. Özellikle çağdaş kapitalist ve bilgisayarlaşmış toplumlarda bilgi en önemli üretim gücü haline gelmiştir (Lyotard, 1984). Araştırmanın bu kısmında işte bu çerçeve içerisinde büyük verinin ortaya çıkışı tarihselleştirilmektedir.

2000'li yılların başında dot-com balonunun patlaması sonucu bilişim sektöründeki şirketlerin işletme stratejilerini değiştirmesi büyük verinin ortaya çıkışı açısından kritiktir. Bu dönüşümde yazılım ve çevrimiçi servisleri bir meta olarak üretip satma stratejisi bir kenara bırakılmış ve ücretsiz veya küçük bir meblağ karşılığı kullanıcılara abonelik sağlamak ana strateji haline gelmiştir. Örneğin Facebook global bir krizin yaşanmaya devam ettiği 2011 yılında 51.2% kar açıklamıştır (Fuchs, 2015). Yine benzer olarak Google 2010 yılında 8.5 milyar dolar kar açıklamıştır (Fuchs, 2012). Fuchs'a göre bu karlılığın en büyük sebeplerinden biri, Thrift'in de ifade ettiği gibi, metaların veya platformların çoğunlukla kullanıcılar tarafından üretilmesidir. Kullanıcılar bu durumdan maddi anlamda herhangi bir karşılık almadığından Fuchs bu durumu, teorik olarak, sonsuz bir sömürü durumu olarak nitelendirir (Fuchs, 2015).

Bu bahsedilen değişim internet temelli, hizmet olarak yazılım paradigması altında geliştirilen platform ve ürünlerle örneklendirilmiştir. Hizmet olarak yazılım veya platformun karlı bir işletme stratejisi olması bu yazılım, platform ve servislerin kullanıcılarını sürekli bir gözetim altında tutmasına ve kişisel verileri toplamasına bağlıdır. Yatırım yapılan teknolojik aygıtlar sisteminin kar beklentisi olduğundan, toplanılan veri metalaştırılarak kapitalin akümülasyon sürecine dahil edilmek zorundadır. Nigel Thrift'in bilen kapitalizm(knowing capitalism) içerisinde meta ve meta ilişkilerinin değişmesi hakkında yaptığı saptamalar tam bu konuda önemlidir. Thrift bunu durumu şöyle ifade etmektedir:

1. Metalar artık hem üreticiler hem de tüketiciler tarafından üretilmektedir.
2. Metalar bir "deneyim ekonomisi" içerisinde üretilmekte ve bu ekonomi kullanıcıların metalara daimi olarak yatırım yapmasını gerektirmektedir.
3. Metalar artık sürekli, daimi bir şekilde üretilmekte veya "geliştirilmektedir" ve bu sebepten interaktif olmak zorundadırlar (Thrift, 2005).

Sistem verimlilik prensibi içerisinde çalıştığından, metalar, özellikle dijital servis ve platformlar, veri toplama araçlarına dönüşmüşlerdir. Bu metalar sayesinde daimi bir bilgi akışı döngüsü ortaya çıkmaktadır. Kullanıcıların gün geçtikçe metalara zamanları ve kendi haklarındaki verilerle yatırım yapması beklenmekte ve bu süreç içerisinde metalar ve meta ilişkileri sürekli yeniden yaratılmaktadır

(Thrift, 2005). Thrift'in saptamaları Lyotard'ın sunduğu çerçeve içerisinde düşünüldüğünde bu durum anlam kazanmaktadır; yaratılan bilgi akışının iki amacı vardır (1) birincisi veri toplama araçlarını geliştirmek iken, (2) ikincisi kapitalin akümülasyonunu devam ettirmektir. Toplanan verinin değerlenmesi iki süreç içerisinde gerçekleşmektedir. Veri hizmet olarak yazılım anlayışı içerisinde bir alışveriş birimi olarak ortaya çıkmaktadır. Hizmeti sağlayan kurumlar çoğu zaman kullanıcılardan sağladığı ürünü veya hizmeti satın almasını istemektense kişisel veri karşılığı hizmeti "ücretsiz" sunma yolunu seçmektedirler. Bu alışveriş içerisinde kişisel veri bir para birimi fonksiyonu görerek değerlenmektedir. Verinin değer kazandığı ikinci süreç ise toplanan verinin (1) kullanıcılara kişisel reklam sağlama amaçlı analizi, (2) temizlenmiş ve analizi yapılmış verinin satılması ve (3) toplanan verinin veri toplama araçlarını daha verimli hale getirmekte kullanılması adımlarından oluşmaktadır. Bu iki durumda da asıl amaç her zaman Thrift'in işaret ettiği gibi kapitalin akümülasyonunu devam ettirmek ve Lyotard'ın iddia ettiği gibi sistemin verimlilik anlamında optimizasyonunu sağlamaktır. Bu durum ise yukarıda söylendiği gibi kapitalizm, bilgi üretimi ve teknoloji arasındaki spesifik ilişki sonucu ortaya çıkmaktadır. Bilgi üretiminin meşruiyetini sağlayan faktör çağdaş kapitalist ve bilgisayarlaşmış toplumlarda ise bu sürecin ta kendisi olarak ortaya çıkmaktadır. Sonuç olarak büyük veri de işte bu süreç içerisinde ortaya tamamen metalaşmış ve geleneksel sosyal bilimleri değişmeye zorlayan özelliklerini bu metalaşmada bulabileceğimiz bir bilgi kaynağı olarak ortaya çıkmıştır.

Bu sürecin tezin problemi bağlamında önemi ise şunlardır. Analizin ilk kısmında büyük verinin herhangi bir amaçla toplanmadığı iddia edilse de ikinci kısmında ortaya çıkan sonuç verinin toplanışının arkasında yatan sebebin kapitalist bir motivasyon olduğudur. Büyük verinin toplanması yüklü miktarlarda yatırım ve çok sofistike bir teknolojik araçlar sistemi (apparatus) gerektirdiğinden, büyük verinin toplanmasının arkasındaki motivasyon ancak kapitalist olabilir. Büyük verinin toplanmasındaki amaç bilimsel değil kapitalist olduğundan da büyük veri bilimsel bir problemin empoze ettiği organizasyon ve yapılandırmalardan mahrumdur. Boyd ve Crawford'un (2012) ve Kitchin'in (2014a) ve birçok farklı düşünürlerin ifade ettiği büyük verinin geleneksel sosyal bilimlerin teorileri, metodları ve metodolojileri ile çalışılamıyor olmasının arkasındaki en büyük sebeplerden biri budur. Büyük verinin bir önceki kısımda sıralanan özellikleri onun metalaştığı süreçler içerisinde anlaşılabilir ve bu süreçler içerisinde anlaşılmak zorundadır. Büyük verinin geleneksel sosyal bilimlerin metodları ve metodolojileri ile çalışılamıyor oluşunun başlıca sebeplerinden biri büyük verinin metalaşmış doğasıdır. Literatürde bu kısım çoğunlukla gözardı edilmiş ve büyük verinin özellikleri kendinden verili bir şekilde alınmıştır. Analizin ikinci kısmının "Büyük Veri'yi Bağlamı İçerisinde Anlamak" (Understanding Big Data in Context) bölümü büyük verinin özelliklerinin içinde şekillediği süreç içerisine izlenebilirliğini açık etmektedir.

Fenomenal bir alan olarak dünya ve veri kendi başına varolan ve bir anlam taşıyan bir şey olmadığından ve öncelikle kavramsallaştırmaya ihtiyaç duyduğundan (Kuhn, 1970, Boyd & Crawford, 2012) büyük verinin sosyalin objektif ve ham bir temsili görülmesi problemlidir. Bilgisayımsal sosyal bilimin, bilgisayımsal metodolojilerin popülerleşmesi ve bunun bilimsel bir devrim olduğu iddiaları Kuhn'cu anlayış ile ters düşmektedir. Kuhn'un teorisinde bilimsel devrim ancak izole bir şekilde anlaşılan bilimin kendi pratiği içerisinde ortaya çıkan anomalilerin akümülasyonu sonucunda oluşan bir kriz sonucu mümkün olabilir (Kuhn, 1970). Halbuki elimizdeki vaka bilimlerin içerisinde pratik edildiği toplumların ekonomik ve sosyal yapıları ile iç içe olduğunu ortaya koymaktadır. Bilgisayımsal sosyal bilim uzun süredir var olmasına rağmen ancak büyük verinin ortaya çıkışı sonucunda yeteri kadar popülerleşmiştir. Fakat, Savage ve Burrows'un da ifade ettiği gibi, bu durum sosyal bilimlerin Kuhn'cu anlamda kendi dinamikleri sonucu değil aksine bir fenomenal alanın, hallice metalaşmış bir fenomenal alanın, içinde bulunduğumuz kapitalist ve bilgisayarlaşmış toplumun dinamikleri sonucu açılmasının sonucudur.

Sonuç olarak bu tez sosyal ve ekonomik faktörler ile bilgi üretimi arasındaki ilişkiyi problematize etmiş ve tartışmıştır. Bu bağlamda son yıllarda sosyal bilimlerde gittikçe popülerleşen bilgisayımsal metodolojilerin yaygınlaşmasını kendine örnek bir vaka olarak almıştır. Bilgisayımsal metodolojilerin popülerleşmesinin varoluş koşullarından biri büyük verinin ortaya çıkışıdır. Büyük verinin toplanması arkasındaki teknoloji ile iç içe geçmiş kapitalist motivasyon onun metalaşmış doğasını göstermekte ve sosyal bilimlerde ontolojik bir kaymayı işaret etmektedir. Bilgisayımsal sosyal bilimin ortaya çıkışı literatürde bilimsel bir devrim olarak tartışılsa da bilimi izole bir şekilde ele alan Kuhn'cu teorik çerçeve buna izin vermemekte ve aynı sebepten sosyal bilimlerdeki son gelişmeleri anlamak açısından yetersiz kalmaktadır. Lyotard ve Thrift'in sunduğu çerçevelerle yapılan bu tezdeki analiz bize şunu göstermektedir; bilimlerin fenomenal dünyalarının değişmesi veya yenilerinin açılması sadece bilimin kendi içerisindeki dinamiklere bağımlı değil aksine direk olarak bilimsel bilgi üretiminin pratik edildiği yapının koşullarına bağlıdır.

B. THESIS PERMISSION FORM/TEZ İZİN FORMU

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences	
Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences	X
Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics	
Enformatik Enstitüsü / Graduate School of Informatics	
Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences	

YAZARIN / AUTHOR

Soyadı / Surname : AK
Adı / Name : ONURHAN
Bölümü / Department : SOSYOLOJİ

TEZİN ADI / TITLE OF THE THESIS (**İngilizce** / English) : COMPUTATIONAL PARADIGM SHIFT: A HISTORICIZATION OF RECENT DEVELOPMENTS IN SOCIAL SCIENCES

TEZİN TÜRÜ / DEGREE: **Yüksek Lisans** / Master [X] **Doktora** / PhD [ ]

1. **Tezin tamamı dünya çapında erişime açılacaktır.** / Release the entire work immediately for access worldwide. [X]
2. **Tez iki yıl süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of two years. * [ ]
3. **Tez altı ay süreyle erişime kapalı olacaktır.** / Secure the entire work for period of six months. * [ ]

** Enstitü Yönetim Kurulu kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir.*
A copy of the decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.

Yazarın imzası / Signature .... **Tarih** / Date ....