Inferring causation from big data in the social sciences

Virginia Ghiara

University of Kent

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

2019

Word count: 72,465

I, Virginia Ghiara, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Date:

Signature:

Abstract

The emergence of big data has become a central theme in scientific and philosophical discussions. A main tenor in the literature is that big data can drastically change the way in which causal studies are conducted. My thesis aims to explore how big data can be used to establish causal relationships in the social sciences. The beginning of the thesis will focus on data-driven studies and will investigate some of the limitations that characterise this type of study. This analysis will lead me to identify three key challenges of big data for causal studies in the social sciences. The first challenge is how to overcome the limitations of data-driven causal studies. This challenge is motivated by the observation that, regardless of how sophisticated they are, causal data-driven methods can suffer from bias. The second challenge is how to understand the role of ethnographic, qualitative data in causal studies based on big data. This challenge appears vital in the social sciences, where some researchers remain hesitant about the use of data-driven methods and try to defend the importance of qualitative, 'thick' data. The third challenge is how to use big data, in the social sciences, to obtain evidence of causality that goes beyond correlations. This challenge is strongly associated with the idea that, in order to establish causation, both the presence of a correlation between the cause and the effect, and the presence of a mechanism linking the cause and the effect need to be established. This idea, originally proposed by Russo and Williamson (2007) and known by the name of the Russo-Williamson thesis, will be discussed in detail to provide a solution to the first challenge. I will argue that researchers should comply with such a thesis to overcome the limitations of data-driven causal studies in the social sciences. Next, I shall examine the discussions on mixed methods research to claim that qualitative ethnographic data can be used both to collect evidence of social mechanisms, and to help researchers to obtain a comprehensive understanding of the phenomenon under study. Finally, I shall argue that big data can be used, in specific circumstances, to collect evidence of entities and activities constituting causal mechanisms, and that big data might be used to identify sociomarkers, the social version of biomarkers, to trace causal processes that evolve over time.

Acknowledgements

This thesis would not have been written without the support of Jon Williamson and David Corfield. Our conversations about philosophy of science, causality and the social sciences always challenged me to formulate better arguments. Jon taught me immensely about clear philosophical writing. He has always pointed to me new enriching professional opportunities, encouraging me to develop new projects. David has helped me not to lose sight of the "social sciences", giving me excellent food for thought. Thanks to him I discovered what Adverse Childhood Experiences are and I managed to develop my idea of sociomarkers more clearly.

The intellectual environment is incredibly important when working on a PhD project. I would like to thank the EBM+ group and the Evidence Seminar Group for all the stimulating arguments that we have had. Particular thanks must go to Daniel Auker-Howlett, Micheal Wilde, Veli-Pekka Parkinen, Christian Wallman, Graeme Forbes and James Hoctor. Special thanks also to Alyx Robinson, who kindly proofread one of my published articles.

Doing philosophy of science discussing openly with scientists is an incredibly stimulating experience. I have to thank the LIFEPATH group, and in particular Paolo Vineis, Mauricio Avendano, Oliver Robinson and Jessica Laine, for this opportunity.

These acknowledgements would not be complete without vigorously thanking Federica Russo, Phyllis Illari and Brendan Clarke. Phyllis and Brendan welcomed me into their department and have repeatedly provided remarkably insightful feedback on this project. Federica showed me how beautiful it can be to work in philosophy of the social sciences and has always helped me think big. Her support has fundamental during this project.

Friendship and philosophy were beautifully intertwined in the discussions with Mara Floris, Osman Caglar Dede and Stefano Canali. I must thank them as well as my friends Valentina Baldovino, Clara Paglialonga, Davide Tosches and Lucia Modena for supporting me and bringing me joy and happiness in the past three years.

I must thank my sisters, Arianna, Benedetta and Margherita, my mum Graziella and Matteo for their unconditional support. Finally, I thank my adventure partner, Antonio. Thank you for supporting me, stimulating me and accompanying me every day.

List of Figures

List of Tables

Glossary

Constitutive mechanisms: mechanisms that can explain the phenomena by describing the organizations of entities and activities that actualise the capacity making the phenomena-to-be-explained possible (p. 146)

Etiological mechanisms: mechanisms that, through time, lead to the production of the phenomena-to-be-explained (p. 146)

Mixed methods research (MMR): the combination of quantitative and qualitative methods in the same study (p. 77)

Process tracing: approach to trace the causal chain connecting the putative cause and the putative effect (p. 140)

Russo-Williamson Thesis (RWT): both the identification of a correlation, or difference-making relationship, between the cause and the effect, and the identification of a mechanism linking the cause and the effect are required in order to establish causation (p. 49)

Thick data: in general described as ethnographic, qualitative data collected by researchers (not in automated way) and characterised by contextual complexity which help to recognise not only what people do, but also the reasons behind their activities (p. 42)

Table of Contents

1 Introduction

1.1 The emergence of big data

In this thesis, I aim to explore how big data have changed the way in which causal studies are conducted in the social sciences. Over the last few years, the term 'big data' has become pervasive, although in several discussions it is not clear what the difference between data and big data is. A careful analysis of the literature, indeed, shows that big data can be defined in different ways, depending on which aspect is highlighted (Kitchin & McArdle, 2016; Schutt & O'Neil, 2013, p. 24).

One of the most popular claims on big data is that such data possess a cluster of characteristics that makes them different from traditional data. Among these features, the most important are claimed to be the so-called 'Three V's': volume, velocity and variety (Laney, 2001). The first 'V', volume, can be interpreted in different ways: it might denote the amount of data produced, the size of each datum, or the total storage required by a dataset[1]. Velocity, furthermore, is considered another vital aspect of big data. With this term, researchers usually indicate the speed at which data are created and at which they can flow from one dataset to another. Finally, the third 'V', variety, stands for the numerous formats in which data can be collected. In the social sciences, for instance, relevant data might include structured data such as GPS coordinates and numerical indicators, and unstructured data like satellite images and videos[2].

After Laney's proposal, four additional 'Vs' have been proposed in the literature to denote some characteristics or challenges of big data (Khan et al., 2014): validity, veracity, volatility and value. Validity is referred to the correct use of such data, while veracity concerns the truthfulness of data, that may contain bias and inaccuracies, or might be very noisy. Volatility denotes the extent to which data can remain available and reusable over time. Finally, value regards the type of value that can be attributed to data: obviously, data have value because they might provide scientific evidence, but it might also be argued that nowadays data have economic and ethical value. Data and the insights

[1] In all these cases, what should count as a 'large volume' remains ambiguous, given that it depends on the ability to generate, store and disseminate data. What was considered a 'large volume' of data some decades ago, indeed, may not be defined as such nowadays.

[2] Data collected through qualitative studies are in general not included in the definition of big data. As I will examine in chapter 2, section 2.4, in the social sciences data collected through qualitative methods are in general considered 'thick data'.

obtained from them can indeed be traded in the marketplace and can contain personal information on human subjects (for more on this discussion see Borgman, 2016; Kitchin, 2014a; Mayer-Schönberger & Cukier, 2013).

As argued by Kitchin and McArdle (2016), the features represented by such 'Vs' cannot be identified in all the datasets that are generally thought to contain big data. The analysis of existing big datasets, in fact, leads to the conclusion that it is impossible to find a cluster of features shared by *all* big data.

Some examples can help to clarify this observation. First, datasets containing sound sensor data, in general considered big data, consist of amounts of data that are often smaller than the amounts contained in traditional census datasets. In this case, the amounts of data produced are not bigger than the amounts of traditional data. Second, in some cases (like for call records data and GPS data), the size of a big datum is similar, if not smaller, than the size of a traditional datum. Third, in numerous cases big data contained in datasets are not characterised by variety. Datasets containing data generated by sensor devices or financial data, for instance, generally contain only numeric, structured data.

Overall, the 'Vs approach' provides a broad list of characteristics of big data from which researchers can benefit, and some key challenges that are now at the heart of several methodological and ethical discussions.

In this thesis, I shall adopt the notion of big data I recently developed (Ghiara, 2016, p. 32), according to which big data are in general produced through automated processes and, for this reason, can be created very frequently and almost in real time, as described by the "V" standing for velocity.

My consideration can be combined with the observation that big data have significantly changed scientific practice. As proposed by Leonelli (2014b), the novelty of big data does not reside just in some features that big data have, but in two key changes that have characterised scientific research in the last two decades. The first change concerns the status attributed to data: data are now considered the outputs of scientific research rather than mere side products. To give some examples, new infrastructures such as Figshare and new funded projects such as Understanding Society (https://www.understandingsociety.ac.uk) have been developed with the explicit intention to disseminate data as research outputs in their own. The second change consists

in the development of approaches, infrastructures and skills to format, circulate, retrieve, use and interpret data.

1.2 Big data and causal reasoning

My thesis is focused on how big data are used and interpreted in studies aimed at explaining causal phenomena in the social sciences. This topic has recently been at the heart of several discussions in both the social sciences and philosophy of science. On the one hand, social scientists are wondering how to use the available data to explore causal phenomena and what challenges are associated with big data studies, on the other hand many philosophers of science are studying the links between big data and concepts such as causality and explanation. Questions on the use of big data in causal studies, hence, seem to offer fertile ground to develop scientifically-informed philosophical discussions that can, at the same time, both improve philosophical reflections and clarify scientific practices.

The thesis combines philosophical discussions with scientific arguments and case studies. The starting point is the argument, supported both by philosophers and social scientists, that big data do not offer magical solutions to the challenges that characterise causal studies. Data-driven studies are based on inferential (and often automated) forms of reasoning: they use large amounts of data to create models from them. Yet, data-driven results face the same problems that have already been recognised before the emergence of big data. Some social scientists, furthermore, have added that often data-driven studies do not give due importance to qualitative, ethnographic data, although in the social sciences qualitative approaches have produced valuable insights when used to study causal relationships.

These considerations will be analysed in detail in chapter 2, and will lead me to identify three key challenges of big data for causal studies in the social sciences: i) how to overcome the limitations of data-driven causal studies, ii) how to understand the role of qualitative, ethnographic data in causal studies based on big data, iii) how to use big data, in the social sciences, to obtain evidence of causality that goes beyond correlations.

Together with chapter 2, chapters 3 and 4 constitute the first part of the thesis, and examine the first two challenges from a theoretical perspective, drawing some conclusion from conceptual discussions that can be found in the philosophical and social sciences'

literature. More specifically, chapter 3 discusses a way to overcome the limitations of data-driven causal studies, while chapter 4 clarifies how qualitative ethnographic data can improve causal studies based on big data.

Chapter 5 and 6 constitute the second part of the thesis and provide a detailed analysis of two case studies. The case studies are used both to argue in favour of the claims discussed in Part I and to prepare the ground for the proposal that will be discussed in chapter 8 of Part III.

Finally, chapter 7 and 8 compose the third part of the thesis. Such chapters are both focused on how big data can be used, in the social sciences, to obtain evidence of causality that goes beyond correlations. Chapter 7 explores the existing literature on constitutive mechanisms to discuss how constitutive mechanisms are conceptualised and how big data might help to study them. Chapter 8 starts by observing how data have been used in scientific research to trace etiological mechanisms, and then provides a new approach to collect evidence of etiological mechanisms through sociomarkers, the social version of biomarkers.

Overall, the thesis offers three main take-home messages: first, regardless of how sophisticated they are, causal data-driven methods can suffer from bias and their results are in general insufficient to establish causation. To properly establish causation, it is necessary to use both evidence of a correlation between the cause and the effect, and evidence of the mechanism responsible for such a correlation.

Second, the need for mechanistic evidence sheds light on one possible use of qualitative ethnographic data, that can be analysed to collect this type of evidence. Furthermore, qualitative ethnographic data, also called thick data, might be used to study different ontological categories such as those associated with local and general causation, and to study social phenomena from different epistemological perspectives. The importance of such data, therefore, should not be underestimated even in the era of big data.

Third, studying big data does not necessarily mean finding mere correlations between the cause and the effect. Big data can be used to shed light on causal mechanisms, as I argue analysing the use of biomarkers. In the social sciences, the notion of sociomarkers might help researchers to understand how measures obtained from big data can be employed to trace etiological mechanisms. In rare circumstances, furthermore, big data might help to obtain evidence of constitutive mechanisms.

1.3 Methods

My thesis is empirically grounded on the study of existing scientific projects such as Wood's study on the reasons why peasants in El Salvador decided to join rebel movements (2003) (in chapter 4), Garip and Asad's study concerning the phenomenon of Mexico–U.S. migration (2016) (in chapter 5), and in the analyses conducted within the project LIFEPATH's about how socio-economic conditions influence health (in chapter 6). The decision to use real-life scientific examples is consistent with the branch of philosophy of science known by the name of the 'Philosophy-of-science in practice', defined as:

> "philosophy directly engaged with scientific research through interaction with scientists about philosophical problems" (Boumans & Leonelli, 2013, p. 259)

Since the aim of my thesis is to investigate real scientific challenges associated with big data, one of the first methodological considerations I had to address concerned the best strategy to *select* research projects. As Boumans and Leonelli argued, one of the pivotal aspects of the Philosophy-of-science in practice is the interaction with scientists: this idea led me to focus on two aspects that helped me to identify what I argue are good case studies. First, I considered whether the researchers involved in such projects were engaged in conceptual and methodological discussions. Second, I considered whether it was possible to receive feedback on my arguments from researchers involved in those or similar scientific projects.

The former aspect (whether the researchers involved in such projects were engaged in conceptual and methodological discussions) helped me to select, among a list of putative case studies, those in which particular attention was spent on epistemological and conceptual questions.

As for the latter aspect, in one case (the case study explored in detail in chapter 6), I managed to directly interact with the researchers involved in the project. Such interactions were informal discussions about the aim of their project and the methodological and conceptual challenges associated with it, and formal presentations during joint workshops (such as the Causality Workshop at Imperial College London), which allowed me to present my research findings and to get feedback on my interpretations.

In the other cases, I managed to obtain feedback from researchers working in the same fields by publishing my material in scientific journals. The material presented in chapters 4 and 5 was published in one article in the *Journal of Mixed Methods Research*. During the reviewing process I received extensive comments from 3 anonymous reviewers and 2 editors who conduct mixed-methods studies. The material presented in chapter 8, moreover, has been published in Longitudinal and Life Course Studies, and is the result of an interesting discussions with different researchers working both in epidemiology and in the social sciences.

In general, the Philosophy-of-science in practice approach runs the risk of being contaminated with the problem of cherry-picked case studies. Selected case studies might be unrepresentative for different reasons (for instance, only those researchers more interested in conceptual questions might be interested in interacting with philosophers), and this might lead to inferential mistakes if the case studies are taken to be the illustrative cases of general scientific practices.

Although it is extremely difficult (if not impossible) to completely rule out this problem, in my thesis I try to reduce the risk of cherry-picking in two ways. First, I selected different case studies from diverse disciplines in the social sciences. The quantity and variety of cases chosen help me to argue with more conviction that the claims proposed in this thesis can be applied to a broad range of research practices. Second, the claims proposed in the thesis are not proposed as descriptive claims that are applicable to all the possible scientific studies based on big data. Rather, they are proposed as possible (but not unique) ways to overcome some challenges of big data and are claimed to be based on careful considerations based on interesting case studies.

Part I

Methodological debates in the social sciences and philosophy

The emergence of big data has become a central theme in scientific and philosophical discussions. Both inside and outside academia, there has been growing interest in the potential of big data, as illustrated by the recent surge of dedicated funding, policies and journals. The main tenor of the discussion, shared by academics and non-academics, is that big data allow researchers to think and to do science in a new way.

This idea has been sometimes intensified in the claim that, in the era of big data new statistical methods, rather than existing theories or new approaches to theorising, will drive the practice of science (Anderson, 2008; Mayer-Schönberger & Cukier, 2013). According to these authors, the massive amounts of data now available and the sophisticated algorithms that can be used for data analysis allow scientists to do something unprecedented: to use data for discoveries without taking into account any theoretical element about causation. Causal theories are claimed to be redundant for two reasons: first, sound explanations can be developed just examining the abundant correlations that can be found between data, consequently causal theories are no longer a guide for scientific research; second, big data are making scientists move away from aspiring to understand the reasons *why* something is happening. In the era of big data, scientists just aspire to understand *what* is happening:

> "Causality won't be discarded, but it is being knocked off its pedestal as the primary fountain of meaning. Big data turbocharges non-causal analyses, often replacing causal investigations" (Mayer-Schönberger & Cukier, 2013, p. 68)

However influential and exciting, the idea that a data-driven science does not need causal theories and causal investigations has been criticised by scientists and philosophers (Canali, 2016; Floridi, 2012; Kitchin, 2014a; Leonelli, 2014b; Titiunik, 2015), who have cast light on the limitations of big data and on the importance of causal reasoning.

Although these considerations have paved the way for a crucial debate on the use and novelty of big data, more work needs to be done to completely address the questions about big data. In particular, how to use big data in causal studies is still a matter of debate. The overall aim of my thesis is to offer a new answer to this question by examining how big data can be used for causal studies in the social sciences. Part I of the thesis, made of the next three chapters, will provide some clarifications about what problems and solutions have been discussed for causal discoveries in the social sciences and in the philosophy of the social sciences.

Chapter 2 will start by illustrating one of the statistical methods used with big data for causal analysis, Bayesian Networks. This method will be described in detail to show one common way in which data are analysed quantitatively. Next, some of the problems that characterise data-driven studies will be described and applied to Bayesian Networks to clarify how such issues can emerge in real-life data-driven studies. To this, some considerations about the limitations of big data specifically in the social sciences will be added, and the notion of 'thick' data will be introduced. Overall, these discussions will lead to the identification of three key challenges of big data for causal studies in the social sciences: i) how to overcome the limitations of data-driven causal studies, ii) how to understand the role of qualitative, ethnographic data in causal studies based on big data, iii) how to use big data, in the social sciences, to obtain evidence of causality that goes beyond correlations. These challenges will be discussed in detail in section 2.5.

Chapter 3 will provide a solution to the first challenge. It will be argued that evidential pluralism, in the formulation proposed by Russo and Williamson (2007), and now known by the name of the Russo-Williamson thesis, can offer a solution to the problems associated with causal data-driven studies in the social sciences. In order to comply with such a thesis, social scientists need to establish both a correlation between the putative cause and the putative effect, and a causal mechanism linking the cause and the effect and producing the statistical relationship between them. While some philosophers of the social sciences have shown scepticism about the usefulness of the Russo-Williamson thesis, I shall argue that it is not only useful, but also in line with how social scientists establish causation.

Chapter 4 will tackle the second challenge by exploring the advantages of combining big and thick data in causal studies. I shall claim that, since thick data are in general defined

as qualitative data, the advantages of such a combination can be explored through the analysis of mixed methods research, based on the combination of quantitative and qualitative approaches. To begin with, I will show that, in mixed methods research, the aim is often to establish the presence both of a correlation and of a causal mechanism between the putative cause and effect, and that this is in accordance with the Russo-Williamson thesis. In these terms, the use of qualitative data is in general justified by claiming that they allow for the identification of causal mechanisms. Next, I shall argue that qualitative data are also used to problematise quantitative findings: the combination of different types of data is used to mix different ontological categories and different epistemological assumptions. In this way, researchers can obtain a comprehensive understanding of the phenomenon under study. Thick data, I shall conclude, can hence be used both to collect evidence of social mechanisms, and to help researchers obtain a comprehensive understanding of the phenomenon under study.

Overall, the following three chapters will help to clarify how the methodological discussions within the social sciences and philosophy of the social sciences communities can be associated with the challenges brought about by big data.

2 No magical solutions with big data

2.1 Introduction

Uncovering causal relationships is vital in many aspects of our daily life: we want to know what affects our health, what might exert an effect on a particular political system, what causes economic crises and so forth. It follows that causal discovery is a problem of interest in almost all the scientific disciplines. Over the last few years this interest has been associated with the discussions regarding the emergence of 'big data' and the possibility of automated forms of reasoning. From the beginning of this century, indeed, researchers inside and outside academia have experienced what has been called a 'data-deluge'. Everywhere we look, it is clear that the quantity of information available is soaring: to give an example, according to some IBM analysts and to the researchers at the International Data Corporation, the total amount of data in the world is now doubled every two years (Giczi & Szőke, 2018; Guo, 2017).

Researchers from different disciplines have tried to find out the best ways to take advantage of this data deluge. Many have argued that new sophisticated data techniques can be used to infer patterns and regularities in the data (Jensen & Nielsen, 2007; Mazzocchi, 2015; Spirtes et al., 2000). Some philosophers, however, have shown that, in real-life scientific practice, there is no such a thing as a *purely* direct inductive approach (Leonelli, 2014b; Pietsch, 2016). Inferential reasoning from big data, it has been claimed, always requires some theoretical commitments concerning the phenomenon under investigation and the ways in which data are modelled. Many researchers, furthermore, have identified some problems that big data studies might be unable to tackle (see for instance Cartwright, 2007; Floridi, 2012; Hitchcock & Sober, 2004; Kitchin, 2014; Leonelli, 2014b; Raghupathi & Raghupathi, 2014; Titiunik, 2015).

This debate does not represent a sharp division between those who affirm and those who reject the value of inferring patterns from big data. There is a general agreement, indeed, that new statistical methods can make a contribution to the analysis of data. What is seen with scepticism, instead, is the optimism of those who would like to pursue automated forms of scientific reasoning, where the roles played by human activities would be progressively reduced and algorithms would analyse data "better and faster than humans" (Jensen & Nielsen, 2007, p. 24). This position has often been identified with the

expression ‘data-driven science’, according to which inductive and automated forms of reasoning from data should be recognised as a crucial form of scientific inference (Leonelli, 2012a). The scepticism, moreover, is particularly strong when researchers discuss the use of big data for causal studies (Canali, 2016; Clark & Golder, 2015; Kitchin, 2014a; Leonelli, 2014a; Taylor et al., 2008; Titiunik, 2015; Triantafillou et al., 2017). The reason motivating such a reaction is that even though statistical algorithms are becoming more and more sophisticated, there are some difficulties that still today might cause relevant bias in data-driven causal discoveries. The aim of this chapter is to explore such difficulties.

The chapter is structured as follows: in section 2.2 I shall describe Bayesian Networks as an example of machine learning algorithms used to quantitively analyse data. I shall explain why Bayesian Networks can be considered an important statistical approach in the era of big data, and I will show how Bayesian Networks work and the assumptions built into such an approach. In section 2.3 I shall discuss some of the main problems that might threaten the correctness of machine learning’s analyses and for which the solution is not always obvious. More specifically, I will show how machine learning algorithms, in particular circumstances, can generate results that suffer from the so-called Simpson’s paradox, how the high dimensionality of data can have an impact on the capacity to build causal models, and how the quality of data and meta-data remains a challenge even with the most sophisticated algorithms. While such problems are not specific to Bayesian Networks, I shall discuss their impact on Bayesian Networks’ analyses to clarify how they can lead to biased results in real-life situations. Section 2.4 will describe another problem related to big data studies, namely the gap between big, often quantitative, data, and contextual ethnographic data and will show how social scientists are expressing their worries about the limited importance given to the latter type of data. In section 2.5, then, three different challenges of big data will be discussed: i) how to overcome the limitations of data-driven causal studies, ii) how to understand the role of qualitative, ethnographic data in causal studies based on big data, iii) how to use big data, in the social sciences, to obtain evidence of causality that goes beyond correlations. Section 2.6 will conclude the chapter by arguing that new philosophical investigations can help to overcome such challenges.

Overall, the discussion of this chapter will provide some evidence in support of the idea that there are no ‘magical solutions’ with big data: machine learning algorithms can

enhance scientists' capacity both to extract insights from data and to obtain information concerning specific causal relationships, but bias as well as lack of contextual knowledge might still threaten data-driven results. This observation will pave the way for further philosophical investigations to tackle the challenges proposed at the end of the chapter.

2.2 Bayesian Networks

In the last thirty years several techniques have been proposed to infer causal relationships from observational data, and a new interdisciplinary field, named "causal modelling", has become particularly important in the study of causal inference. The end of the twentieth century, indeed, was characterised by an explosion of works on this topic, and several important contributions emerged both from fields such as statistics, philosophy and computer science, and from more subject-specific areas like epidemiology and econometrics (for more information see Hitchcock, 2009).

Several sophisticated statistical approaches have been developed to analyse causal relationships in large datasets. Just to mention some of them, data are now analysed through the Principal Component Analysis, Diffusion Maps, Regression Analysis, the Granger Causality Approach, the Structural Equation Modelling and Bayesian Networks. In particular, the set of approaches known by the name of Bayesian Networks (BNs), given its capacity to handle uncertainty in complex domains[3] (Constantinou & Fenton, 2018) and the possibility of combining BNs with prior knowledge (Heckerman et al., 1995), has become a popular approach in big data studies. New big data projects based on BNs strategies have been funded both in Europe and the U.S. (see for instance the project European funded project BAYES_KNOWLEDGE and the American project Learning Big Bayesian Networks), and many researchers have explored the applications of BNs machine learning algorithms to big data (see for instance Gogoshin et al., 2017; Tang et al., 2016; Wang et al., 2014).

BNs were first developed in the 1980s, but gained popularity from the 1990s and continued to flourish to the present day thanks to some interdisciplinary discussions involving statisticians, computer scientists and philosophers (Korb & Nicholson, 2004; Neapolitan, 2004; Pearl, 1998, 2000; Spirtes et al., 2000; Williamson, 2005, 2010). A

[3] In BNs the uncertainty concerning the model is represented by a probability distribution over the possible DAGs.

regular Bayesian Network consists of three components: a finite set V = {V1, ..., Vn} of variables, a directed acyclic graph representing the probability distribution of each variable ∈ V conditional on its parents, and the Markov Condition.

A directed acyclic graph (DAG) consists of a set V = {V1, ..., Vn} of variables, that are used as nodes, and directed edges connecting some pairs of nodes, often called arrows. The graph is acyclic because there are no directed cycles standing for mutual relationships (for instance A → B, B → A), like the one represented in Figure 1.

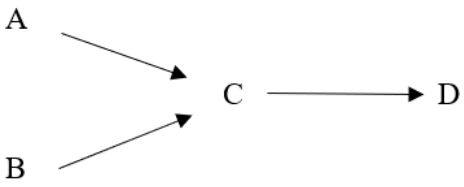

Figure 1. A directed acyclic graph (DAG).

In BNs, researchers often make free use of the terminology of kinship to denote the different relationships between the vertices. In Figure 1, for instance, C has two parents (A and B, which are its direct causes), while D has 3 ancestors (A, B, C) and no children (therefore no descendants).

The graph is linked to the probability distribution of each variable ∈ V conditional on its parents by means of the so-called Markov Condition, according to which any variable ∈ V is probabilistically independent of all other variables apart from its descendants, conditional on its parents.

Often, graphs in BNs are given a causal interpretation and are called *causal graphs*. The key assumption behind this causal interpretation associates causal graphs with probability distributions and is formalised in the so-called Causal Markov Condition:

> **Causal Markov Condition**: Each variable is probabilistically independent of its non-effects, given its direct causes

The Causal Markov Condition is part of the definition of causal BNs, and claims that, if we know the value of the *direct causes*, we should not consider the indirect causes. For instance, in Figure 1, D only depends on C, and it is causally independent of A and B once we conditionalise on its parent C by holding its value fixed. This condition must

always hold if the causal net is to be a BN model: if the Causal Markov Condition is not satisfied, the formalism cannot be interpreted as a Causal Bayesian Network.

The Causal Markov Condition specifies how to obtain probabilistic independencies between variables. However, inferring causation from statistical data in a correct way requires also some knowledge about dependence relationships between variables. This is why, in causal BNs, researchers have proposed the Faithfulness Condition:

> **Faithfulness Condition**: All the probabilistic dependencies and independencies among V are derivable from the causal graph through the Causal Markov Condition

According to the Faithfulness Condition, all probabilistic dependencies and independencies characterising the set of variables under study should be represented in the DAG, otherwise the distribution would be unfaithful to the DAG.

It is important to specify that such a condition, unlike the Causal Markov Condition, does not have to necessarily hold in BNs: even if the Faithfulness Condition is not satisfied, the causal net can still be considered a Bayesian net. If the Faithfulness Condition is satisfied, however, researchers know that the dataset under study has exactly and only the probabilistic independencies shown in the DAG. In other words, the use of this condition has an important inferential advantage: through it, it is possible to say that the model generated from the data implies exactly the dependence and independence relations that are characterising the dataset under study. When the Faithfulness Condition is not satisfied, on the other hand, we cannot be sure that the model obtained represents all and only the dependencies that really characterise the set of variables. The consequence is that, if the Faithfulness Condition does not hold, we might be more likely to draw wrong conclusions about the causal relationships in our dataset.

In the BNs literature, a stock example where the Faithfulness Condition is not satisfied is the DAG representing the relationships between the use of contraceptive pills, thrombosis and pregnancy (Hesslow, 1976, p. 291). On the one hand, contraceptive pills lower the chances of pregnancy; on the other hand, both contraceptive pills and pregnancy increase the probability of thrombosis, as illustrated in Figure 2.

In this case, we have two paths. The first one goes from contraceptive pills to the reduction of the possibility of pregnancy and, hence, the reduction of the risk of

thrombosis (pills → $^{-}$ pregnancy →$^{+}$ thrombosis). The second one goes from contraceptive pills to the risk of thrombosis (pills → $^{+}$ thrombosis).

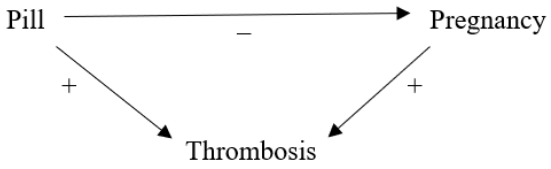

Figure 2. A directed acyclic graph (DAG) showing the causal relationships between contraceptive pill, pregnancy and thrombosis.

Let us suppose that these two paths are perfectly balanced: the presence of the first path might hide the operation of the second path, and vice versa. The two paths would cancel each other, with the consequence that any probabilistic dependence between contraceptive pills and thrombosis would vanish (i.e. there would not be a probabilistic dependence between pills and thrombosis). Yet, since the DAG shows a path, the probabilistic distribution would be unfaithful to the DAG.

In the statistical and philosophical literature, this and similar problems have been discussed in relation to data-driven methods such as BNs. Originally, BNs were conceived of as belief networks, and the probabilistic relationships were assumed to represent the appropriate degrees of belief for an agent given the data (Pearl, 1988). When BNs are used for machine learning algorithms, however, the probabilistic relationships are in general simply induced by the frequencies observed in the dataset. The DAGs used to infer causal relationships, consequently, are the result of data-driven analyses. In what follows, I shall discuss some of the general limitations that can undermine the usefulness of data-driven studies. Furthermore, to illustrate how these limitations can pose a threat to causal models in specific cases, I shall illustrate the ways in which they can cause bias in BNs.

2.3 Problems with causal modelling

2.3.1 Simpson's paradox

Suppose you are a doctor reading a paper on a promising new treatment that seems to be very effective. You are excited by the discovery, and you look up the data from the trials to learn more about that. You look at the data concerning male patients and find out that

actually there is no correlation between treatment and recovery. “Well”, you will probably think, “this drug must work very well for female patients”. Then you look at the data about women, and you discover that, also in this case, the treatment does not seem to be effective. How is this possible? How can a drug be bad both for male and female patients, but at the same time good for people? It is clearly problematic if two different analyses induce you to take two opposite actions (give or not the treatment) based on the same data.

This problem is known by the name of ‘Simpson’s paradox’ and was discovered by the statistician Edward Simpson (1951), from whom it took its name. It is based on a statistical phenomenon according to which, in some datasets, subgroups with a common trend (for instance a negative trend) show the reverse trend (a positive trend) when they are aggregated. To illustrate this paradox, let us consider the example proposed by Meek and Glymour (1994, p. 1012). Suppose the data were collected from a study involving 990 patients who divided themselves into a control group (610 patients) and a treatment group (380 patients). Table 1 shows the number of recoveries in the two groups:

	Control Group	Treatment Group
Alive	260	240
Dead	350	140

Table 1. Survival rates in the control and treatment groups.

By comparing the data about the control group and the treatment group in Table 1, it is possible to conclude that there is a positive association between treatment and survival. Indeed, while in the control group only 260/610 patients survived (43%), in the treatment group the survival rate was 63%.

Suppose that now we look at the same data by distinguishing between male and female patients, as shown in Table 2 and 3.

	Male	
	Control Group	Treatment Group
Alive	160	40

Dead	320	80

Table 2. Survival rates in the male population.

	Female	
	Control Group	Treatment Group
Alive	100	200
Dead	30	60

Table 3. Survival rates in the female population.

In Table 2 it is possible to see that there is no correlation between treatment and recovery among males: in the treatment group 40/120 (33,3%) of patients survived, in the control group 160/480 (33,3%) of patients recovered from the disease. The percentage of recovery, hence, was the same both in the treatment group and in the control group.

In Table 3, furthermore, it is shown that 200/260 (77%) of female patients who were treated survived, the same percentage that was found in the control group, where 100/130 (77%) of the untreated women recovered. Overall, hence, there was no association between treatment and survival among both male and female patients.

There is an explanation for this paradox: when data about women and men were aggregated, the percentage of recovery was calculated by taking into account the weighted average. In other words, the analysis considered how many patients were in each group. Given that women were more likely to recover and, in that specific situation, a larger proportion of women, if compared to the proportion of men, was treated, the results showed that the survival rate in the treatment group was higher than in the control group.

Simpson's paradox is not a rare problem that can be found only in fictional situations. In 1973, the Associate Dean of the graduate school of the University of California Berkeley, by looking at the admission rate data, hypothesised that there was a sex bias in the admission process. The data were indeed shown in a table like Table 4, and it was clear that the denial rate among women was particularly high if compared to the denial rate among men.

Applicants	Admitted	Denied
Female	1494	2827
Male	3738	4704

Table 4. Acceptance rates to the University of California Berkeley in 1973.

Some years later, in a famous paper, Bickel, Hammel and O'Connell (1975) explored the dataset, observing in detail the percentages of admission for each department. It was thanks to the data analysis performed department by department, that they were able to claim that there was no bias towards women in the admission process. Indeed, by considering the admission rate per department they generated a table similar to the following:

Department	Male Applications	Male Admissions	Female Applications	Female Admissions
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	191	28%	393	24%

Table 5. Acceptance rates per department to the University of California Berkeley in 1973.

As shown in Table 5, the problem with the admission rate was that female applicants tended to apply more to departments that were more difficult to get into, as the data about department C clearly illustrate. By considering the data properly partitioned, hence, it appears not only that there was not a bias *against* women, but that there was a small (but statistically significant) bias *in favour* of women.

While the examples described above are two of the most quoted examples of Simpson's paradox, it can be noted that several gender bias studies and analyses of sports' data have provided very nice examples of this phenomenon. For instance, it has been shown that evidence in support of gender bias in science funding is likely to suffer from this paradox (Albers, 2015). From the analysis of basketball's data, in addition, researchers can claim that a team has the worse shooting in each category, and yet the best overall field goal percentage (Ma & Ma, 2011). In baseball, furthermore, a player A can have a higher batting average for three consecutive seasons if compared to another player B, and yet by combining the three seasons and considering how often each player played, the result is that player B has the higher average (Pearl & Mackenzie, 2018). More examples of this paradox, both in the natural and the social sciences, can also be found in Selvitella's recent paper entitled 'The ubiquity of the Simpson's Paradox' (2017).

It is now possible to examine how Simpson's paradox can cause bias in the specific case of BNs. Let us consider again the treatment example discussed above: the treatment apparently worked for the whole population but was ineffective when analysed in the male and female subpopulations. By analysing the overall population divided into a treatment and a control group, researchers would consider just two variables, T (treatment) and R (recovery), and would obtain a causal DAG showing an arrow between T (treatment) and R (recovery), as illustrated in Figure 3(a). This would lead them to give a causal interpretation to the arrow and to establish that the treatment causes recovery. The probabilistic dependence between T and R, however, is actually caused by the common cause G, gender, as illustrated in Figure 3(b). Due to the way in which the control and treatment groups are partitioned, the variable G might not be added in the dataset, with the consequence that the subset of variables analysed might lead scientists to infer a causal relationship that does not exist in reality.

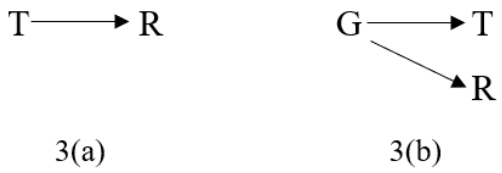

Figure 3. The possible DAGs between treatment, recovery and gender. 3(a) illustrates the probabilistic dependence between T (treatment) and R (recovery), if G (gender) is not considered; 3(b) illustrates the

probabilistic dependencies between G (gender), T (treatment) and R (recovery).

Someone might argue that Simpson's paradox causes problems especially when researchers analyse small datasets, given that there is a major risk of not considering relevant variables. In big data studies, on the contrary, scientists are able to analyse simultaneously numerous variables, therefore the risk of ignoring important subpopulations could be drastically reduced. This consideration is particularly relevant in specific cases such as medical studies, where it is known that patients' characteristics such as age and gender can influence physical conditions and recovery, and where such variables are generally available.

In other cases, however, it might be difficult to imagine how to partition data or, in other words, what subpopulations could be relevant to the research question. Human behaviour studies, for instance, in general rely on heterogeneous observational data generated by subpopulations of different sizes. Such behavioural data can now be collected quite easily (for instance, data about consumer behaviours are now collected through the Internet and by platforms such as Amazon, while travel behaviours can be monitored through the use of geolocated data), but at the individual level they often appear very sparse and noisy. It is for this reason that researchers might prefer to aggregate data and to analyse behaviours at the population level. Some examples are the use of population-level data for online activities (such as shopping online) and for diurnal and seasonal mood rhythms (the changes in mood through the day or in different seasons) (Lerman, 2018).

The problem that emerges, however, is that when hundreds or thousands of behavioural data are aggregated, it is not easy to identify the relevant subpopulations. In an online social network, for instance, a large number of variables such as sex, age, occupation, income, education, nationality, average online activities, the number of words per post, and the number of followers might be relevant in order to rule out the possibility of confounders. Some of these factors could be difficult to measure. Furthermore, such variables might be collected in different datasets, and getting access to them could be difficult or very expensive. In a similar situation, what to include or not would finally depend on the researchers' decisions (with the risk of using inappropriate measures or ignoring important factors). Finally, even when all the possible relevant variables are available and easily measurable, scientists are not immune to inferential errors: too many

variables might indeed cause the curse of dimensionality, which is what I explore in the next section.

2.3.2 High dimensional datasets

Most of the datasets created in recent years are characterised by high dimensionality, an expression that refers to the situation in which a dataset is characterised by a large number of features. When the number of features is larger than the number of observations, datasets are characterised by the 'curse of dimensionality' (Bühlmann & Geer, 2011, p. 1). Suppose we have a dataset containing 100 images (observations) with a high resolution. Each image is composed of thousands or even millions of pixels (in the case of Flickr, for instance, each image can contain 2048x2048 pixels), and each pixel within the image can be understood as a feature of the image. This means that we have a huge number of pixels (features) within each image (observation), and the total number of images is very small if compared to the total number of pixels. The database is hence characterised by high dimensionality. This kind of situation can be found in several contexts: financial datasets, for instance, can contain observations measured daily, hourly, even every minute or second, and for each time slice such datasets have hundreds of features. Similarly, social networks allow for the collection of hundreds of features for each individual using them. Scientific advancements in medicine, finally, have led to the current situation in which thousands of features are collected for each individual, instead of having, as in the past, a big sample with a low dimension (Zeng et al., 2016).

It would be reasonable to hold the intuition that the analysis of high dimensional datasets through machine learning algorithms should offer more accurate results. In reality, however, it is very common that the opposite happens. When datasets are characterised by the curse of dimensionality, as the number of variables increases, the number of plausible combinations of variables explodes exponentially. This happens because a fixed number of data points become increasingly 'sparse' if the dimensionality is increased. Figure 4, for instance, represents two cases, one in which we have two dimensions, one in which we have three dimensions: when the dimensionality increases, also the number of sides increases. Consequently, while the data points in two dimensions are sufficient to find a pattern, the same data points in three dimensions are too sparse to enable researchers to select one pattern among all other possible patterns.

Let us consider the case of BNs: as the number of nodes increases, the size of the search space of the relationships between causal nodes grows exponentially in dimensions. If the number of data points, furthermore, is very small if compared to the number of nodes, the confidence in the probability dependences represented in the DAG becomes very low: in other words, we cannot be sure that the complex DAG we are observing represents the correct causal relationships between the nodes under study.

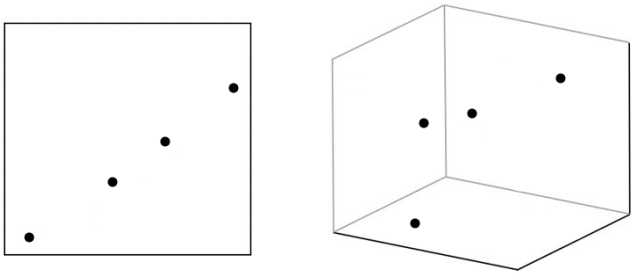

Figure 4. Data points represented in two-dimension space and in three-dimension space.

A phenomenon closely related to the curse of dimensionality is the phenomenon known by the name of 'overfitting'. When a new algorithm is developed, scientists use training data to train it. Overfitting occurs when the model produced by the algorithm is very accurate on the training data but is much less accurate on the real data. In Figure 5, for example, the curved line best follows the training data if compared to the dashed line, but it might be too dependent on such data. The algorithm producing the curved line, therefore, might have a higher error rate when used with new data. This happens because the set of training data points is too small if compared to the variables analysed, and the algorithm runs the risk of modelling not only the general patterns in the data, but also the idiosyncrasies of that specific data set that are unlikely to recur in further data (Hitchcock & Sober, 2004). This, hence, would cause the model to poorly perform when new data are analysed. For instance, in BNs overfitting typically takes the form of a fully linked DAGs, where the number of arrows is so high that all the nodes of the DAGs are linked to each other.

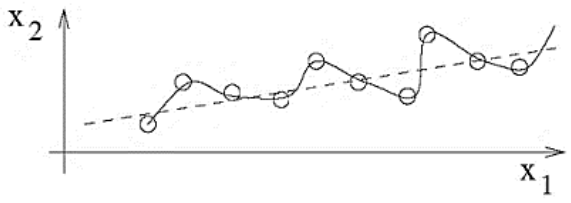

Figure 5. An example of overfitting (Silipo, 2007, p. 287).

Such considerations bring to the fore what Floridi has called the 'small pattern' problem of big data (Floridi, 2012): the vast amount of information available entails that it can be more difficult than in the past to spot where the new relevant patterns lie. Given the growing number of dimensions available for each research question, for instance, how to select the dimensions that can really help both to uncover important patterns and to avoid the curse of dimensionality? To give an example, thousands of genes (features or parameters) are monitored for each person, but for a specific disease only a fraction of them are biologically relevant. It is crucial, hence, to identify those parameters (i.e. dimensions) that can be used for prediction and diagnosis.

In the statistical and computer science literature several methods have been proposed to avoid the curse of dimensionality. One of the most common strategies is to reduce the dimensions of the datasets. Feature transformation techniques, for instance, are data pre-processing methods that allow for the transformation of the original dimensions of a data set into a more compact set of features or parameters, maintaining at the same time as much information as possible (a typical example is Principal Component Analysis). Feature selection algorithms, on the other hand, are based on an alternative approach to dimension reduction and select among the available dimensions the most relevant subset of parameters. Finally, wrapper algorithms use the classification accuracy of some classifier as a criterion to reduce dimensionality (for more details see Cunningham, 2008; Guyon & Elisseeff, 2003).

To avoid some risks associated with the curse of dimensionality and to select the right 'small patterns', furthermore, in general researchers work with supervised machine learning algorithms, trained with labelled training datasets containing input objects and the desired output value attached to such objects. To give a simple example, an algorithm can learn to classify animals (such as dogs and cats) after being trained on a dataset containing photos properly labelled with the corresponding species and some identifying features.

There are cases where researchers use unsupervised algorithms, that are thought to be able to recognise processes and patterns without any human guidance. However, in order to improve the reliability of data-driven results, in most of the cases researchers prefer to train the algorithm. In some situations, moreover, scientists' knowledge is used also to

guide algorithms' decisions after the training. As described above, for instance, in genetic studies each observation (such as each mRNA sample) can have hundreds or thousands of features (genes). Let us suppose researchers know that the expression of some specific genes depends on environmental conditions, but they are not sure about what gene expression varies under which conditions. To identify the most useful features, researchers are working on statistical techniques that can incorporate external information (for instance, a candidate list of relevant genes could already exist before and researchers could decide to give to such candidates a higher priority) (Liu et al., 2016).

Despite the use both of supervised algorithms and of techniques to reduce the dataset's dimensions, however, it is still difficult to rule out all the problems associated with the curse of dimensionality and overfitting. Racial algorithmic bias can offer some illustrating examples to understand the consequences of such problems.

A case in point is that facial recognition algorithms are very likely to discriminate minorities, and that this problem is caused by the high dimensionality of the training datasets. This phenomenon is due to the fact that in the Western society the data used to train algorithms often over-represent one population (for instance white people) and under-represent minorities (like Afro-American people). Due to this under-representation, the number of variables examined can be larger than the number of samples in the data. As a consequence, algorithms might start to overfit the data associated with the under-represented population, modelling also idiosyncrasies, as illustrated above in Figure 5. This can cause critical situations: to give an example, in some cases facial recognition algorithms are not able to categorise members of the under-represented population as 'persons' because they do not have all the idiosyncrasies required by the algorithms to be classified in that way. This has led to controversial situations where recognition algorithms categorised members of the under-represented population as animals.

2.3.3 Data and meta-data quality

Imagine you have enormous amounts of data about what people search on the Internet. You assume that a relevant proportion of those people with flu-like symptoms would go to the web to look for information on what to do. Thanks to your data, you can select the search terms that people are more likely to use, and then build a sophisticated algorithm

to track where and with what frequency such search terms are used in a specific geographical area.

This is exactly what happened when the Google Flu Trend project was developed in 2008. Google engineers used Google data to identify the search terms more likely to be correlated to flu by considering both the search queries trends from 2003 to 2007, and the real local patterns regarding the number of people with influenza-like illnesses in those 4 years. In this way, they selected a list of 45 terms that people searched on the Internet during the flu peak weeks. The Google Flu Trend algorithm was then developed to identify when and where such terms were searched through Google[4]. As many studies have already reported (Butler, 2013; Lazer et al., 2014), the project was far from being a success. In 2009, the algorithm badly underestimated the Influenza Virus A (H1N1) Pandemic (Cook et al., 2011), while in 2010 it was demonstrated that the information coming from that algorithm had the same accuracy than the results obtained with a fairly simple projection forward based on traditional data (Lazer et al., 2014, p. 1203).

As anyone can guess, the main problems of Google Flu Trend had to do with the initial assumption and the data used. It is indeed questionable whether so many people actually look for their symptoms and for information about influenza virus on the Internet, and it is even more questionable whether the quality of the data used was sufficiently high. These two problems are strictly linked one to the other: indeed, if only a small proportion of people looked for information about flu symptoms and remedies on Google, the data could be unrepresentative of the population under study. For instance, it might happen that only specific subpopulations (such as parents of new-born babies) search for flu symptoms on the Internet, while other members of the population (such as those aged 65 and over) do not rely on the Internet but prefer to contact their doctors.

After the failure of the program, furthermore, it turned out that the data selected and analysed through the algorithms had several problems. First, some scientists found out that Internet search behaviours changed during the Influenza Virus A (H1N1) Pandemic: the complications of that flu were different from the general influenza complications, consequently people searched for terms not included in the original Google Flu Trend list (Cook et al., 2011) and these data were not taken into account by Google Flu Trend. Second, the data-generating process used by Google was not stable: both this process and

[4] Google never released the algorithm and the raw data behind Google Flu Trends.

the way in which Google recommended some searches were changed over time to improve Google's service. The direct consequences were that data collected through different algorithms were not comparable, and that their reliability was affected by changes in search recommendations (Lazer et al., 2014).

The description of this case helps me to highlight another problem for which, I argue, there are no magical solutions: even with the best algorithms, data need to be properly cleaned and curators need to organise them by attaching additional information known as meta-data. Such meta-data, in general, consist of detailed information concerning the origin of the data, the methods used to collect them, the research aims and the protocols used during the collection (Leonelli, 2014a; Taylor et al., 2008). If something goes wrong, if data are not sufficiently clean or if meta-data are not sufficiently detailed, even the best algorithms will not produce reliable results.

In the Google Flu Trend example, for instance, two problems could have caused biased results. On the one hand, all or some data could have been biased. The quality of the data used by Google Flu Trend, consequently, would have been insufficient to ensure the reliability of the results. For example, the first algorithm used by Google might have generated low-quality data before being improved. To avoid this problem, researchers should have used only those data collected through the second (improved) algorithm.

On the other hand, the problem could have been caused by the lack of information concerning the way in which data were collected. In such a case, the insufficient information 'attached to' the data would have been responsible for the failure of the program. Scientists could have avoided this problem by adding a new variable in the dataset to distinguish between the subgroup of data generated by means of the first algorithm and those data collected through the second algorithm, or curators might have properly used meta-data to put the dataset in order.

In the literature, one of the earlier discussions on the core innovations brought in by the emergence of big data has been focused on the idea of messiness. According to some authors (Leonelli, 2014b; Mayer-Schönberger & Cukier, 2013) we should not underestimate the fact that big data are very messy, vary in quality and can be stored in so many different datasets. This means that, in order to assemble big data and make it possible to integrate them, it is in general required a huge amount of manual labour. Leonelli (2012b, 2014b) pointed out the importance of these human decisions by

describing the different problems that database curators can encounter when new biological databases are created. In other scientific fields, similar discussions emerged when the researchers responsible for new databases started to clarify how data are repaired, curated and documented. For instance, on the website of the UK Data Archive (http://www.data-archive.ac.uk) it is possible to read how curators clean and prepare the new data that will be added to the dataset.

Curators spend a considerable amount of time on adding meta-data that are made available to datasets users. The assumption behind this kind of activity is that researchers need some information about their data before analysing them through machine learning algorithms. Meta-data can help researchers to avoid mixing data obtained in incompatible ways, like in the Google Flu Trend case, and can also allow researchers to assess the quality and the suitability of the data in relation to their specific research questions. For instance, researchers might decide to conduct new studies because they think that the method used to collect data is not sufficiently reliable, or might choose particular data because the research goal for which they were collected is very similar to the research question they are trying to answer.

Adding meta-data, however, is not a trivial task. Due to the lack of standard terminology to describe how data are collected, capturing the relevant aspects of data production can be remarkably challenging. For instance, some experimental practices or algorithms' characteristics might not be intelligible to researchers coming from different fields, or some pieces of information regarding the data collection process might be omitted even if, for certain purposes, they might be vital parameters (Leonelli, 2014a). Consequently, it is not rare to have datasets whose meta-data are partly incomplete, from which it follows that it is not rare to have data studies based on unsuitable data. This is one of the main reasons why, even with big data and machine learning algorithms, data quality can still threaten the possibility of good statistical results.

Some examples can help to clarify this problem. In a famous work, Petricoin et al. (2002), reported having found a method to distinguish between serum samples from women with ovarian cancer, serum samples from healthy women and samples from women with a benign ovarian cancer. The algorithm used for the diagnosis was trained with molecular data obtained through mass spectrometry, called spectra. During the training it identified 50 normal spectra and 50 cancer spectra, then it predicted 116 spectra, identifying 47 out

of 50 normal spectra, all the 50 cancer spectra and all the 16 benign spectra as 'other'. The result was so exciting that the U.S. Congress was urged to increase the funding to develop the diagnostic test (Check, 2004). However, when Baggerly, Morris and Coombes (2004) tried to replicate the study, they did not obtain the same exciting results. It was finally discovered that the three types of spectra used by Petricoin et al. had been pre-processed differently, and such differences (that were not associated with the biology of cancer) were those recognised by the algorithm.

Another case regards some studies that used racial data from the National Health Index to explore possible causal linkages to mortality. It was discovered that, in many occasions, the officials who completed the death certificate determined the racial status based on their own judgments rather than asking to the members of the family, and that relevant percentages of American Indians, Asian Pacific Islanders and Hispanics were identified into another racial category (Williams, 1999). The direct consequence of this problem was that any study based on such data could contain bias.

Problems in data quality can have dramatic, even expensive consequences: in 1999 NASA lost a Mars orbiter costing $125 million because engineers combined data measured with the metric system of millimetres and meters, with data measured with the Imperial system of inches, feet and pounds (De Veaux & Hand, 2005, p. 235).

After the examples described above, it is easy to imagine a situation in which BNs, due to the bad quality of data or meta-data, could produce causal graphs with distorted relationships. Like in the other cases, the solution is not straightforward: the processes of data production and curation require careful examinations in order to avoid substantial mistakes like those described in this section.

2.4 Big and thick data

While the problems described above can be found in numerous disciplines, there is another aspect of the discussion about the role of big data for causal discoveries that deserves attention and that regards in particular the social sciences.

Many social scientists have claimed that neither the increasing number of variables nor the existing computational techniques can solve the basic problems characterising causal inference. Such claims, however, have not been supported only by methodological discussions about problems like Simpson's paradox. The main tenor of this literature is

that big data cannot provide the contextual understanding that is required to causally investigate a social phenomenon.

In 2014, for instance, political scientists extensively discussed this limitations of data-driven methods in a roundtable entitled ‘Big Data, Causal Inference, and Formal Theory: Contradictory Trends in Political Science?’ at the Annual Meeting of the Midwest Political Science Association. The discussion continued during a symposium with the same title, during which Titiunik (2015) argued that, to identify causal relationships, it is in general required to have both substantial background knowledge and contextual insights into the phenomenon under study. During the symposium, furthermore, Monroe et al. (2015) claimed that it is the responsibility of social scientists to take a central role in the big data studies, shaping the questions they want to answer through big data according to their contextual knowledge and establishing whether the answer can be convincing or not.

Similar considerations were proposed also by other researchers who claimed that, if data are not embedded in contextual knowledge, data-driven studies are likely to obtain ‘anaemic’ or unhelpful results (Bornakke & Due, 2018; Kitchin, 2014b). An example used to justify this claim is the way in which big data are analysed in smart cities studies. Often, it is argued, smart cities data-driven studies ignore the social theories and causal hypotheses proposed in the literature, with the consequence that data-driven results are in general reductionist and do not take into account the cultural and political dimensions that can influence certain individual behaviours (Kitchin, 2014b).

According to some social scientists, finally, traditionally social sciences have given priority to the identification of causal mechanisms, while data-driven studies’ interest is in the recognition of stable associations and correlations, with the consequence that big data are now challenging the social sciences about the relevance of mechanistic discoveries (Ashworth et al., 2015; Iwashyna & Liu, 2014).

What emerges from this overview is a general worry about the types of data used in big data research, and the approach employed to study such data. To balance this tendency towards data-driven studies, some researchers both inside and outside academia have developed the notion of ‘thick data’.

2.4.1 Thick descriptions, thick data and thick concepts

The idea of 'thick data' emerged from the idea of 'thick descriptions', that goes back to Gilbert Ryle (1968) and Clifford Geertz (1973). Ryle argued that thick descriptions do not illustrate human activities just as bodily motions, but characterise them as activities with an intentional dimension. The best way to clarify this idea is to think about some of the examples proposed by Ryle. Let us suppose we see different boys winking for different reasons: the first boy winks involuntarily, the second one winks conspiratorially to a friend, and does it in a slow and conspicuous way. Finally, the third boy imitates the second boy to entertain his cronies: he winks awkwardly, but he is not himself clumsy. In these three cases, the same 'thin' description, 'the boy winks', can be applied. Nevertheless, if we want to distinguish between the three actions, we need to provide thick descriptions.

Such descriptions were not conceptualised by Ryle as a list of different activities performed by the person. To give an example, the thick description of what the second boy does could not be a list of five separate things such as i) winking intentionally, ii) to a specific person, iii) with the aim to impart a message, iv) according to a common code, v) without the other friends being aware of it. Rather than being a set of different things, the activity performed by the second boy is understood by Ryle as a complex thing, that can be successfully performed only if all these aspects are successfully carried out. The difference between the simple activity of the first boy, and the complex activity of the second one is this success-failure condition (the second boy can wink successfully only if all those activities are fulfilled). Such a condition, finally, is what we should make explicit in our thick description. Geertz followed Ryle's proposal, and argued that the aim of anthropology is to explain cultures and phenomena through thick descriptions (1973). Factual (thin) descriptions, Geertz argued, are not only insufficient to understand a culture, but can also be misleading: this is why anthropologists develop thick descriptions that can incorporate cultural information.

Recently, these considerations have been used to discuss the notion of 'thick data'[5]. Thick data are in general described as ethnographic data characterised by contextual complexity

[5] It is interesting to observe that the notion of 'thick descriptions' proposed by Ryle and Geertz did not inspire only the idea of 'thick data'. After Ryle's and Geertz's works, the idea of 'thick descriptions' was used to develop the notion of 'thick concepts' (Williams, 1985). The first proponent of this notion, Williams, argued that thick concepts are both world-guided and, due to the evaluation attached to them, action-guiding: their uses and applications are guided by people's experiences and, at the same time, they

which help researchers to recognise not only what people do, but also the reasons behind their activities (Bornakke & Due, 2018, p. 2). The notion of 'thick data' has been proposed in contrast with the notion of big data: thick data, it is claimed, are in general collected by real individuals from small samples and are not produced through automated processes and characterised by high velocity (Blok & Pedersen, 2014; Bornakke & Due, 2018; Latzko-Toth et al., 2016).

In addition, although the differences between big and thick data have not yet been explored in detail, one underlying assumption seems to be that thick data are mostly (if not only) qualitative. Onwuegbuzie and Leech (2007, p. 244), for instance, claimed that in qualitative research, new credibility can be gained by collecting thick data sufficiently detailed to enhance our understanding of certain phenomena. Blok and Pedersen (2014, p. 3), furthermore, argued that big and thick data need to be combined in order to fill the gap between 'hard' quantitative evidence and 'soft' qualitative data. This link between qualitative and thick data can be partly explained by the fact that the same notion of thick description was developed to explain the insights that can be obtained through ethnographic and qualitative methods such as participant interviews and field observations (Geertz, 1973; Ryle, 1968). It might be argued, furthermore, that dense, contextual information would be hardly contained in quantitative data.

While in the introduction of this thesis I have observed that variety is one of the features that, according to the literature, characterise big data, it is worth noting that the type of unstructured data in general described in the literature are images, videos, audios and social media texts that are collected automatically with high frequency, with no need for qualitative social scientists. Thick data, instead, are described as contextual data produced by human observers embedded in particular contexts.

To clarify what thick data are, let us consider some cases described in the literature. The first example of thick data discussed by social scientists is the example provided by Wang (2013), an ethnographer who worked for Nokia in 2009. During her period at Nokia, she conducted ethnographic works in China and she collected evidence supporting the idea that low-income Chinese consumers had begun to take an interest in smartphones, and

provide reasons for or against certain types of actions. In general, the immediate examples of thick concepts we imagine are ethical concepts such as courage and promise. It might be argued, however, that the notion of thick concepts can be applied also to other categories of concepts, such as to the aesthetic concepts (for instance elegant and gaudy) and epistemic concepts (like wise and relevant) (for more discussions see Kirchin, 2013).

would have willingly paid more than in the past for this new type of mobile phones. The conclusion of her work was that Nokia needed to change its business model: rather than producing niche products for high-income users, Nokia should have started to make affordable smartphones for low-income consumers. When it was informed, however, Nokia did compare the ethnographic evidence coming from the small sample studied by Wang with the quantitative data collected from millions of users, and concluded that it was better to ignore Wang's observations because in the existing datasets there were no signs of such a trend. This surprised Wang and decisively influenced the future of Nokia given that, as we now know, Chinese smartphones sales started to grow exponentially some years later (Ma, 2017, p. xi).

Another interesting study based on big and thick data can be found in Denmark, where the Faculty of Social Sciences at the University of Copenhagen and the Department of Informatics and Mathematical Modelling at the Danish Technical University have started a joint program called 'The Copenhagen Social Networks Study'. Statisticians, computer scientists and social scientists have examined together data obtained by continuously recording social interactions among an entire university class made of more than 1000 students. Using smart devices collecting Bluetooth data, GPS data and social networks data, researchers obtained a large dataset that was combined with thick observational data collected by an anthropologist 'embedded' in the class for an entire year. Big and thick data together, according to the principal investigators, offer the foundations for what they call 'complementary social science' (Blok & Pedersen, 2014, pp. 3–4).

The discussions on thick data, in general, are in accordance with what proposed by The Copenhagen Social Networks Study researchers. There is indeed a shared assumption that thick and big data have complementary strengths and weaknesses. To give some example, Bornakke and Due (2018, p. 3) and Latzko-Toth, Bonneau, and Millette (2016, p. 201) argued for the complementary nature of thick and big data, and recognised that they are far from being the first to assert it. Despite the general agreement, the advantages associated with the combination of big and thick data are still partly unexplored. More specifically, how to use thick data in causal studies based on big data is still a challenge.

2.5 Three challenges of big data

The problems described above lead to three different challenges associated with causal studies based on big data in the social sciences.

The first challenge is *how to overcome the limitations of data-driven causal studies.* In section 2.3 I have described some of the issues that still threaten data-driven results: Simpson's paradox, the curse of dimensionality and overfitting, the presence of low-quality data. While numerous researchers are trying to develop a solution to such problems, a general agreement about the strategy to tackle them has not been found yet.

The second challenge is *how to understand the role of thick data in causal studies based on big data*. This challenge is particularly relevant nowadays, since some researchers show hesitation towards the emergence of computational research programmes in the social sciences and try to defend the vital role of thick data. This challenge resembles the more general challenge concerning how, in causal studies, qualitative data can be used in combination with quantitative data to produce rich results. While the challenge concerning qualitative and quantitative data has been recently discussed within the mixed methods research community (a community of social scientists sharing the idea that the best way to conduct social research is by mixing quantitative and qualitative approaches), how the study of thick can be combined with big data to improve causal studies is still unexplored.

The third challenge is *how to use big data, in the social sciences, to obtain evidence of causality that goes beyond correlations*. In general, big data are analysed through statistical methods to identify correlations between variables. It might be possible, however, to use big data to obtain mechanistic evidence. This possibility would allow researchers to study big data to obtain different forms of evidence in support of a causal claim. Until now, this possibility has not been directly explored in the traditional social sciences literature, where discussions on big data can be generally identified as statistical debates about the use of computational strategies (see for instance Monroe, 2013; Shu-Heng, 2018), and as considerations about the use of big data to obtain social indicators and proxies (see di Bella et al., 2018). In philosophy, instead, some researchers have paved the way for novel discussions about the use of data in causal analyses. In particular, new philosophical discussions have emerged to examine how causal mechanisms can be studied, and how large-scale biological datasets can help to identify evidence of causal processes in social epidemiology (Baumgartner & Casini, 2017; Canali, 2016; Illari & Russo, 2016; Krickel, 2018; Vineis et al., 2017).

These three challenges will guide my dissertation. The solution to the first challenge (how

to overcome the limitations of data-driven causal studies) will be offered in the next chapter. Chapter 3 will consider one epistemological account of causation, known by the name of the Russo-Williamson thesis, developed in philosophy of medicine, and will apply it to the social sciences. I shall argue that this epistemological account can reduce the risks of bias in causal analysis, then I shall explore whether the Russo-Williamson thesis can really be used in the social sciences, and whether it is in line with what some social scientists already do. Chapter 5 and 6, furthermore, will describe two real-life studies that comply with the Russo-Williamson thesis.

In chapter 4, I shall analyse the debates in mixed methods research to identify the specific reasons why qualitative approaches and data are claimed to be important and complementary to quantitative approaches and data in social research. This analysis will be then used to clarify some of the ways through which qualitative thick data, if combined with big data, can improve causal studies in the social sciences. These considerations will be finally used to provide a solution to the second challenge.

Finally, the solution to the third challenge will be articulated in chapters 7 and 8. Chapter 7 will explore whether big data can be used to collect evidence of constitutive mechanisms, while chapter 8 will bring together the social epidemiological and philosophical literature to discuss how big data can help researchers to trace etiological processes.

2.6 Conclusion

Over the last few years, the excitement due to the rise of big data has been viewed with suspicion in philosophy, especially when the discussion was about causal discoveries. Numerous philosophers have indeed highlighted the vital role that researchers still play in the process of big data analysis (Canali, 2016; Floridi, 2012; Leonelli, 2014a). Similar doubts have been expressed in the social sciences, where several researchers have defended the role that thick data can play in the era of big data (Blok & Pedersen, 2014; Bornakke & Due, 2018; Latzko-Toth et al., 2016).

These considerations have led me to identify three challenges that will be addressed in this dissertation: i) how to overcome the limitations of data-driven causal studies, ii) how to understand the role of qualitative, ethnographic data in causal studies based on big data, iii) how to use big data to obtain evidence of causality that goes beyond correlations.

These challenges will be investigated in relation to the social sciences, where big data from social networks, geolocation tools and digitalised activities are radically changing the kinds of information that social scientists can use. To answer the first question, in the following chapter I shall provide a detailed overview of one of the approaches through which researchers can strengthen their confidence in causal models: the approach known by the name of the Russo-Williamson thesis.

3 Taking the Russo-Williamson thesis seriously in the social sciences

3.1 Introduction

From economics to demography and political science, it is in general recognised that statistical models play a crucial role in the social sciences when the aim is to establish causal relationships. This role has become even more prominent after the emergence of big data, as it is shown by the growing body of computational social science research (Hox, 2017; Lazer et al., 2009). As described in chapter 2, however, several social scientists remain hesitant about the use of data-driven methods. When discussing the rise of big data, for instance, Clark and Golder (2015, p. 66) mentioned sampling populations, confounders and overfitting as some of the reasons why political scientists tend to have some reservations about data-driven studies. Similarly, many social scientists generally curb their enthusiasm about social media data and consider how sampling bias is likely to influence studies based on Twitter or Facebook information (Culotta, 2014; Mislove et al., 2011). To give an example, Gayo-Avello (2011) illustrated that sampling problems due to age bias can influence data studies aimed at predicting political elections from Twitter sentiment. To avoid such problems, contextual considerations, as well as the analysis of the adequacy of the results should, according to some social scientists, be a crucial part of the process of scientific inquiry (see for instance Bornakke & Due, 2018; Kitchin, 2014; Monroe et al., 2015; Titiunik, 2015).

In this chapter I shall describe one strategy that can help social scientists to test the adequacy of data-driven results and that can be used to tackle some of the data-driven problems described in chapter 2: the approach known by the name of the Russo-Williamson thesis (henceforth RWT), according to which a causal claim can be established only if it can be established that there is a correlation between the cause and the effect, and that there is a mechanism linking the cause and the effect that is responsible for such a correlation (Russo & Williamson, 2007). The applicability of Russo and Williamson's idea was hugely debated in relation to biomedical research (Clarke et al., 2014; Gillies, 2010; Illari, 2011; Weber, 2009; Williamson, 2018). In the social sciences, however, the debate has never been shaped rigorously, and while some philosophers and social scientists have advocated the use of different kinds of evidence for causal

discoveries (Brady & Collier, 2004; George & Bennett, 2005; Little, 2015; Moneta & Russo, 2014; Steel, 2004), Reiss (2009) and Claveau (2012), two philosophers of the social sciences, have criticised this approach. With this chapter, I aim to defend RWT from criticisms and to show its importance in the social sciences.

The chapter is structured as follows. In section 3.2, I shall summarise what RWT is and I shall describe how this approach can be applied to the social sciences. In section 3.3, I shall reconstruct four main criticisms of this thesis proposed in the philosophy of the social sciences literature: namely i) RWT is not used in the social sciences, ii) RWT does not work, iii) causal concepts do not overlap and iv) there are better alternatives to RWT. In section 3.4, I shall provide a defence of RWT based on two general considerations: i) the strategy proposed by RWT is not infallible and contemplates exceptions, and ii) RWT appears perfectly in line with the research methods used in the social sciences. Section 3.5, finally, will offer some examples in which RWT might solve problems of data-driven studies. Section 3.6 will conclude the chapter.

3.2 The Russo-Williamson thesis

The thesis proposed by Russo and Williamson requires both the identification of a correlation, or *difference-making* relationship, between the cause and the effect, and the identification of a *mechanism* linking the cause and the effect to establish causation.

The origins of the distinction between difference-making and mechanistic relationships can be traced back to the philosophical discussions on causation that emerged in the second half of the twentieth century. On the one hand, in the so-called 'difference-making accounts', a causal claim is established when there is an appropriate difference-making relationship between the cause A and the effect B (see for instance Eells, 1991; Lewis, 2004; Reichenbach, 1958; Suppes, 1970). Such a relationship is understood either as a probabilistic relationship (A causes B only if A raises the probability of the occurrence of B), or as a counterfactual relationship (A and B are actual events, and if A had not occurred, then B would not have occurred). Evidence in support of difference-making relationships is hence understood as any evidence proving that i) the effect would not occur if the cause did not occur; ii) the cause *is correlated* to the effect: evidence from randomised controlled trials, experiments or quantitative analyses counts in general as difference-making evidence. As an illustration, the claim that "smoking causes lung cancer" can be based on evidence showing that i) intervening on smoking behaviours

results in a decrease of cancer rates; ii) when smoking rates decrease, lung cancer rates decrease too; iii) smoking increases the probability of developing lung cancer. On the other hand, according to the mechanistic accounts, a causal relationship between A and B corresponds to a causal mechanism, or process, linking A and B. This idea has been spelt out in different ways: according to some philosophers (see for instance Dowe, 2007; 1992; Reichenbach, 1958; Salmon, 1984, 1997), a causal mechanism is a spatio-temporal process that should be understood through the language and tools of physics, such as energy and momentum. According to others, (Craver, 2006, 2007; Glennan, 2002; Machamer et al., 2000) a causal mechanism is an organised system made of entities and activities. In these terms, for instance, the claim that "smoking causes lung cancer" can be supported by evidence of a sufficiently well understood biological mechanism made of entities (such as proteins and genes) and activities (such as protein expressions or genetic mutations) that links smoking and lung cancer.

It is useful to expand a little more on RWT: this thesis requires evidence to support the claim "there is a correlation between the cause and the effect", and evidence to support the claim "there is a mechanism linking the cause and the effect"[6]. In many cases, such a requirement is associated with the use of different kinds of studies. Indeed, while difference-making evidence is in general obtained through randomised controlled trials or statistical studies, evidence of mechanisms is often obtained through a broad group of mechanistic evidence-gathering methods, among which experiments, case studies and observational studies. It is however important to distinguish between types of evidence and types of evidence-gathering methods, because they do not always coincide. Mechanistic evidence is *not necessarily* collected through mechanistic evidence-gathering methods, similarly difference-making evidence is not always obtained through randomised controlled trials or statistical studies. Illari (2011, pp. 5–6), to give an example, argued that mechanistic evidence can be obtained through repeated trials, as happened in the trials conducted by Crick et al. (1961) on chemical mutagens to crack the genetic code. There might be cases, consequently, where mechanistic claims are supported by evidence obtained from statistical studies or randomised controlled trials, and cases where correlational claims are supported by evidence collected in mechanistic

[6] The amount of evidence should be sufficient to establish the presence of a correlation and the presence of a mechanism. In the case of mechanisms, mechanistic evidence is likely to be partial and to shed light only on some aspects of the mechanism. If researchers have good reasons to think that this information is sufficient, they might decide to establish the presence of the mechanism.

studies. The difference between types of evidence and types of methods is represented in Figure 6.

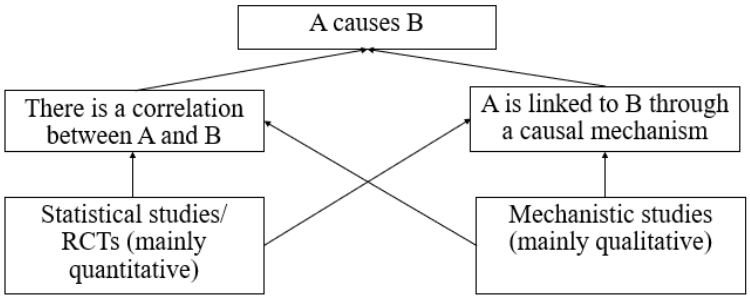

Figure 6. RWT requires the collection of evidence supporting both the claim "there is a correlation between the cause and the effect", and the claim "there is a mechanism linking the cause and the effect". It does not require the use of difference evidence-gathering methods. Furthermore, it does not assume that mechanistic claims are supported only by mechanistic studies, and that correlational claims are supported only by statistical studies. A mechanistic claim, for instance, might be supported by evidence obtained through statistical studies or randomised controlled trials.

3.2.1 The Russo-Williamson thesis in the health sciences

Originally, Russo and Williamson put forward an account of evidential pluralism in the health sciences. They developed their thesis by considering what claims need to be established in causal studies, and did not explicitly discuss the methods required to establish such claims. In other words, they focused their attention on the top of Figure 6, without discussing pluralism of evidence-gathering methods. This point was further clarified by Illari (2011, pp. 3–6).

While the collection of evidence supporting a probabilistic relationship was a process widely discussed in the medical sciences, where for a long time randomised controlled trials had been considered the 'gold standard' (Bothwell et al., 2016), how to understand the notion of evidence in support of a mechanistic relationship remained questionable. Recently, Williamson (2018) has clarified that, in their account, the term 'mechanism' can be understood broadly in three ways: i) as a complex system consisting of entities and activities organised in such a way that, together, they are responsible for the phenomenon under study (as described by Machamer et al., 2000); ii) as a mechanistic process through

time and space along which a signal is propagated (Dowe, 2007; Reichenbach, 1958; Salmon, 1997); iii) as a combination of a complex system and a process.

Some examples can help to clarify this point. In his fascinating autobiography *In search of memory* (2006), Kandel described the discovery of the complex molecular system of long-term memory, organised in entities (neurons, proteins and genes) and activities (such as protein movements and gene expressions). As illustrated in Figure 7, an environmental stimulus such as the presence of a dangerous animal in the room activates the release of serotonin, that in turn acts on the sensory neuron to increase cyclic adenosine monophosphate and causes protein kinase A and MAP kinase to move to the nucleus, activate the binding protein CREB-1 (that activates gene expression) and deactivate the binding protein CREB-2 (that suppresses gene expression). The activation of gene expression changes the function and the structure of the cell, facilitating in this way the growth of new synaptic connections.

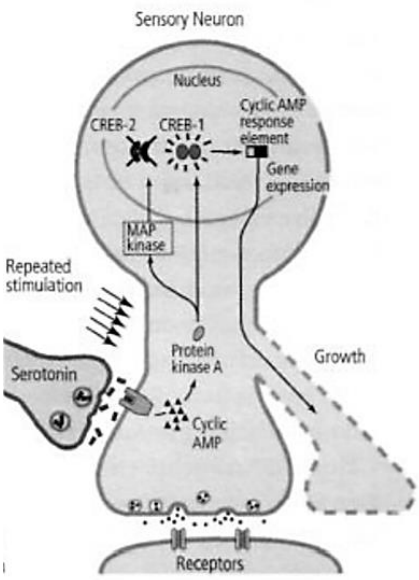

Figure 7. The complex molecular system of long-term facilitation composed of entities (neurons, proteins and genes) and activities (like protein movements and gene expressions) (Kandel, 2006, p. 204).

Regarding the notion of causal processes, furthermore, a case in point is the propagation of an electrical signal from an artificial peacemaker to the appropriate part of the heart.

The account proposed by Russo and Williamson was purely epistemological: RWT, they argued, is a thesis about how to properly *establish* a causal relationship. The epistemological account was proposed as both descriptive and normative. It is claimed to

be descriptive because it captures a typical approach used in the health sciences to establish causality. The International Agency for Research on Cancer, for instance, regularly brings together experts to establish whether particular exposures are carcinogenic to humans, and its overall evaluation is based on a variety of evidence, among which epidemiological (difference-making) studies and mechanistic information (Wilde & Parkkinen, 2017).

It is claimed to be normative because it is needed to reduce considerably the risk of bias when causal claims are established (this point was confirmed by a later formulation of the thesis in Clarke et al., 2014). In chapter 2 I have discussed some of the reasons why machine learning algorithms (as well as other difference-making methods) might detect false correlations between variables: in some circumstances, sample bias might cause Simpson's paradox and subgroups with a common trend might show the reverse trend when they are aggregated. In other situations, low-quality data or the curse of dimensionality can lead researchers to establish a correlation that does not exist in the underlying population, or the absence of relevant data might make scientists ignore an important variable that acts as a confounder. All these errors entail either the inference of a non-existent correlation (like in the case of low-quality data showing a correlation that would not exist if variables were correctly measured), or the inference of a non-existent causal relationship from an existing correlation (like in the case of confounders).

Even causal studies based only on the presence of a causal mechanism might generate inferential mistakes, among which the most famous is the problem of 'masking'. Establishing the presence of a causal mechanism is not sufficient to establish causation because there may also exist further counteracting mechanisms whose effects interact with (or mask) the effect of the first mechanism, with the consequence that any difference made through the established mechanism is cancelled out by these counteracting mechanisms (Illari, 2011). Steel, for instance, offered an example examining the relationship between exercise and weight loss: on the one hand, there is a well-known mechanism whereby exercise can cause more calories to be burned; on the other hand, exercise increases appetite, and this can lead people to eat more than usual (Steel, 2008, p. 68). In other words, the effect of one mechanism might mask the effect of the other.

Establishing both the presence of a correlation and the presence of a causal mechanism can help to reduce these forms of errors, and this is why Russo and Williamson argued

that their thesis has a normative dimension. Establishing the presence of a causal mechanism, for instance, can allow researchers to uncover bias caused by Simpson's paradox, or can help them to detect problems associated with the quality of data. Establishing the presence of a correlation, instead, can help to tackle the problem of masking by recognising whether the cause, via the established mechanism, does make a real difference to the effect.

3.2.2 RWT in the social sciences

Russo and Williamson developed their claim by considering how researchers infer causality in the health sciences, nevertheless they expressed their confidence in the possibility of extending the epistemological claim also to other domains. For, instance, they recognised that:

> "evidence in the social sciences is very diverse too. The social sciences look for mechanisms, which can be social, biological or mixed, and for dependencies between variables of interest [...]. Yet in using both mechanistic and probabilistic evidence scientists do not aim at establishing different causal claims" (Russo & Williamson, 2007, p. 169)

According to this claim, when social scientists establish casual relationships, both difference-making and mechanistic claims are in general established. Like in the health sciences, various types of mechanistic evidence can be collected because social causal mechanisms can be conceptualised in different ways. Researchers might look for complex systems made of organised entities and activities, but they could also search for causal processes linking the cause and the effect through time and space. In particular situations, furthermore, they might find a combination of them.

Unlike in the health sciences, however, it is difficult to find social phenomena that can be understood as causal processes à la Salmon (Dowe, 2007; 1992; Reichenbach, 1958; Salmon, 1984, 1997), given that most of the causal phenomena in the social sciences cannot be spelt out in terms of physical quantities. Values and beliefs, for instance, play often a crucial role in social phenomena, and are not associated with any physical object. When social scientists talk about causal processes, consequently, they often refer to a group or chain of events that leads to specific effects. Maxwell (2004a, p. 5), for instance, argued that causal processes are "processes by which some events influence others",

while Little (2018, p. 415) described them as a combination of social conditions, constraints, or circumstances that together bring about a causal effect.

An illustrating example that shows how a mechanism can be conceptualised as a causal process is offered by Skocpol's political study (1979). By analysing social revolutions in France (1788-9), China (1911-1916) and Russia (1917), Skocpol developed a reconstruction of the common causal chain that led, in the three different contexts, to a revolution. The key steps of the causal chain were, according to her study, i) the growing of political pressure from more economically developed countries; ii) an agrarian economy (peasants as a major presence) and iii) a non-autonomous State (for instance, the dominant class had strong political leverage and could contrast the State's decisions). Together, these factors caused a State breakdown. For instance, in China the breakdown followed the transfer of power to the regional level that had occurred in the final years of the old regime, before 1911. At the same time, iv) peasant autonomy and v) landlord vulnerability led to a massive peasant revolt. Together, finally, the State breakdown and the peasant revolt caused a social revolution, as illustrated in Figure 8.

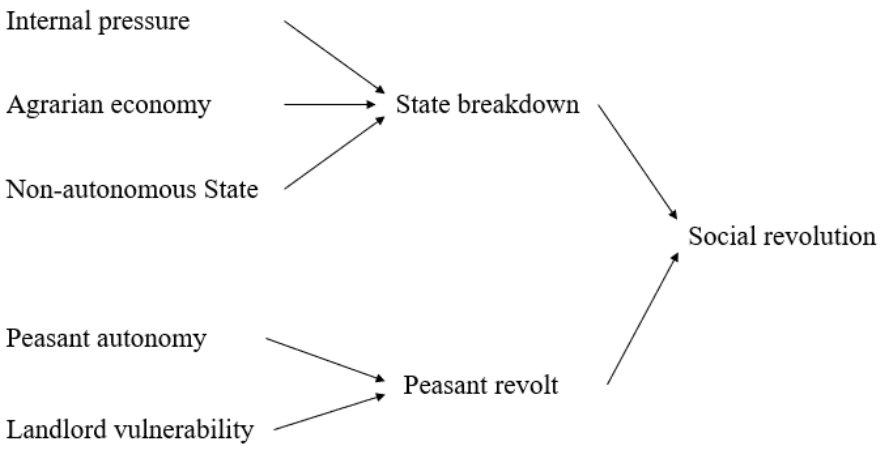

Figure 8. The causal process leading to social revolutions studied by Skocpol (1979).

As concerns an example of an organised system with its own entities and activities, it is possible to analyse the structure of the well-known segregation mechanism studied by Schelling (1978). According to his studies, individual discriminatory preferences, all together, constitute the neighbourhood structure, that can in turn be characterised by a phenomenon of segregation. The initial mechanistic assumption is that the society is composed of individuals who belong to a specific group and who are able to discriminate

(i.e. to distinguish between their own group and other groups). In Schelling's mechanistic model, all individuals will be happy to live in a mixed neighbourhood, but will be unhappy if they have minority status. By interacting with each other, each individual—who is free to move if he is discontent with the composition of his neighbourhood—will look for the neighbourhood with the preferred composition. Of course, if an individual abandons his neighbourhood, his activity changes the composition of both his past and future neighbourhoods, with consequences for the others (for instance, some individuals of the same group, now in minority, might decide to leave their neighbourhood). Due to the continuous activities of the agents, everyone will end up living in segregated neighbourhoods. This would be a consequence of the simple individual preference not to live as a minority. In such a case a social mechanism consisting of entities (individuals) and activities (their movements between one neighbourhood and another), would produce the relevant effect on the state of segregation of a population. Figure 9 can help to illustrate it: each spot represents an individual. The overall group is composed of two groups (the red and the blue). After the first individual movements, all individuals react to them with the consequence that, in the end, the two groups are clearly divided.

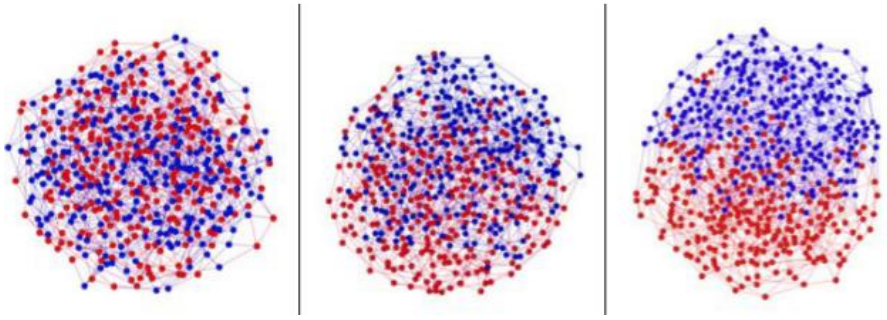

Figure 9. The complex social system of segregation. Figure from Cortez and Rica (2015, p. 64).

From a normative point of view, evidence supporting the presence of either a complex mechanism or a spatio-temporal process, and evidence showing a difference-making relationship, would be required by RWT to establish a causal claim in the social sciences.

3.3 Four criticisms of RWT

Numerous causal studies in the social sciences apparently establish both correlational and mechanistic claims, however some philosophers argued that RWT does not work very well in such a field of study (Claveau, 2012; Reiss, 2009). Their criticisms are based both

on descriptive and normative observations and can be divided into 4 separate arguments. All these arguments are about the type of claims established and do not discuss pluralism of evidence-gathering methods. In other words, they focus only on the top of Figure 6.

3.3.1 Against descriptive RWT: social scientists do not use it

RWT contains a descriptive claim according to which scientists establish a difference-making relationship as well as a mechanistic relationship between the cause and the effect when they establish causation. Claveau (2012) disagreed with this observation, and based his criticism on a concrete case study: the analysis of the causal relationships, across countries, between unemployment, the unemployment benefit system and the employment protection legislation. To examine RWT, the author reconstructed how economists established causal claims concerning the causes of unemployment (Claveau, 2012, p. 808).

Three causal relationships were accepted within the scientific community: (i) unemployment benefits increase unemployment, (ii) the strictness of the unemployment benefit eligibility conditions—linked to 'job search intensity'—reduces unemployment, iii) the strictness of the employment protection legislation has no net effect on unemployment. The first and the third claims were supported by evidence of difference-making relationships obtained through regression analyses, and by evidence of causal mechanisms collected by means of micro-data studies focused on potential causal pathways between the cause and the effect (such studies enable researchers to look, for instance, at the firing and hiring firms' behaviours in cases of a stricter employment protection and of weaker legislation). Nevertheless, the second claim ("the strictness of the unemployment benefit eligibility conditions—linked to 'job search intensity'—reduces unemployment") was established by identifying only the presence of a causal mechanism, without collecting difference-making evidence (Claveau, 2012, p. 812).

Claveau argued that this happened because economists did not find a measure of the strictness of unemployment benefit eligibility (SUBE) that was comparable across countries: the most suitable measure was the cost of 'active labour market policies', or ALMPs, the system of programmes intervening in the labour market to help unemployed people find work. This variable, however, did not include only SUBE, but many other factors such as placement services and subsidised training (Claveau, 2012, p. 812). Given that the cost of ALMPs was not determined only by SUBE, this was recognised as a bad

proxy for it. Some economists tried to disaggregate ALMPs expenditures into five components, among which 'Public Employment Service', that was thought to be a good proxy for SUBE (Bassanini & Duval, 2006, pp. 81–83). The coefficient of that component, however, was proved to be very unstable, consequently any statistical analysis based on it was not accepted as good evidence of a different-making relationship. Due to this limitation, economists decided to establish a causal claim just considering some evidence obtained through mechanistic studies. This situation, hence, is in contrast with the descriptive form of RWT.

3.3.2 Against normative RWT: RWT does not work

Suppose we want to understand if watching violent TV programs causes violent behaviours. We start by searching for a causal mechanism, and we find out that, in a sample population, consumers identify themselves with aggressive characters and start confusing reality and fantasy, with the consequence that they begin to believe that the depicted scenarios are realistic. This, in turn, leads them to react violently even in real-life situations. Now that we have evidence of a mechanism in support of our causal hypothesis, we look at the correlations between violent TV programs and violent behaviours in the total population under study, and we collect evidence of a difference-making relationship. It seems, hence, that we can establish our causal claim.

The problem, however, is that the correlation is actually brought about by a common cause that was not measured. In other words, the correlation that we are considering is spurious. This happens because, rather than having just one causal mechanism linking violent TV programs and violent behaviours, in the population under study there are two acting mechanisms that mask each other, as described in Figure 10. The second mechanism, for instance, might lower the violence in some of the consumers because TV consumption acts as a deterrent. In such a way, the effects of the two mechanisms cancel each other out and, if it were not for the spurious correlation, we would not find any difference-making relationship between violent TV programs and violent behaviours.

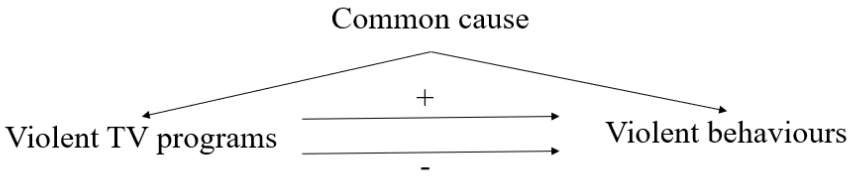

Figure 10. The relationship according to which watching violent TV programs increases violent behaviours is masked by another causal mechanism whereby watching violent TV programs decreases violent behaviours. The remaining probabilistic relationship between watching violent TV programs and violent behaviours is brought about by a common cause.

This is the fictitious example proposed by Reiss (2009, p. 28) to justify his criticism: RWT does not work. Unlike Claveau's analysis, this consideration poses a problem to the normative RWT claim. How can we say that social scientists should use RWT, if it can lead to incorrect causal claims like in the example above?

3.3.3 Against normative RWT: non-overlapping causal concepts

This first criticism proposed by Reiss is strictly linked to a second consideration, according to which different concepts of causation do not necessarily overlap. Reiss introduced this idea discussing the errors made in the violent TV programs example:

> "In our example, the hypothesis we can hope to establish or reject on the basis of statistical evidence is a probabilistic one: in a causally homogeneous population, is violence correlated with the consumption of violent TV programs? (Answer in the example: no.) Using mechanistic evidence, by contrast, we can hope to establish or reject a mechanistic hypothesis: is there, in some individuals, a continuous mechanism from "input variable"—TV consumption—to "output variable"—violence? (Answer in the example: yes.)" (Reiss, 2009, p. 31)

To clarify his claim, Reiss described some common situations that are characterised by this problem: those where omissions have causal power, and those where there is a problem of masking. To begin with, he considered that there are several cases in the social sciences where omissions are claimed to cause phenomena: negligence in civil law, for instance, is in general understood as a failure to act, and is often claimed to be the cause of some effects (such as the complainant's harm) even if it is impossible to find a physical mechanism between them. In addition, he observed, the presence of a mechanism does not exclude the possibility that the overall influence of the cause to the effect is null. Masking situations can hence be interpreted as cases where causal concepts do not overlap.

Reiss argued that, due to these situations, knowing that there is a correlation between X and Y should not allow us to conclude that, if the relationship is causal, there should also be a mechanism linking X and Y. Similarly, knowing that there is a mechanism should not allow us to infer that, if the relationship is causal, we should find also a correlation. Even in this case, Reiss' criticism was directed against normative RWT: if this problem were proved to be true, RWT would be considered inadequate in the social sciences. Consequently, social scientists should not be asked to collect both evidence of mechanisms and evidence of difference-making relationships to establish causal claims.

3.3.4 Alternatives to RWT: conceptual pluralism and conceptual monism

Russo and Williamson observed that social scientists tend to establish the presence both of a correlation and of a mechanism to claim that the relationship between two factors is causal. This, according to them, cannot be explained by the existing monistic or pluralistic accounts of causality. On the one hand, existing accounts of conceptual monism would not be able to explain why more than one type of relationship is established, on the other hand existing accounts of conceptual pluralism would entail that scientists use different types of evidence to support different kinds of causal claim (evidence of correlation to support a causal claim that sees causation as a difference-making relationship, evidence of a mechanism to support a causal claim that sees causation as a mechanistic relationship). Neither monistic nor pluralistic accounts, if we consider the existing possibilities, would be able to account for the use of different types of evidence in support of a unique casual claim[7].

Both Reiss and Claveau criticised this argument, even if in diametrically opposed ways: Reiss defended conceptual pluralism, Claveau supported the use of conceptual monism in the social sciences.

Starting from the consideration that the overlap between the presence of mechanisms and of correlations is only an empirical truth that needs to be confirmed a posteriori, Reiss concluded that we should not accept RWT. The best option we have is to accept that each causal method *defines its own concept* of causation: for statistical methods, "X causes Y" means that there is a correlation between X and Y, nothing more, while for a mechanistic approach "X causes Y" means that there is just a mechanism between the cause, X and

[7] While Russo and Williamson argued for a metaphysical account of causality that they called the epistemic theory of causality, the form of evidential pluralism proposed by RWT does not say anything directly about the concept of causality that researchers should endorse (see also Williamson, 2018).

the effect, Y[8]. Evidence of a correlation and evidence of a causal mechanism, therefore, would not support the same causal claim, but different causal claims, one for each of the accounts used. To defend the applicability of his form of conceptual pluralism, furthermore, Reiss argued that the fact that social scientists do not specify the concept of causation they endorse should not be interpreted, like Williamson and Russo did, as a supporting factor in favour of the idea that there can be just one concept of causation bearing relationships with different kinds of evidence. Social scientists are pluralist regarding the concepts of causation, and this ambiguity has only caused equivocations.

Claveau, on the contrary, argued that the use of RWT can be justified within a monistic account. He argued that it is very rare, in the social sciences, to conduct a good difference-making study with quality high enough to rule out the risk of bias. It is for this reason, hence, that social scientists embracing a difference-making account in general prefer to use also evidence supporting the presence of a mechanism:

> "it would be an anomaly for the counterfactual-manipulationist account if other (mechanistic) evidence for a specific claim was demanded even though this same claim was already backed by the best difference-making evidence. This proposition has however little practical relevance since difference-making evidence is perhaps never of the 'perfectly adequate' type" (Claveau, 2012, p. 809)

Both Reiss and Claveau disagreed with the claim that neither existing forms of conceptual monism nor existing forms of conceptual pluralism can account for the use of a plurality of evidence in a single study, but there is a fundamental difference between their objections. Reiss argued that scientists use different types of evidence to support different causal claims: conceptual pluralism was considered an alternative account that clarifies how social scientists use evidence. Claveau, instead, accepted that different types of evidence are used for a single causal claim, and justified such an approach highlighting the methodological limitations of the (monistic) difference-making account that, according to him, social scientists embrace when they study causality. In his view, until researchers will not be able to improve their methods, different kinds of evidence will be used to support a single causal claim.

[8] Reiss recognised, however, that if the mechanism under study is the only one that links the two variables, scientists expect the variables to be correlated as well (Reiss, 2009, p. 30).

3.4 A defence of RWT in the social sciences

In this section, I shall propose an answer to each criticism discussed above. To begin with, in section 3.4.1 I shall examine the descriptive claim according to which there are situations where social scientists do not establish the presence both of a difference-making relationship and of a causal mechanism, and I will show that this criticism can be solved by arguing i) that researchers can establish the presence of a correlation without conducting statistical studies; ii) that researchers did not establish causation or that iii) *normally* social scientists comply with RWT, even if there are exceptions. In section 3.4.2 I shall explain that RWT requires that researchers *properly* establish the presence of a correlation, and that such a thesis is claimed to lower the risk of methodological bias in causal studies. It is not expected, consequently, that RWT cancels all the possible inferential mistakes when a causal claim is properly established. This consideration can be used to answer Reiss' criticism according to which RWT might lead researchers to establish wrong causal relationships. Next, in section 3.4.3 I shall argue that in general causal concepts overlap, while in section 3.4.4 I will show why RWT should be preferred to existing accounts of conceptual pluralism, and why it is possible to accept both RWT and existing forms of conceptual monism. A careful analysis of RWT and real examples from the social sciences will support my defences of RWT in the social sciences both at the descriptive and at the normative level.

3.4.1 Social scientists comply with RWT

Russo and Williamson's descriptive claim is based on the observation that, in the social sciences, researchers tend to use a plurality of evidence to support and establish causal relationships. As described by Claveau himself, for instance, for two of the three causal claims established, economists collected both evidence of difference-making relationships and evidence of mechanistic relationships.

From this claim, however, it cannot be excluded that, under particular circumstances, social scientists can establish a causal claim using just one type of study. As illustrated in Figure 6 in section 3.2, there are cases where either mechanistic studies can be used to establish correlations, or statistical studies can be used to establish the presence of mechanisms. Illari (2011, pp. 5–6), for instance, described a case where researchers established the presence of a mechanism by collecting evidence from repeated trials (Crick et al. 1961).

This consideration can be used to develop a possible answer to Claveau's criticism. It could be argued that, in the case described by Claveau, the mechanistic studies conducted by researchers were sufficiently good to allow them to establish both the presence of a mechanism between the cause and the effect, and the presence of a correlation linking the cause to the effect. In the context described by Claveau, this answer seems plausible: economists established the presence of a mechanism using theoretical models, simulations, and micro-data studies conducted in several countries (see Fredriksson & Holmlund, 2006, pp. 370–377). In addition, some difference-making evidence was collected by means of randomised controlled trials. Although evidence from trials was not sufficient, according to researchers, to establish the presence of a correlation, the combination of this evidence with the set of information obtained through mechanistic studies could have been considered sufficient to establish the presence both of a mechanism and of a correlation.

There is a second possible answer that can be offered to Claveau: due to the impossibility of establishing a correlation, economists did not establish the presence of a causal relationship but only argued that the strictness of the unemployment benefit eligibility conditions *is likely* to reduce unemployment. In the review of the literature published by Fredriksson and Holmlund (2006) cited by Claveau, for instance, the authors considered that the available evidence was not wholly conclusive and concluded that:

> "more stringent search requirements *are likely* to speed up transitions to employment" (Fredriksson & Holmlund, 2006, p. 377)

If one of these two answers is accepted, then it might be possible to conclude that the case proposed by Claveau does not pose a problem for RWT.

Someone, however, might reject both these answers, and might argue that economists actually established causation without having established correlations. This position would cause a problem to RWT because it would be understood as an exception, a case in which RWT did not hold.

My answer to this argument would be twofold. On the one hand, I would consider that exceptions do not pose a real problem to RWT. While in their first discussion of RWT (2007), Russo and Williamson did not specify it, recently Williamson (2018) has clarified that *normally*, in order to establish a causal claim, one needs to establish both correlation and mechanism (Williamson, 2018, p. 1). It is hence possible to have cases where

causation is established even if either the presence of a mechanism or the presence of a correlation is not established. In these terms, Claveau's case study can be accepted as one of the exceptions where a causal claim is established without establishing the presence of a difference-making relationship.

On the other hand, I would argue that, as it will be explained in detail in section 4.3, Williamson has changed RWT having in mind some situations where either a difference-making relationship or a causal mechanism *does not exist*. Claveau's criticism, instead, simply shows a case where researchers *were not able to collect sufficient evidence to establish correlation*. Their decision to establish causation appears motivated by the fact that their good mechanistic studies and some (insufficient) difference-making evidence from randomised controlled trials led them to believe in the presence of an unmeasurable correlation. Although it remains *possible* that causation was actually studied without establishing the presence of a correlation, it seems more plausible that the combination of mechanistic evidence and evidence from randomised controlled trials allowed researchers to establish the presence of a correlation.

It is important to note that my answer according to which the causal claim was actually established without establishing the presence of a difference-making relationship accepts Claveau's case study as an exception. It is assumed that, generally, this practice is not used in causal studies. It is hence helpful to provide real cases that illustrate how, in general, causal relationships in the social sciences are established by establishing both the presence of a mechanism and the presence of a correlation. To begin with, we can analyse the causal debate about the relationship between natural resources and civil war. As explained by Ross (2004), several studies identified a strong correlation between natural resources and civil war in different cases. Collier and Hoeffler (1998), for instance, argued that states that massively export primary commodities have a higher risk of civil war if compared to resource-poor states. Buhaug and Gates (2002), furthermore, discovered a strong correlation between the incidence of natural resources in the conflict zone, and the scope of the conflict (in terms of the conflict's geographical area). Ross argued that statistical correlations cannot make a causal claim complete and persuasive, and considered that the correlation between natural resources and civil war could be spurious:

> "[…] both civil war and resource dependence might be independently caused

> by some unmeasured third variable, such as a weak rule of law. A state where the rule of law is weak might be unable to attract investment in its manufacturing sector, and hence would depend more heavily on resource exports; this state might also face a heightened risk of civil war through a different process. The result could be a statistically significant correlation between resource dependence and civil war, even though neither factor would cause the other" (Ross, 2004, p. 36)

It was this reason that motivated social scientists to look for causal mechanisms able to account for such a correlation. Since several mechanistic hypotheses were proposed in the literature, Ross verified their presence in thirteen cases of civil wars. He focused his attention on seven mechanistic hypotheses (to give an example, one of the hypotheses was that the availability of natural resources offers the population of resource-rich zones an incentive to form a separate state, and this increases the probability of civil war). He found out that five of these mechanisms operated in the cases under study, and identified also four unforeseen mechanisms linking resource wealth and civil war. The possibility of observing the hypothesised causal mechanisms at work led Ross to conclude that there was sufficient evidence in all the thirteen cases to support the claim according to which natural resource wealth is causally linked to civil conflicts (Ross, 2004, p. 61).

Further examples of RWT in the social sciences will be discussed studying Wood's study on the reasons why peasants in El Salvador decided to join rebel movements (2003) (in chapter 4), Garip and Asad's study concerning the phenomenon of Mexico–U.S. migration (2016) (in chapter 5), and in the analyses conducted within the project LIFEPATH's about how socio-economic conditions influence health (in chapter 6).

Overall, Claveau's criticism can be solved in three ways. First, it can be argued that researchers do not need a statistical study to establish the presence of a correlation between the cause and the effect. In this case, the mechanistic studies conducted by researchers were sufficiently good to allow economists to establish both a correlation and a mechanism between the cause and the effect. Second, it can be claimed that this case is perfectly consistent with RWT because, due to the lack of difference-making evidence, researchers did not establish causation but only claimed that the strictness of the unemployment benefit eligibility conditions *is likely* to reduce unemployment. Finally, if it is confirmed that researchers did not establish a correlation but established causation, it

might be argued that, in general, researchers need to establish the presence both of a correlation and of a mechanism in order to properly establish causation, but there can be exceptions. In such a context, Claveau's example can be considered an exception.

3.4.2 RWT works in the social sciences

My answer to Reiss' criticism is that RWT requires that, in order to properly establish causation, researchers properly establish the presence of a correlation and of a mechanism. In Reiss' example, however, it can be argued that correlation was not properly established because the study did not allow researchers to claim that there was a probabilistic dependence between watching violent TV programs and violent behaviours *conditional on potential confounders*[9] (Williamson, 2018, p. 4). Researchers did not rule out the possibility of confounders, otherwise they would have discovered the presence of the common cause responsible for the correlation. This consideration is sufficient to conclude that, in Reiss' example, researchers did not properly establish correlation and, consequently, did not comply with RWT. Therefore, the inferential mistake described by Reiss is not due to RWT, and there is no reason to reject the normative dimension of RWT.

Reiss might reply that there are cases where researchers properly establish a correlation after having ruled out all the possible confounders they were able to hypothesise. However, they may not have been able to imagine a specific possible confounder Z, with the consequence that they might not check if the probabilistic dependence between the cause and the effect is conditional on Z. This might lead them to establish the presence of a correlation between the cause and the effect even if the correlation is caused by Z.

In this context, my response would be that RWT is a thesis about how to properly establish causation and to reduce the risk of bias. In some situations, researchers can conduct high-quality studies, rule out the possibility of confounders and properly establish causation. Establishing causation, however, is fallible: it follows that researchers can properly establish a causal relationship that, in reality, does not exist. The normative dimension of RWT is based on the claim that evidence of difference-making relationships and evidence of mechanisms have different weaknesses, therefore through their combination scientists can *reduce* the probability of incorrect causal inferences (Illari, 2011, p. 146). It is not

[9] It is important to specify that RWT does not require that researchers test all the confounders that can be found, but only to think about possible confounders and to verify their absence. Background information, in general, plays a vital role in this process.

claimed that RWT helps to rule out all possible inferential mistakes: properly establishing both the presence of a correlation and the presence of a mechanism does not guarantee the correctness of the causal inference, it only *reduces the possibility of methodological bias* that might lead to inferential mistakes.

Consequently, even if Reiss could offer real-life examples in which causal relationships are properly established but there are inferential mistakes, this would not cause substantial problems to RWT.

3.4.3 Overlapping causal concepts

The possibility of having different non-overlapping causal concepts can be seen as a big obstacle to RWT. However, as I shall show, in at least some cases RWT can allow for situations where causality is apparently not accompanied by difference-making relationships, or a causal link does not act through a causal mechanism.

Initially, Russo and Williamson did not address explicitly this problem, but Williamson (2018) has recently discussed it, admitting that RWT contemplates exceptions. Such exceptions, however, are rare enough not to pose a real problem to RWT. If we consider Reiss's examples, moreover, it is arguable that at least one of the examples used in his paper should be understood as a real exception.

Let us consider the case of masking. To begin with, it is curious to see that Reiss used this problem as a reason to reject RWT. It was exactly the problem of masking, indeed, one of the weaknesses described by the authors to show how the collection of different kinds of evidence could lead to less bias in causal inference. Discovering only one of two counteracting mechanisms linking X and Y would provide just a partial mechanistic explanation of the effect Y, and it would make impossible to claim either that the cause X has a positive or negative effect on Y, or that the cause X has a null effect on Y.

The first response, hence, might be that masking does not necessarily entail that the two concepts of causation—mechanistic and difference-making—are not overlapping. Each acting mechanism exerts an effect and produces a difference-making relationship, the problem is that counteracting mechanisms and their probabilistic effects interact, and it might be then difficult to identify the difference-making relationships associated with every acting mechanism. In some cases, this difficulty can be overcome: in Reiss's example, for instance, scientists would probably find a difference-making relationship

(i.e. watching violent TV programs increases violence) in the sample population used for the mechanistic study. This relationship, however, would be hidden in the total population because another part of that population shows exactly the opposite relationship (i.e. watching violent TV programs decreases violence).

In other cases, two or more mechanisms can act in the same population, therefore it might be impossible to isolate their distinct probabilistic effects. I have mentioned above the example proposed by Steel (2008, p. 68): a person who does exercise causes calories to be burned; but can also cause her appetite to increase. If researcher considered a group of people, hence, it would be possible to find in each person involved in the study both the mechanism through which exercise reduces bodily fat by burning calories, and the mechanism whereby exercise increases bodily fat by increasing appetite and making a person eat more than usual.

This, however, should not lead scientists to believe that there is a mechanism without a corresponding difference-making relationship. Rather than being cases where the causal concepts do not overlap, these could be understood as cases where scientists have failed to find evidence showing the distinct difference-making relationships.

This first answer should be accompanied by a further consideration. The problem of masking has been discussed several times in the social sciences. I have already mentioned Steel's example of exercise and weight-loss (Steel, 2008, p. 68), another example was proposed by Elster (1993), who observed that the 'sour grapes mechanism', according to which individuals' desires are adjusted in accordance with the difficulties related to their achievement, is often masked by an opposite mechanism whereby people desire what they cannot have, precisely because it is impossible to have it.

Unlike what argued by Reiss, in general, social scientists do not consider the claim "there is a mechanism linking X and Y" as sufficient to establish any kind of causation until the possibility of masking is ruled out. Also Claveau, for instance, claimed that:

> "since the causal claim was established solely with mechanistic evidence, it must mean that the specialists are quite confident in the reliability of mechanistic evidence for this specific claim—i.e., they don't expect missing paths, general equilibrium effects and country heterogeneity to be major problems." (Claveau, 2012, p. 812).

Social scientists are aimed at recognising when counteracting mechanisms act together.

They do not want merely to discern one of them, they want to understand the conditions under which they act separately or together (see Elster, 1993, p. 2). This goal is clearly illustrated by the fact that one of the chapters of George and Bennet's book *Case Studies and Theory Development in the Social Sciences* was used to develop new strategies to study the complex interactions between different, contrasting mechanisms (George & Bennett, 2005, pp. 233–262).

It could be concluded that the fact that social scientists want to recognise when a mechanism is masked is in accordance with the idea that evidence in favour of a type of causal relationship (mechanistic, in this case) should entitle scientists to believe in the presence of another type of causal relationship (difference-making). Social scientists want to know that the mechanism shown by their evidence is not masked before establishing causation. The absence of masking phenomena, hence, would lead them to believe that there should be also a difference-making relationship between the cause and the effect. In addition, the absence of a difference-making relationship associated with the mechanism under study would lead social scientists to look for further counteracting mechanisms.

Reiss's masking example, once examined in depth, does not seem to be a case where the two causal concepts do not overlap. But what about causal relationships where the causes or the effects are omissions?

Regarding omissions that cause something, while it is easy to find difference-making evidence in support of a causal claim such as "the omission X causes Y", it is impossible to find processes linking omissions or absences and phenomena. From this, it would follow that there are some situations where the notion of cause as a difference-making relationship does not have a corresponding notion of cause as a mechanistic relationship. Reiss focused on the social sciences, but it should be noted that even in the health sciences there are several cases involving absences or omissions (for instance failing to treat a disease can cause the patient's death). If Reiss were right, consequently, this would be a problem for RWT in general.

As Williamson (2018) has recently discussed, however, there is a solution to this problem that can be applied to both these fields: in the case of absences or omissions, it can be argued that, if things had been different and *if the cause had been present*, it would have

been possible to find a mechanism between the present cause and the putative effect. For instance, diligence in civil law could, through a causal process, cause a positive effect on other people, while treating a disease could cause the recovery of the patient. Hence, even in the case of causally relevant omissions, where actual mechanisms cannot be found, it should be easy for scientists to establish the existence of counterfactual mechanisms.

As concerns situations where the omission is not the cause, but the effect, the solution might be different. Let us consider some cases: the presence of an authority figure can cause in a person a lack of moral reasoning, failing a treatment might cause the absence of coagulation. In such cases, scientists might explore counterfactual mechanisms as suggested by Williamson (2018). This strategy, however, might involve complex transformations where, in order to consider real effects, the causes themselves have to be turned into omissions or absences. Alternatively, researchers can turn such omissions and absences into the presence of a phenomenon. For instance, the lack of moral reasoning caused by the presence of an authority figure can be described as a situation of total obedience to authority (the social psychologist Milgram called this condition the 'agentic state' (Milgram, 1974)). Hemophilia, moreover, is known as the condition in which a person's blood does not clot properly. This, hence, would allow scientists to establish the presence both of a difference-making relationship and of a mechanism between the cause and the effect.

In general, causal relationships where omissions and absences are either the causes or the effects, allow scientists to collect, if not evidence of an existing causal mechanism, evidence of a counterfactual causal mechanism. It might be argued that particular cases do not enable scientists to collect real or counterfactual mechanistic evidence. In most of the cases, however, for a difference-making link, scientists can find a corresponding existing or counterfactual mechanism. The situations where this is not possible are so sporadic that it can be just claimed that, *normally*, causal concepts are overlapping.

3.4.4 Overcoming conceptual pluralism and conceptual monism

Two criticisms need to be addressed in this section. The first criticism is associated with a defence of conceptual pluralism and was proposed by Reiss in his analysis of RWT. The second criticism, supported by Claveau, is based on the assumption that conceptual monism can accommodate the use of a plurality of evidence.

Let us consider conceptual pluralism first. My response to Reiss' discussion is that RWT is to be preferred to conceptual pluralism not only because it is not true that RWT does not work, but also because this is how social scientists use causal evidence to establish just one causal claim. In the social sciences, discussions on causality often start with the general consideration that "correlation does not imply causation", and that the existence of a correlation and the identification of a mechanism support the same causal claim. A leading proponent of this idea was Elster, who claimed that:

> "Causal explanations must be distinguished from assertions about correlation. Sometimes we are in a position to say that an event of a certain type is invariably or usually followed by an event of another kind. This does not allow us to say that events of the first type cause events of the second" (Elster, 1989, p. 5)

More recently, George and Bennet discussed a very similar idea:

> "Case studies can also explore the possible causal mechanisms behind the correlations or patterns observed in statistical studies, providing a check on whether correlations are spurious or potentially causal and adding details on how hypothesised causal mechanisms operate." (George & Bennett, 2005, p. 34)

The limitations of correlations were discussed also by Elsenbroich, who suggested that:

> "Seeing the social sciences as concerned with mechanisms means to not allow "black-box explanations" such as statistical correlations. Although statistical correlations *can be used as evidence for causal associations*, they are not an explanation in themselves as they do not lay open the "cogs and wheels" operating to produce the phenomenon in question." (Elsenbroich, 2012, p. 5, emphasis added)

Finally, a very similar point was proposed also by Friedrichs a couple of years ago:

> "Without an identifiable mechanism, covering laws and statistical correlations […] cannot be interpreted as causal." (Friedrichs, 2016, p. 4)

These and similar claims appear in contrast with Reiss' descriptive argument according to which each causal method *defines its own concept* of causation. In general, social

scientists disagree with the idea that "X causes Y" can be interpreted just as "there is a correlation between X and Y", nothing more. Similarly, their concern about masking suggests that, for many of them, it is not enough to establish the presence of a mechanism to establish causation. From a descriptive stance, hence, RWT is a feasible and accepted approach.

From a normative perspective, furthermore, Reiss rejected RWT because the overlap between the presence of mechanisms and of correlations is, in his view, just an empirical truth that needs to be confirmed a posteriori. As I argued in section 3.4.3, however, cases where causal concepts do not overlap are very rare, especially if we include the existence of counterfactual mechanisms. Given that Russo and Williamson do not exclude that, in some cases, RWT might be impossible, it can be concluded that, in normal cases, there are not reasons to reject it.

Furthermore, even if RWT were proved to be impossible, this would not allow for the conclusion that conceptual pluralism is the right solution in causal studies. Alternative accounts, such as monism (as described by Claveau), could be defended. It would be hence necessary to show why conceptual pluralism should be preferred. Similarly, the claim that social scientists already use different concepts of causation separately would require more examples: it is not clear what made Reiss believe that we should consider as ambiguous all the cases where social scientists seem to use a single notion of causation and different types of evidence. RWT, indeed, would fit perfectly those situations.

Finally, Reiss claimed that the type of causal evidence to use should be determined on the basis of the purposes: prediction, for instance, is likely to require just a probabilistic account of causation (2009, p. 36). It might be accepted that there are situations (especially outside academia) where correlations are used alone to predict phenomena: consumer behaviours, for instance, can be predicted just analysing behavioural data, without establishing the presence of a mechanism. It could be added, however, that in such cases, researchers *do not need to use causal concepts at all*: probabilistic evidence is used to support a (non-causal) claim according to which an event X will happen in the future.

It is now time to turn towards conceptual monism. The consideration proposed by Claveau, as the author himself recognised, does not reject the use of different types of evidence within a single study. Rather, Claveau argued that both difference-making and

mechanistic evidence can be used to support a single difference-making causal claim. His account allows RWT because, it is claimed, difference-making methods are not sufficiently reliable to produce evidence free from bias. Only if these methods improved, social scientists could stop using evidence of mechanisms. This, however, appears to be only a remote possibility, in Claveau's words:

> "it would be an anomaly for the counterfactual-manipulationist account if other (mechanistic) evidence for a specific claim was demanded even though this same claim was already backed by the best difference-making evidence. This proposition has however little practical relevance since difference-making evidence is perhaps never of the 'perfectly adequate' type" (Claveau, 2012, p. 809)

As already explained, RWT is not associated with a specific notion of causation. Both Russo and Williamson's epistemic proposal, and Claveau's justification based on an existing form of conceptual monism can hence be accepted as possible causal accounts that allow scientists to embrace RWT.

3.4.5 Descriptive and normative RWT in the social sciences

In the previous section I have provided some evidence in support of the claim that social scientists comply with RWT when establishing causation. As I discussed in section 3.2.1, however, RWT in the health sciences is not considered just a descriptive account. The normative dimension of the thesis is a crucial aspect that Russo and Williamson (2007) highlighted several times. This dimension appears particularly relevant also in the social sciences, especially given that most of the criticisms I have discussed above targeted it.

As I discussed above, the criticisms of the normative dimension of RWT can be rejected. Moreover, the examination above of some methodological discussions in the social sciences has clarified that social scientists are well aware of problems such as confounding, overfitting and masking (Clark & Golder, 2015; Elster, 1993; George & Bennett, 2005; Steel, 2008). Due to the presence of such problems in the social sciences, I argue that the normative form of RWT applies to causal studies in the social sciences.

Like in the health sciences, RWT is capable of discriminating between strong and weak causal knowledge, and can guide social scientists in determining in which circumstances a particular causal claim is properly established. The combination of evidence to support

the claim “there is a correlation between the cause and the effect”, and evidence to support the claim “there is a mechanism linking the cause and the effect”, indeed, reduces the possibility of bias that might lead to inferential mistakes and helps researchers to establish more reliable causal claims.

3.5 RWT and big data

Once it is accepted that RWT can be applied to the social sciences, it remains to explore how it can improve causal data-driven studies tackling some of the problems described in chapter 2. As I mentioned above, establishing the presence of a causal mechanism can allow researchers to uncover bias such as incorrect causal relationships produced by Simpson’s paradox or by low-quality of data. Some examples can help to illustrate this idea.

Let us start with Simpson’s paradox. Over the last few years, numerous events have increased the public awareness of disproportionate police violence against minorities in the U.S. This phenomenon, furthermore, has been linked to disparities in the justice system, and many protests have risen in the country to ask statistics about disparities against minorities. Suppose some researchers decide to collect information about trials in the U.S. and to create a large dataset containing information about the type of assault, the race of the person accused of assault and the outcome of the trial (convicted or not convicted). Researchers analyse data about thousands of trials and discover not only the absence of discriminations against minorities, but also the presence of discriminations against white people in all types of assault. Before establishing a causal relationship between being white and having a high possibility of being convicted, however, they decide to look for possible causal mechanisms that might explain such a correlation.

Through the use of mechanistic studies, they try to spot any evidence showing discriminations towards white people, but they do not find sufficient evidence to establish that there is a causal mechanism by which being white greatly increases the chance of conviction. However, they collect interesting evidence in support both of i) a mechanism of discrimination towards minority and ii) a mechanism by which having a white victim raises the chance of conviction. Due to these observations, researchers decide to add a new variable in their database: the race of the victim. This finally allows researchers to obtain an interesting result: it is right that white defendants have a higher overall conviction rate, but the reason is only that they mostly have white victims, and having a

white victim causes an increase in the possibility of conviction. If the overall group is divided into two subgroups (one in which the victim is part of a minority, and one in which the victim is white), it is possible to conclude that white defendants always have a lower conviction rate. In such a case, by collecting further evidence in support of these discrimination mechanisms, researchers could establish both the presence of causal mechanisms and the presence of correlation. These claims, consequently, would allow researchers to establish that being part of a minority is causally linked to a great chance of conviction.

In this and similar cases, the RWT requirement to establish both correlation and mechanism can help researchers to spot problems linked to Simpson's paradox. This idea has been proposed also by Kievit et al. (2013), who argued that:

> "to be able to draw causal conclusions, we must know what the underlying causal mechanisms of the observed patterns are, and to what extent the data we observe are informative about these mechanisms." (Kievit et al., 2013, pp. 2–3)

and that the problems associated with Simpson's paradox can be avoided:

> "by explicitly proposing a mechanism, determining at which level it is presumed to operate (between groups, within groups, within people), and then carefully assessing whether the explanatory level at which the data were collected aligns with the explanatory level of the proposed mechanism" (Kievit et al., 2013, p. 6)

Another possible aspect that might cause biased results in a data-driven study is the low quality of data. This problem appears particularly relevant if applied to social networks data, which allow researchers to obtain massive amounts of information about a population that might be unrepresentative of the overall population and that might act differently online and offline. An illustrative case might be a study based on Twitter data whose aim is to understand how living in urban areas influences well-being. Let us suppose researchers collect thousands of geotagged tweets in two months, and identify some key terms linked to happiness such as 'smile', 'happy', 'rainbow' and 'love'. This study allows them to state that living in urban areas positively influences happiness, if compared to the level of wellbeing in the countryside.

The search for a mechanism responsible for such a correlation, nevertheless, leads to conflicting evidence: researchers collect numerous information in support of a mechanism whereby living in urban areas increases the level of stress and causes a decrease of happiness. Consequently, researchers decide to search for further mechanisms that could mask the effect of that mechanism, causing the positive effect of living in urban areas, but decide also to examine more in detail the methods and data used in their study. In such a way they discover that the majority of Twitter users live in urban areas, with the consequence that people living in the countryside are unrepresented in the sample population. Furthermore, researchers might conclude that their method for identifying the inherent 'happiness' of a tweet is characterised by substantial errors (such as the inability to recognise sarcasm). While such a case is just a fictitious example, there are real causal studies that have been criticised for the low quality of the data used. To give an example, a Twitter data-driven study similar to the one described here has been published by Mitchell et al. (2013), and has been often criticised because of its unrepresentative sample population (see for instance Jensen, 2017; Malik et al., 2015).

3.6 Conclusion

In this chapter, I have argued that RWT is one of the solutions that social scientists can adopt to test the adequacy of data-driven results and to recognise bias in statistical analyses. To support the use of RWT, I have responded to the main criticisms concerning its application in the social sciences. I have identified four general issues discussed in the literature: i) RWT is not used in the social sciences, ii) RWT does not work, iii) causal concepts do not overlap and iv) there are better alternatives to RWT.

I have shown that, in general, RWT is used successfully in the social sciences, and that it is supposed to reduce inferential mistakes, not to eliminate all of them. Next, I have explained that there are cases where, only prima facie, causal concepts do not overlap (like in the case of masking or in some cases of omissions). As for the alternatives to RWT, I have argued that Reiss's argument is not sufficient to justify the claim that social scientists should embrace conceptual pluralism rather than RWT. Furthermore, I have highlighted that Claveau's defence of monism does not rule out the possibility of using different types of evidence, therefore it can be considered a possible account for RWT. Finally, I have provided two examples in which RWT can allow for the recognition of errors in data-driven results.

4 Thick and big data: learning from mixed methods research

4.1 Introduction

The emergence of big data has in general been associated with the development of statistical tools able to identify causal relationships between quantitative variables. In the social sciences, characterised by a long tradition of both quantitative and qualitative approaches, this tendency has caused different reactions.

As described in chapter 2, one of the emergent ideas is that studies based on big data cannot account for the importance of contextual, qualitative information in the social sciences: big data are in general analysed quantitatively, without considering the complex social contexts from which such data are collected. This idea has led some researchers to discuss the role played by ethnographic, qualitative data in causal studies, and to propose the notion of thick data in analogy with big data (see for instance Blok & Pedersen, 2014). Interestingly, thick and big data have not been dichotomised as conflicting forms of data: social scientists have argued for the equal importance of big and thick data, and for their combined use, especially when the aim is to uncover causal phenomena (Blok & Pedersen, 2014; Bornakke & Due, 2018; Curran, 2013; Ford, 2016; Latzko-Toth et al., 2016). The underlying assumptions is that, while big data, mainly quantitative, can help to obtain statistical significance, to generalise a claim and to recognise the relevance of a phenomenon within a population; thick data, mainly qualitative, can be used to obtain contextual insights into a phenomenon.

The proposal for a combined use of big and thick data goes in the same direction as the recent mixed methods research (MMR) approach, according to which qualitative and qualitative methods should be mixed to strengthen their results. However, while MMR researchers have spent the last few decades trying to clarify the ways in which quantitative and qualitative data can be combined and can complement each other, proponents of thick data have not clarified yet how thick data could improve data-driven studies. For this reason, in this chapter I shall examine the discussions in the MMR literature to find arguments in favour of the combination of quantitative (big) and qualitative (thick) data[10]. The underlying assumption guiding this examination is that, even if the terminology used

[10] The material discussed in this chapter has been partly published in Ghiara (2019).

is sometimes different, some of the MMR discussions could be applied to the big/thick data debate to clarify how thick data can enhance data-driven studies.

The chapter is organised as follows: in section 4.2, I shall describe how MMR is conducted and why MMR discussions can be used to examine the role of thick data in data-driven studies. In section 4.3 I shall argue that the study of both quantitative and qualitative data can often lead to the identification of a correlation and of a causal mechanism linking the putative cause and effect, as required by RWT. In this sense, hence, qualitative contextual data are used to gather evidence of a causal mechanism. In section 4.4 I shall observe that quantitative and qualitative data can be used together to combine different ontological categories and different epistemological perspectives. In section 4.5 I shall clarify that, from MMR discussions, it is possible to identify three different ways in which thick data can improve data-driven studies: thick data, when used with big data, can i) offer evidence of causal mechanisms, ii) explore ontological categories otherwise difficult to study just with big data and iii) cast light on some dimensions of social phenomena that would be invisible in typical big data studies. Section 4.6, finally, will conclude.

The aim of the chapter is not to discuss the philosophical literature on pluralism and on case-based research, but to shed light on some philosophical insights that can be developed by considering methodological and conceptual discussions within the MMR community. A growing body of philosophical works on such topics is reaching social scientists and is discussed within scientific communities.

For instance, philosophers such as Longino (1990), Giere (2006), Mitchell (2013), Chang (2014), and Ankeny and Leonelli (2016) proposed similar positions arguing that scientific research is in general based on various forms of pluralism about theories and methods, and the numerous references to their works clearly illustrate the impact of their philosophical analyses. In addition, philosophical analyses of how scientific generalisations can be based upon observational studies, such as those proposed by Morgan (2012) and Boumans (2015), have brought to the fore the question whether local cases can be used to establish general claims.

While I share with all these authors the idea that questions about paradigms and case-based research have an important philosophical dimension that has to be further investigated, in my chapter I will take a different approach. My aim will be to inform the

MMR debate casting light on how social scientists are using their quantitative and qualitative data. This, in turn, will help me to provide a solution to the second challenge proposed in chapter 2, namely how to understand the role of thick data in causal studies based on big data.

This chapter will also help to bridge the gaps between i) social sciences' discussions on MMR, ii) social sciences' discussions on thick data and iii) philosophical discussions on causality (especially on RWT). On the one hand, it will establish that qualitative data in MMR, and thick data in new big data studies, if properly analysed, can help to recognise the presence of a causal mechanism, while quantitative data or big data are generally used to identify correlations between variables. On the other hand, it will claim that thick data, like qualitative data in MMR, can help researchers to develop a more comprehensive image of causal phenomena by offering insights from different ontological or epistemological perspectives.

4.2 Mixed Methods Research

Mixed methods research is the combination and integration of quantitative and qualitative methods in the same study. This type of study is not new in the social sciences (Maxwell, 2016), and some of the earlier MMR studies are still considered prominent MMR examples. A case in point is the Hawthorne studies that explored whether people work more efficiently under various environmental conditions (Mayo, 1993; Roethlisberger & Dickson, 1964). By performing some studies at the Western Electric's factory at Hawthorne, a suburb of Chicago, researchers combined quantitative data coming from statistical analyses and qualitative data gathered through interviews and observational studies to analyse the factors affecting productivity. Researchers concluded that the workers' productivity was not only affected by the changes in physical conditions and in the workplace, but also by the awareness of being observed.

Another similar example can be found in the Lynds' Middletown studies that examined the ways in which the white population of a typical American city changed between 1890 and 1925 (Lynd & Lynd, 1929, 1937). To understand how the economic growth and the process of industrialization influenced the citizens' life in an 'average' mid-size American city, the authors conducted both interviews and statistical studies. This allowed them to identify some aspects of the daily life that changed due to an emerging 'consumer culture'.

Both the Hawthorne and the Middletown studies are still cited as good exemplars of MMR studies (Biddle & Schafft, 2015; Denscombe, 2008; Hunter & Brewer, 2003). Despite this long tradition, however, it was only in the last few decades that social scientists began to consider MMR as a research methodology with a recognised name, a distinct identity and specific debates (Denscombe, 2008). In particular, the recent MMR literature has often discussed the possible ways in which quantitative and qualitative methods can be mixed. Tashakkori and Teddlie (2003), for instance, identified nearly 40 types of MMR strategies that can be distinguished according to the way in which data are gathered, the relevance given to different types of methods, the purpose, and the priority assigned to diverse inferences coming from particular types of data.

Among these MMR strategies, one of the most common approaches entails the collection of both quantitative and qualitative data. In this case, MMR studies can be characterised either by the concurrent use of data collection methods (employed to gather data of different types simultaneously), or by the chronological or sequential use of different data-gathering methods. The methods used to collect data, in addition, are in general associated with different types of data analysis. Generally, quantitative data are examined by means of quantitative, statistical methods, while qualitative data are explored qualitatively to clarify the insights obtained from contextual information.

An illustrating example is the Child STEPS Effectiveness Trial (CSET) carried out by the Research Network on Youth Mental Health (Chorpita et al., 2013) and described in detail by Palinkas (2014). The project was focused on children aged between 8-13 who had been referred for treatment of problems involving depression, disruptive conduct, anxiety, or any combination of these. CSET involved ten clinical service organizations in Honolulu and Boston, 84 therapists, and 174 children. Therapists were assigned randomly to one of three groups: in the first group, children were treated with the usual treatment procedures; in the second and third groups, youth participants were treated with at least one of the following treatments: cognitive-behavioural therapy (CBT) for anxiety, CBT for depression, and behavioural parent training (BPT) for conduct problems. However, while in the second group therapists used full treatment manuals, in the third group they used a modular approach, which allowed them to develop targeted practices for each child by using only some of the treatments' components. Children who met the study criteria were then assigned randomly to one of these three groups of therapists.

Quantitative data were collected from the trials and studied through regression analyses. The results showed that modular treatments had significantly superior outcomes in improving children's behavioural health than standard treatment and usual care. In other words, the quantitative approach offered evidence of a strong correlation between modular treatments and improvements in children's depression, disruptive conduct or anxiety. These results were used to inform some qualitative studies based on semi-structured interviews and focus groups. These studies involved 38 therapists and were designed by following the quantitative findings obtained using regression analyses.

Qualitative findings confirmed the correlation and provided causal evidence regarding some processes through which modular treatments can improve children's behavioural health more than standard treatment and usual care. For instance, qualitative data cast light on the fact that protocols often did not meet the needs of the children and their parents, whereas modular treatments allowed therapists to identify with children the most immediate issues to address. This also helped to avoid treatments that children and parents would have been unable or unwilling to perform due to cultural and logistical reasons. Furthermore, qualitative approaches helped researchers to obtain detailed information about the therapists' perception of their role and the relations between their convictions and their professional activities. Because of the therapists' belief in the importance of flexibility, for instance, it was found out that modular treatments increased the therapists' motivation. Finally, modular treatments, the therapists argued, improved the therapeutic alliance with the children, with positive consequences at the psychological level (Palinkas et al., 2013).

In this example and in similar causal MMR studies, quantitative and qualitative data appear to have different roles, and this diversity resembles the diversity that social scientists recognise between big and thick data. Quantitative data are numerical information used to identify correlations between variables, qualitative data contain contextual information that helps researchers to explore how the correlations emerged. The relevance of MMR for the thick data discussion, however, does not depend only on the fact that the quantitative/qualitative distinction is very similar to the big/thick data distinction. A more important parallel between the MMR discussions and the debates on thick data can be found by observing that in both cases the general aim is to convince quantitative researchers of the importance of qualitative data and approaches. MMR debates have emerged to defend the crucial role played by qualitative research in the

social sciences (Karasz & Singelis, 2009). For this reason, several discussions have been focused on the ways in which qualitative methods can advance quantitative studies in the social sciences. Likewise, proponents of thick data aim to cast light on the relevance of this form of information to counter the growing tendency to carry out purely quantitative big data studies (Latzko-Toth et al., 2016, p. 200).

Furthermore, another aspect in support of the parallel between MMR discussions and the big/thick data debate is the fact that MMR researchers often refer to thick data and thick descriptions when they describe their qualitative findings. Williams (2013), for instance, described her research on career changes among people studying education in Australia and observed that:

> "A mixed methods approach was chosen as the most appropriate to answer my question. This would enable a picture of career change student teachers as a group to emerge from quantitative data, while the particular narratives and experiences of individual career changers […] would be gained through a more in-depth rich, 'thick' data set" (Williams, 2013, p. 19)

Similar claims can be found also in Robinson and Medelson's MMR study on multimedia content and the impact of online features on viewers' attitudes (2012) and in Martin's research on the development of affective skills in nursing students (2017).

Together, these observations lead to the suggestion that MMR debates can offer useful insights to conceptualise the role that thick data might play in data-driven studies.

4.3 Qualitative and thick data to establish causal mechanisms

In MMR, one of the most common ideas proposed to support the importance of qualitative data is the claim according to which such data are needed in causal studies to identify causal mechanisms. To give an example, Weiss (1995) claimed that:

> "quantitative studies support an assertion of causation by showing a correlation between an earlier event and a subsequent event. An analysis of data collected in a large-scale sample survey might, for example, show that there is a correlation between the level of the wife's education and the presence of a companionable marriage. In qualitative studies we would look for a process through which the wife's education or factors associated with

her education express themselves in marital interaction." (Weiss, 1995, p. 179)

More recently, Maxwell (2004a) discussed the role of different methods in education research and argued that:

> "qualitative methods have distinct advantages for identifying the influence of contextual factors that can't be statistically or experimentally controlled, for understanding the unique processes at work in specific situations" (Maxwell, 2004a, p. 9)

The MMR literature is characterised by numerous similar claims, according to which qualitative approaches contribute to the identification of causal mechanisms or processes responsible for social phenomena (see also Cyr & Mahoney, 2016, p. 442; Fearon & Laitin, 2008, p. 1167; Yoshikawa et al., 2013, pp. 8–9).

A parallel claim could be made in the case of thick data. Section 2.4.1 of chapter 2 has clarified that in the literature thick data are defined as mostly (if not only) qualitative. This, I explained, depends on the fact that the idea of thick data derives from the discussions on thick descriptions, that were claimed to be obtained through qualitative data such as interviews and field observations (Geertz, 1973; Ryle, 1968). It could be suggested, furthermore, that the dense information that, according to the thick data proponents, are provided by such data, could hardly be contained in numerical data. The qualitative nature of thick data allows for the parallel between what is said in MMR, and what could be said for thick data: in both cases, indeed, researchers consider qualitative information and its role in causal studies. Maxwell (2004b), for instance, by exploring how qualitative data can be used in causal studies, discussed the notion of rich data, that resonates with the notion of thick data, and claimed that:

> "Rich data […] are data that are detailed and varied enough that they provide a full and revealing picture of what is going on and of the processes involved (Becker 1970:51ff.). In the same way that a detailed, chronological description of a physical process (e.g., of waves washing away a sand castle or the observations of patient falls described above) often reveals many of the causal mechanisms at work, a similar description of a social setting or event can reveal many of the causal processes taking place." (Maxwell, 2004b, p. 254)

According to this position, the qualitative nature of some data in MMR and of thick data can help researchers to identify operating causal mechanisms responsible for specific causal phenomena. The contextual dimension of thick data, moreover, is consistent with some of the debates about the study of causal mechanisms in the social sciences. As argued by some researchers, indeed, contextual knowledge of social environments can be crucial to uncover causal mechanisms (Dalkin et al. , 2015; Falleti & Lynch, 2009; Steel, 2004). Notably, one of the first examples used to describe the identification of a causal mechanism in the social sciences was the study conducted by the anthropologist Bronislaw Malinowski (1935) to explore whether the number of wives was causally liked to prosperity among Trobriand chiefs. As recognised by Steel, (2004), Malinowski managed to identify the causal mechanism responsible for the correlations between the number of wives and wealth by conducting observational studies and by living in close contact with the Trobriand people. The collection of what could be called 'thick information', helped Malinowski to identify two social practices: (i) brothers were required to donate generous gifts of yams to the households of their married sisters (when a sister married a chief, moreover, the gift was larger than usual); (ii) yams were the principal means by which a chief financed his political project. This information, consequently, allowed Malinowski to establish the presence of a causal process according to which the possession of more wives caused the ownership of more yams; and, in virtue of the practice, a chief with more yams was likely to have a greater influence among his society.

Both qualitative data in MMR and thick data, consequently, can be associated with the collection of evidence of mechanisms. Nevertheless, while there is a general agreement that this type of evidence is important for causal studies, its relevance varies according to whether it is claimed that *it would be better* to establish the presence of a causal mechanism, or that *it is necessary* to establish such a presence in causal studies.

In the former case it might be argued that, by providing evidence of causal mechanisms, qualitative data in MMR and thick data in big data studies help to develop more comprehensive causal explanations of social phenomena. This idea is sometimes discussed in the MMR literature, where it is claimed that the combination of quantitative and qualitative methods provides richer and more complete causal explanations (Shaffer, 2018, p. 11). In this sense, the overall idea is that qualitative data offer further information

about causal relationships that can be established also collecting just correlational evidence.

In the latter case, when qualitative methods make available evidence of mechanisms otherwise not accessible, the role of qualitative data is claimed to be essential in causal studies. There are several discussions in the MMR literature that support such a position. Libman (2012), for instance, argued that, in political science, researchers are in general sceptical about the use of pure statistical or econometric studies of causal phenomena. For this reason, the mixture of quantitative and qualitative approaches is essential in order to:

> "describe a possible causal mechanism, which would explain the link between variables, and also provide at least partial evidence in favor of this mechanism. This evidence can be derived from qualitative research […] econometrics as such is considered as merely an indication of possible existence of causal effects, but not as proof" (Libman, 2012, pp. 179–180)

Similar ideas were suggested also by Trinchero (2014) and Barnes and Weller (2017), who discussed the aim of MMR and considered that:

> "If the research objective refers to establish the existence of a causal relationship between factors (as in the hypothesis "Motivation to learn can improve scholastic achievement") […] quantitative experimental design can be useful to quantify the increase of scholastic achievement probably due to a corresponding increase of motivation, but how can we demonstrate that increase of scholastic achievement is really due to increase of motivation and is not imputable to other extraneous factors? The only way is to study the dynamics that led an increase of motivation to promote an increase of achievement. This is possible by a persistent engagement of the researcher in a qualitative in-depth inquiry on the field in which the experiment was performed, to reveal the processes underlying the causal relationship and the conditions under which this relationship may occur." (Trinchero, 2014, p. 47)

> "Instead of using each method to provide separate evidence of the X/Y relationship, it [mixed methods research] envisages a division of labor among methods in which the large-N work estimates causal effects whereas the smaller-N work probes the validity of the large-N work's underlying

> assumptions. Here, combining methods improves our confidence in the underlying X/Y relationship not because the findings of the large-N and small-N analyses converge but rather because the small-N work demonstrates that the large-N work satisfies the requisites for causal inference" (Barnes & Weller, 2017, p. 1021)

According to this position, evidence of mechanisms collected via qualitative methods in MMR studies or through thick data is crucial in order to establish causal relationships. This claim is very much in line with what I have argued in chapters 2 and 3: data-driven studies based only on quantitative analyses are likely to produce biased results. For this reason, even in the era of big data, causal studies in the social sciences should comply with RWT and, in order to establish causation, should collect both evidence of a correlation between the cause and the effect and evidence of a mechanism linking them. As observed in chapter 3, RWT does not say anything about the evidence-gathering methods that should be used to establish causal relationships. It is hence feasible that social scientists obtain the required evidence by analysing qualitative or thick data.

This consideration casts light on the first way in which thick data can improve data-driven studies: if it is accepted that establishing causal mechanisms helps to test the adequacy of data-driven results and to recognise bias in statistical analyses, thick data can be claimed to provide evidence of causal mechanisms otherwise inaccessible algorithmically.

4.4 Pluralistic debates in MMR

Some of the pluralistic debates in MMR help to clarify further ways in which qualitative data, and eventually thick data, can improve causal studies. As said above, MMR is claimed to be based on the mixture of quantitative and qualitative methods within a single study. Many authors, however, have recognised that although in some cases MMR studies mix just methods (such as structured interviews and questionnaires), there are more complex situations where MMR studies are characterised by a mixture of different (and not always explicit) ontological or epistemological assumptions.

Such discussions in MMR have often been shaped in terms of 'mixing different paradigms', where the term 'paradigms' is used to denote something very similar to the 'worldviews' populating the social sciences that were first identified a few decades ago

by Guba and Lincoln (1994)[11]. Guba and Lincoln saw research paradigms as intrinsically associated with ontological, epistemological and methodological assumptions. When scientists embrace one paradigm, such assumptions provide an interpretative framework, or worldview, that defines the nature of the 'world', the possible relationships to the world, and the limits of legitimate inquiry (Guba & Lincoln, 1994, p. 108). In this context, MMR researchers have argued that the combination of quantitative and qualitative methods allows for the combination of different ontological or epistemological stances. This idea has emerged in different ways in the MMR literature.

The debate about the dominant/less-dominant and the equal status designs in MMR is probably the most visible illustration of this idea. According to these discussions, the dominance of one method and one type of data in MMR can be reflected in the presence of a dominant 'paradigm' in the study. The dominant/less-dominant design is described as the situation in which, while conducting MMR studies, social scientists rely on a specific system of ontological and epistemological assumptions, while other systems (or 'paradigms') are used to examine only small components of the whole studies. On the contrary, the MMR equal status design is referred to the situations where more ontological and epistemological stances, recognised as different paradigms, are considered equally relevant to a study (Creswell, 1994, p. 35; Tashakkori & Teddlie, 2003, p. 435).

In addition, many MMR researchers have discussed the reasons why the combination of different ontological and epistemological assumptions could improve MMR studies. Greene (2006) considered such a combination as the main difference between mixed methods and multimethod studies, and argued that it is only through divergence and dissonance that MMR researchers can examine taken-for-granted understanding and problematise findings associated with a single ontological or epistemological perspective. Uprichard and Dawney (2016) took a similar position and observed that only a 'diffractive' approach in MMR, that allows researchers to analyse phenomena through

[11] Although this is not the aim of the chapter, it is worth mentioning that the discussion on paradigms proposed in the MMR community is particularly interesting because several MMR proponents have supported different interpretations of philosophical works such as those of Kuhn (1970). Johnson and Onwuegbuzie (2004), for example, started from Kuhn's definition of paradigm to develop their notion of research paradigms as systems of ontological, epistemological, axiological and methodological beliefs, values, and assumptions that a community of researchers shares for what concerns the nature and conduct of research (see Johnson & Onwuegbuzie, 2004, p. 24). In addition, Morgan (2007) emphasized the way in which the Kuhnian notion of paradigm shifts could enhance our understanding of the changes occurring within the social science; and Denscombe (2008) claimed that Kuhn's notion of paradigms as "communities of practice", unlike many other concepts, manage to accommodate both the pragmatist underpinnings of MMR and the divergencies and diversity characterizing the MMR community.

different (ontological and epistemological) pathways, can lead researchers to fully understand complex social phenomena. Similarly, Johnson (2012) advocated what he called a 'dialectical pluralism' in MMR. He argued that such a metaparadigm would help researchers to appropriately listen to each research questions and purposes, and this would allow for the development of new workable wholes to study social phenomena.

In general, these accounts are based on the idea that, in MMR, qualitative data enrich quantitative studies by shedding light on new unexplored perspectives. This is possible because, it is argued, qualitative and quantitative data are in general both collected and analysed via methods that entail different ontological and epistemological stances. There is an important caveat to this argument. The mixture of different philosophical standpoints in MMR might not lead to a clear integration: assumptions could be too different to be fully merged. This does not mean, however, that MMR researchers should avoid mixing ontological and epistemological accounts. To problematise findings and fully explore social phenomena, MMR researchers have argued, it is important to give voice to paradigmatic dissonances. For this reason, even in the case or irreducible divergences, qualitative and quantitative data should be combined (Johnson, 2012; Uprichard & Dawney, 2016).

To explore whether such discussions can help to recognise other possible ways in which thick data might advance data-driven studies, the following sections will examine specific real-life MMR studies where ontological categories and epistemological perspectives were mixed. The analysis of such studies will help to clarify how the combination of big and thick data could produce similar results.

4.4.1 Combining ontological categories

In particular circumstances, the mixture of quantitative and qualitative data entails the combination of different ontological categories. Let us imagine we have a certain causal phenomenon we want to explore through a MMR study: probably, we will collect and analyse qualitative data to gain an in-depth understanding of the causal process that has led to the phenomenon, while we will collect and study quantitative data to verify whether the correlation between the putative causal factors and the effect is present in the context of study and, perhaps, even in a broader context. In other words, our qualitative data will enable us to claim that a single, complex, instance of causation exists; while our quantitative data will help us to understand how a causal phenomenon operates generally.

While for some researchers these are just considered two levels of causation that can be explored independently, according to others the difference between such notions is so great that they should be understood as two different categories of causation: one category of causation as a single, unique event, and one category of causation as a general relationship that can be found at the population level. The former category can be called 'singular causation', the latter category is known by the name of 'general causation' (for more discussions on this, see Johnson et al., 2017).

The difference between them is made evident in some discussions concerning general causations in the social realm. Discussing the possibility of finding general causation in the social world, Cronbach wrote:

> "Generalizations decay. At one time a conclusion describes the existing situation well, at a later time it accounts for rather little variance, and ultimately it is valid only as history. […] Propositions describing atoms and electrons have a long half-life, and the physical theorist can regard the processes in his word as steady. Rarely is a social or behavioural phenomenon isolated enough to have this steady-state property. Hence the explanation we live by will perhaps always remain partial […] and rather short lived." (Cronbach, 1975, pp. 122–123)

Similarly, Wolcott (1995) wrote:

> "How do you generalize from a qualitative study? [You] might answer candidly and succinctly, "You don't." That is a safe and accurate answer. It is the basis on which American anthropology was founded under Franz Boas. With an empiricism directed toward rigorous historical particularism, Boas insisted that no generalizations were warranted from the study of any particular society." (Wolcott, 1995, p. 163)

The same consideration has been proposed by Lincoln and Guba (2000), who have argued that social phenomena are characterised by local conditions that make it impossible to study general causation: the contextual differences from one situation to another are too relevant to allow for any form of generalisation. Generalising a causal explanation in the social sciences is, according to them, like crossing the same river twice.

What these authors have in common is the assumption that, when studying social phenomena, it is impossible to investigate the ontological category of general causation: unlike in the natural world, in the social realm this type of causation can seldom be found. In other words, the shared assumption is that in social phenomena causation does not operate *both* at the singular and general level, but *only* at the singular, local level.

In many cases, even though such assumptions and such categories are not discussed when the MMR studies are conducted, their presence is tangible. Let us consider Wood's study (2003) aimed at understanding why peasants in El Salvador decided to join rebel movements. To answer such a question, Wood decided to collect qualitative data by means of ethnographic interviews both with peasants and some members of the elite opposed to the peasants (i.e. landlords, military officers, government officials). The data collected from the interviews were explored to obtain a contextual understanding of the phenomenon through the personal and community histories of the individuals involved. The causal relationship explored in this way, hence, was completely embedded into a personal, social, political and historical system: as argued in the book, the decision to give a contribution to the insurgency was determined by specific personal and political circumstances, that led the peasants from different cultural backgrounds to fight for social justice. It was not a mere class conflict, and the civilian supporters were offered very few benefits during the insurgency (Wood, 2003, p. 158). Those who decided to join rebel movements were often motivated by the history of violence against family members and neighbours, by the brutality of the actions perpetrated by the Government and by personal, moral and affective benefits they received through participation (Wood, 2003, p. 229,235). The ontological category of causation addressed through the case studies corresponded to a complex phenomenon existing at the local level: as argued by Wood, indeed, the causal processes identified through her analysis were all *local*, and interacted in intricate ways during the period of the civil war (Wood, 2003, p. 236).

After having established the presence of this local and multifaceted causal relationship, Wood used the evidence obtained to develop a general hypothesis about the role that pleasure of agency could play in similar situations. Since qualitative data were not sufficient to address general causation, Wood obtained a set of quantitative data collected at the end of the war through a rural household survey, as well as elections data and some databases documenting the changes in agrarian property rights during and after the uprising to test her hypothesis according to which:

> "[…] long-subordinate social actors grant to the exercise of agency per se. To act in defiance of unjust authority, to claim recognition as equal subjects whose personhood needs be respected, to act effectively for the realization of essential interests, and to publicly assert the power of collective efficacy may be important reasons for the emergence of insurgent collective action elsewhere as well as in El Salvador" (Wood, 2003, p. 253)

Quantitative data helped her to work on a formal quantitative model based on the dynamics of a coordination (assurance) game. The study of general causation, finally, led her to establish a dialogue with researchers of collective action and social movements, aimed at developing general causal explanations of social phenomena.

Through the use of qualitative and quantitative data, Wood managed to explore the peasants' participation in rebel movements in El Salvador by means of two different ontological categories of causation. On the one hand, by exploring qualitative data, she developed a causal narrative of the complex interactions of personal, historical and political factors that generated the specific phenomenon under study. Such a narrative was consistent with the ontological category according to which causation is a single and complex phenomenon that exists locally. On the other hand, by studying quantitative data, she provided a causal explanation of the general phenomenon whereby repression and a sense of agency can foster mobilization. This idea was in line with general causation, based on the assumption that a causal phenomenon can operate generally, and that its analysis can be used to explain and predict other similar phenomena.

Wood's study was based on the combination of 'traditional' qualitative and quantitative data, with no references to big and thick data, but it is not difficult to imagine a similar study based on large datasets and thick data. Big data are in general used to uncover regularities and causal relationships that can be generalised from the sample population to the population level (Shmueli, 2017). On the contrary, thick data are collected to analyse the specificity of certain phenomena, and the contexts from which they emerged (Latzko-Toth et al., 2016, p. 202).

If we look for similar discussions in the thick data literature, it is rare to find clear claims. Marwick's observation on the way in which quantitative (big data) studies and qualitative (thick data) analyses can help to understand how people use Twitter (2013), is one of the

most explicit claims that goes in the same direction of my proposal. Indeed, Marwick argued that:

> "Quantitative studies often determine connections and networks […] based on statistics and numbers. Instead, qualitative research seeks to understand meaning-making, placing technology use into *specific* social contexts, places, and times." (Marwick, 2013, p. 119)

In addition, Wang's description of what happened with Nokia (2013) can offer some considerations about the use of different ontological categories in thick and big data studies. As described in chapter 2, Wang conducted ethnographic works in China for Nokia in 2009. She collected thick data providing evidence of a new phenomenon according to which low-income Chinese consumers were becoming willing to pay for more expensive smartphones. Nokia, however, did not find any similar pattern by analysing quantitatively its (big) data. It might be argued that the discrepancy emerged because Nokia was looking for a general causal pattern, while Wang was observing some contextual changes that, ultimately, led to the rise of a new phenomenon. The pattern explored by Wang, at the population level, was not statistically significant. In that specific context, however, such small changes triggered a causal chain that ended in a complete inversion of trend for the Chinese market.

If such considerations are accepted, it can be concluded that thick data, when combined with big data, allow researchers to mix different ontological categories such as the categories of singular and general causation.

4.4.2 Combining epistemological perspectives

The quantitative/qualitative distinction of the social sciences descends largely from the difference between the epistemological questions at the heart of quantitative and qualitative studies. Most of the quantitative studies are underpinned by a specific account of what should be accepted as warrantable knowledge: scientists should conduct studies in which the distance between the observer and the observed is preserved with the aim to maintain objectivity. On the contrary, numerous qualitative researchers are in general committed to observing the world from the point of view of the actor, rather than maintaining any distance. Observers should be close enough to the object of study to gain a contextual understanding of the phenomenon, and objectivity is ruled out due to this proximity and to the fact that any form of knowledge is gained within a social context.

It follows that, when quantitative and qualitative approaches are combined, different epistemological perspectives can be embodied within a single study. To clarify how this combination can be achieved in mixed methods studies, let us consider one case that has been discussed over the last few years: the MMR assessment and evaluation of the program Eat Right! Be Fit! recently developed by some researchers of the Midwestern University. As described by Greene (2015), for the evaluation of the program researchers have decided that they will employ two worldviews characterised by different epistemological accounts: one, known by the name of post-positivism, aspiring to generalisable results based on structured quantitative evidence; and one called constructivism, based on the assumption that meaningful understanding should be contextual (Guba & Lincoln, 1994). To comply with such worldviews, researchers will conduct both a quasi-experiment and mini-case studies informed by unstructured ethnographic observations. Both methods (and worldviews) will be employed to assess the implementation of the program and evaluate its effects.

Epistemologically, the quasi-experiment can assess the implementation of Eat Right! Be Fit! through the objective documentation of the nature and extent of activities done in the classrooms: each classroom schedule will be analysed, and the time spent doing physical activities will be systematically reported. Furthermore, researchers will collect data about how many children walk or bike to school, and the kinds of food eaten by them. The mini-case studies, instead, will aim to understand whether and how the program has an impact on the classroom rhythms and learning by means of contextual explorations into the daily life of children (Greene, 2015, p. 612).

The quasi-experiment, hence, will assess the quality of the program by considering only the program implementation, without taking into account the context in which the program is offered, and the reality perceived by the children involved in the program. On the contrary, these aspects will be at the centre of the mini-case studies, whose goal will be to capture how children perceive the program and its purposes.

Similar epistemological discrepancies between such methods can be observed also regarding the decisions about the evaluation process. On the one hand, the quasi-experiment study will collect standardised health measures such as the body mass index, that will be then compared to measures about the same children collected in the past, and to measures about children from the same area not involved in the program. The success

of the program, in this sense, is associated with the percentage of children obtaining specific values in some standard indicators of child health. The aim of the mini-case studies, on the other hand, will not be to capture the bodily changes obtained through the program, but to explore how children are engaged with it. To do so, researchers will interview children, teachers and parents to obtain insights into the way in which the program has an impact on the children's life (Greene, 2015, p. 614).

The epistemological assumptions behind these evaluation processes show substantial differences: the post-positivist worldview's aim is to collect objective and generalisable evidence of the success of the program, the constructivist worldview's goal is to develop an intersubjective understanding taking into account multiple realities such as those lived by the children and the teachers. Such differences could lead to divergent results: through the mini-case studies researchers might evaluate the program as successful, while the quasi-experiment might show the absence of bodily changes in the classrooms involved.

As described in section 4.4, however, the idea that MMR should intentionally allow for disharmony to reach unforeseen insights is now shared and supported by many MMR proponents. This, furthermore, appears in line with one of the main purposes of MMR, known by the name of 'initiation': MMR studies can be conducted to find out paradoxes and conflicts that lead to a reframing of the research questions. This goal is often described using Rossman and Wilson's words (1985, p. 633): "rather than seeking confirmatory evidence, this [initiation] design searches for the provocative".

Even in this case, it is difficult to find clear discussions of the epistemological perspectives used to study big and thick data. There are however some clues that seem to confirm the idea that the combination of big and thick data can entail the combination of different epistemological perspectives. To begin with, when describing the notion of rich data, that that resonates with the notion of thick data, Marx (2008) argued that:

> "Qualitative research is often described as having rich "thick description," in the words of Clifford Geertz. Rather than merely recording events, people, and places, thick description seeks to present and explore the multifaceted complexities of the situation being studied, the intentions and motivations of the actors involved, and the context of the situation." (Marx, 2008, p. 795)

Similarly, Schultze and Avital (2011) observed that:

> "For interviews to live up to their promise as a means of providing insights into people's life world and its meaning, the data that are generated need to be rich in such a way that they lend themselves to thick description […] qualitative studies include scenic details, participants' motivations and intentions, and the web of social relationships in which events happened and individuals took action. In other words, rich data are deeply implicated in bringing to life the human beings that are the center of social science research." (Schultze & Avital, 2011, p. 3)

Such claims cast light on the subjective dimension that can characterise thick data: thick data are not objective pieces of information, but data collected by observing the world from the point of view of the actors. It follows that, when combined with big data, they can give voice to different epistemological perspectives.

4.5 Conclusion: lessons from MMR

The discussions above show three different ways in which qualitative data can improve quantitative studies in MMR. First, qualitative data can be used to obtain evidence of the presence of causal mechanisms. Second, qualitative data can be used to explore ontological categories that would be difficult to study through quantitative data. Third, qualitative data can be collected and analysed from a different perspective with respect to how quantitative data are studied, therefore they can cast light on otherwise invisible dimensions of social phenomena.

As I proposed in section 4.2, there are several similarities between MMR discussions and thick/big data discussions. To begin with, in both cases researchers want to explore the role and importance of qualitative data: many MMR researchers come from qualitative traditions and explore fruitful combinations of qualitative and quantitative data; similarly, proponents of thick data are convinced of the importance of qualitative and observational research. What is more, in some occasions MMR discussions and the thick/big data distinction partly overlap: some MMR proponents, indeed, justify the use of qualitative approaches by claiming that they allow for the collection of 'rich' or 'thick' data (Martin, 2017; Robinson & Mendelson, 2012; Williams, 2013).

If these similarities between MMR discussions and thick/big data debates are accepted, then MMR discussions can be used to cast light on some ways in which thick data can

improve data-driven studies. Like in MMR, it can be suggested that thick data, if combined with big data, enhance causal research in the social sciences through three pathways.

First, thick data can be collected to find evidence supporting the presence of causal mechanisms. This possibility is often discussed in the MMR literature, where it is claimed that qualitative information can offer some insights into the mechanisms causing social phenomena. My suggestion is that thick data, thanks to their qualitative and contextual nature, can be easily used to find evidence of mechanisms. This use, furthermore, would be consistent with what I proposed in chapter 3: in order to overcome some limitations of big data studies, social scientists should establish the presence both of correlations and of causal mechanisms. In this context, thick data might help to comply with this requirement.

Second, thick data can help researchers to focus on ontological categories that otherwise could be neglected. In the case of ontological categories of causation, I have shown that qualitative and quantitative data are in general used for different purposes: quantitative data are studied to explore general causation, according to which casual relationships exist at the population level and can be generalised; while qualitative data are used to establish causal claims linked to the category of singular causation, according to which causal relationships are local and unique events. Thick data, due to their qualitative and contextual nature, might be used to address singular causation. Big data, thanks to their volume, are likely to be used for general causation, to study how social phenomena function at the general level, apart from particulars. For this reason, it can be claimed that thick data might improve data-driven studies by allowing for the development of a full understanding of social phenomena.

Third, another way in which thick data might help to develop a comprehensive understanding of social phenomena is by providing another perspective from which to analyse them. Quantitative data are in general collected and explored with the aim to maintain objectivity, and the same could be argued of big data. To begin with, a large amount of social data is now generated automatically through new devices: our Internet searchers can be transformed in data by means of new algorithms, our location data can be collected automatically from our mobile phone. Furthermore, also the analysis of big data is often conducted in an objective way, as it has been shown in the case of Bayesian Networks in chapter 2. Of course, both in the process of data collection and in the process

of data analysis numerous choices are made by technicians and researchers, and such choices are even more important when data are curated, as I have described in chapter 2 (see also Leonelli, 2012a, 2014b). Nevertheless, the overall aim, when researchers use quantitative data, is to maintain a distance from the object under study. There is no room for the subjective meaning that researchers or individuals in the study would confer to such an object.

Instead, thick data are collected exactly to gain a contextual understanding of phenomena, and are likely to be used to explore individuals' motivations and intentions (Marx, 2008; Schultze & Avital, 2011). Thanks to such data, researchers can obtain contextual insights, and can study how people perceive and interpret social phenomena. By adding thick data to data-driven studies, therefore, researchers might be able to develop a comprehensive understanding of social phenomena. As I described in section 4.4.2, furthermore, such a combination can help to explore further questions about the causal impact or the effectiveness of interventions.

Overall, the parallel with MMR discussions helps to identify *possible ways* to use thick data to improve causal studies based on big data. There might be cases where thick data are combined with big data for other reasons not described here, or cases where two or three of the possibilities discussed in this chapter are considered as parts of the same general goal in single studies based on big and thick data. Furthermore, given the young and partly underdeveloped nature of the thick data/big data distinction, it is likely that future uses and combinations of such data will be analysed within the social sciences in the future. The aim of the discussion proposed in this chapter was simply to shed light on the relevance of thick data in the era of big data, and to provide fertile ground for future studies on the combination of thick and big data.

Part II

Causal discoveries in practice

In Part I, chapters 3 and 4 have helped to provide a solution to two of the challenges guiding this thesis, namely how to overcome the limitations of causal data-driven studies; and how to understand the role of thick data in causal studies based on big data. The considerations motivating the proposed solutions, however, have been mainly theoretical: rather than observing existing research studies, I have drawn my conclusion from some conceptual issues discussed in the literature. In this part of the thesis, instead, such challenges will be explored through two case studies. Such case studies will be used both to argue in favour of some of the discussions offered in Part I and to prepare the ground for the proposal that will be discussed in chapter 8 of Part III.

On the one hand, the analysis of case studies will help to establish a strong relationship between the philosophical considerations proposed in Part I and the real practices in the social sciences. This is particularly relevant for chapter 5, based on the analysis of a MMR study. By discussing a real study in detail, it will be possible to argue that i) social scientists can comply with RWT, ii) qualitative data can provide evidence of mechanisms and iii) mixing qualitative and quantitative data can allow for the mixture of different ontological categories. These claims will hence support the discussions provided in chapter 3 and chapter 4.

On the other hand, the project described in chapter 6 will pave the way for a detailed discussion on the role of biomarkers. This discussion will be crucial, then, to develop the general notion of markers and the concept of sociomarkers in chapter 8. As it will be argued in chapter 8, by using big data to obtain sociomarkers, social scientists might collect new evidence of social mechanism. The second case study, consequently, will provide useful observations to develop a solution to the last challenge of this thesis: how to use big data, in the social sciences, to obtain evidence of causality that goes beyond correlations.

The case studies selected for this part of the thesis have two characteristics in common: they are both aimed at establishing causation, and they are both based on very large databases of data. However, while the study explored in chapter 5 is a MMR study, where

quantitative data are used in combination with qualitative data to establish both statistical and mechanistic relationships; the study discussed in chapter 6 uses a new conceptual framework to obtain evidence from big data.

More precisely, chapter 5 will describe a traditional MMR study on the causes of the phenomenon of Mexico-U.S. migration, where quantitative data collected from 92,527 individuals were analysed together with qualitative data to obtain both evidence of correlations and evidence of mechanisms (Garip & Asad, 2016). Quantitative data were analysed to establish a correlation between people's ties to prior migrants and their decision to migrate, while qualitative data were used to identify three causal mechanisms responsible for such a statistical correlation. Together, quantitative data and qualitative data helped to recognise the causal relationship by which social networks *influence* migration.

Chapter 6, instead, will examine a new epidemiological project called LIFEPATH. This project, like the mixed methods study, is based on large datasets of quantitative data, but what makes it an interesting case study is the novel approach characterising it, according to which data can be used to obtain *markers*. The analysis of LIFEPATH will be used to examine how biomarkers are identified by analysing biological data, and to develop a comprehensive definition of biomarkers. Such a discussion will then inform the discussion on sociomarkers in chapter 8.

5 Explaining migration through mixed methods: the phenomenon of Mexico-U.S. migration

5.1 Introduction

Human migration is often considered one of the most interesting social phenomena and is recognised as an essential area of contemporary research in the social sciences. In the second half of the twentieth century, *international* migration has become one of the main factors leading to social transformations and development in the world (Castles, 2000). After the Second World War, due to the post-war economic boom that stimulated massive labour migrations to the U.S. and Western Europe, international migration started expanding in volume and scope and became one of the main objects of study of the social sciences. International migration began to be considered a complex phenomenon, and numerous studies started exploring its causes, the continuity of migrant flows, its effects, the use of immigrants' labour and the socio-cultural adaptation of migrants (Portes, 1997).

In the studies on the determinants and continuity of migration, this phenomenon has in general been explained through the analysis of the socio-economic and political conditions characterising the home and the receiving countries. The central claims are that migrants could be 'pushed out' of their home countries because of the worsening economic conditions or political discontent. Conversely, migrants might also be 'pulled into' countries offering high wages, good health care, and stable and positive political conditions (Simpson, 2017).

Over the last few decades, however, new network accounts have dominated the discussions about the reasons for growing migratory flows (Garni, 2010). Network theories are based on the assumption that, rather than looking at the socio-economic and political determinants that might initiate migration, researchers should focus on what can perpetuate migration in time and space (Massey et al., 1994). This idea is very much in line with the consideration according to which migration is one of those phenomena that, once initiated, tend to acquire a self-sustaining character (Garcia & Jutila, 1988).

Empirical studies have provided support for such network theories: a case in point is the growing body of statistical studies showing that migration is more likely in communities already characterised by high levels of migration (Curran et al., 2005; Massey et al., 1994). Despite the fact that there is a general agreement about the possible causal

influence of social networks on migration, statistical studies have not been considered sufficient to establish a causal claim due to the difficulty in identifying the mechanism responsible for that correlation. Some researchers have emphasised the role that normative pressures might have in sending communities, and the potential mechanism that might operate at that level (DiMaggio & Garip, 2012). According to a putative normative pressure mechanism, a society in which migration is valued positively would push individuals to conform to the society's demand to avoid social sanctions. Other researchers have shed light on the crucial role that prior migrants can play by providing information or direct help to current migrants (Carrington et al., 1996). According to this second hypothesised mechanism, obtaining useful information from network peers can drastically reduce the cost of migration and can facilitate the process of migration.

Unfortunately, quantitative empirical studies have not helped to answer the question about what causal mechanism generates the observed associations. To complicate the situation, both the normative pressure mechanism and the mechanism associated with the existence of network peers could alone provide an explanation of the correlations resulting from the analysis of migration data, but it could also be possible to explain the pattern as the result of the interactions between such (and further) mechanisms. The lack of evidence of mechanisms has caused ambiguity in the interpretation of correlational evidence (Garip & Asad, 2016, p. 1171), and this ambiguity motivated the mixed methods study described in this chapter. As I shall discuss, the quantitative analysis conducted by Garip and Asad (2016) helped them to collect new difference-making evidence supporting the claim that social networks *influence* migration, whereas qualitative methods allowed them to gather evidence in favour of the different hypothesised mechanisms described above.

The chapter is structured as follow: section 5.2 will discuss the context and the mechanistic hypotheses motivating the mixed methods research. In particular, the section will provide a description of the three mechanistic hypotheses discussed in the literature. In section 5.3 I shall describe in detail how the mixed methods strategy was employed to explore such causal hypotheses. Section 5.4 will offer a review of the main results of the study, while section 5.5 will conclude. The overall argument defended in this chapter will be that the mixture of quantitative and qualitative approaches, in this particular study, helped not only to confirm the authors' causal hypotheses at the contextual level, but also

to strengthen causal hypotheses at a more general level, where different cases of migration flows can be taken into account.

5.2 Mechanistic hypotheses about Mexico–U.S. migration

5.2.1 The phenomenon of Mexico–U.S. migration

In their study, Garip and Asad (2016) employed a mixed methods strategy to explore the social mechanisms responsible for the growing Mexico–U.S. migration. Their research focused on a specific context that has attracted the attention of many researchers: the number of Mexican immigrants has indeed continuously increased over the past four decades, reaching, according to the Department of Economic and Social Affairs of the United Nations (2017), the current number of around 12.7 million of Mexican immigrants residing in the U.S., and making this flow the largest sustained migratory flow in the world.

Mexican migration to the U.S. is far from being a new phenomenon. Historically it was a phenomenon of *temporary* migration, characterised by the migration of unauthorised workers in agriculture who, after a couple of years, returned home. The stream started in the 1900s, when U.S. labour recruiters approached people living in Central-Western Mexico to enrol new workers. 4.6 million Mexicans were recruited through the Bracero program for short-term farm work between 1942 and 1964 (Cornelius, 2001); a further 3 million Mexicans moved to the U.S. without documents in the same period (Passel & Woodrow, 1987).

From the 1990s, the population of Mexican migrants began to increase very fast. After border crossing became more difficult due to the increased border enforcement, this form of temporary migration was partly replaced by a new form of *permanent* migration, with the consequence that the rates of return migration started declining (Massey, 2005; Reyes, 2004). The aspect that has attracted the attention of the authors, Garip and Asad, is that the growing number of migrants in the U.S. contributed to the emergence of strong social ties, that according to some hypotheses, might be crucial for migration. Each new migrant makes the social network connecting migrants in the U.S. to individuals in Mexico expand, and there is evidence that it causes an increase in individuals' migration propensities (Fussell & Massey, 2004; Massey & Zenteno, 2000). Even though researchers have accepted the probabilistic relationship according to which social

networks increase the likelihood of migration, no agreement has been found about the particular mechanisms producing such a statistical correlation. The collection of evidence of mechanisms was the main purpose of Garip and Asad's MMR study (2016).

5.2.2 Three mechanistic hypotheses

In their article, Garip and Asad (2016) disentangled and tested the presence of three different mechanistic hypotheses discussed in the literature. While quantitative data were analysed statistically to confirm the positive association established in the literature between individuals' ties to prior migrants and their decision to migrate, qualitative data were studied to verify the presence of causal mechanisms.

The mechanistic hypotheses tested in their paper were developed by the authors from the considerations proposed by DiMaggio and Garip (2012), who reviewed the literature on network effects and argued that three different general causal mechanisms were conceptualised in the literature: (i) social facilitation, (ii) normative influence, and (iii) network externalities. Such mechanisms were not just discussed in the context of migration, but were proposed as general mechanisms that can increase the probability of an individual to adopt a practice (DiMaggio & Garip, 2012, p. 95). Garip and Asad, hence, began their study by translating these three general mechanistic hypotheses into mechanistic hypotheses about migration.

Social facilitation is the hypothesised mechanism according to which network peers, by providing information, increase the utility or reduce the cost or risk of a new practice (DiMaggio & Garip, 2012, p. 96). Such an effect has been described by Kohler et al. (2001) in women's fertility decisions: when women are uncertain about the use of a new contraceptive device, information provided by friends who have already used it can reduce uncertainty. As for the decision to migrate, Garip and Asad (2016, pp. 1171–1172) hypothesised that network peers (such as family members and community members) can facilitate migration by providing useful information or by offering help that reduces the costs associated with migration. In Figure 11, I have tried to represent graphically the social facilitation mechanism. In order to operate, this mechanism requires a sufficient number of peers involved in the behaviour to exceed a threshold, under which the social facilitation effect is equal to zero. In other words, this social facilitation mechanism can be effective only if potential migrants know from similar situations that receiving help and information from network peers is likely to reduce the cost of migration. The effect

of this mechanism, furthermore, apparently increases at a declining rate with the number of peers: after having reached the threshold of peers engaging in the behaviour, the individual can obtain sufficient information without the need for further discussions with other peers (Kandel & Kao, 2006; Tilly, 2007).

This mechanism, finally, might also generate further effects such as the decision of some young adults to give up potential opportunities in their home country to seek fortune abroad (Winters et al., 2001), or the imposition of non-consensual social obligations on prior migrants, who are forced to help new migrants once they arrive in the U.S. (Portes & Sensenbrenner, 1993). Rather than being part of the mechanism, however, these further effects act as its possible consequences: the reduced costs associated with migration might cause such effects. It is for this reason, that they are not represented in Figure 11.

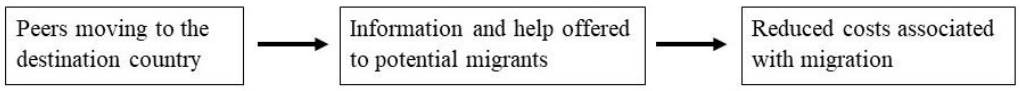

Figure 11. The social facilitation mechanism.

The second plausible mechanism, known by the name of normative influence, was described as the mechanism operating through "the application of positive and negative sanctions upon network members" (DiMaggio & Garip, 2012, p. 108). The propensities to stop smoking or to start a diet, for instance, were described in the literature as the effects of normative influence. In the case of migration, such a mechanism is claimed to be based on the fact that the community can offer compensation or impose sanctions to encourage a type of behaviour. In Figure 12, I have illustrated this mechanism. In heterogeneous communities, some members might support the decision to migrate whereas others might disagree about it. In such a case, the final network effect would depend on the relative proportion of supporters versus opponents (Garip & Asad, 2016, p. 1172). Empirical studies on Mexico-U.S. migration have collected evidence supporting the presence of this mechanism. In diverse settings, indeed, researchers have underlined the importance of a 'culture of migration' (Cohen, 2004; Reichert, 1981; Wiest, 1973), where not only is migration associated with positive ideas, but it is also considered as a sort of rite of passage towards a delineation of the individual's identity.

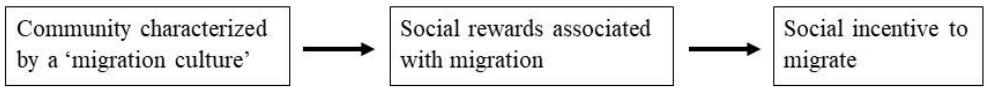

Figure 12. The normative influence mechanism.

The third mechanism that might cause migration is based on the development of network externalities. In general, network externalities can be found in the case of communications technologies and can operate by increasing the value of a specific practice as if it were a function of the number of prior adopters. Software systems such as Adobe Acrobat, which dominates the market by offering Acrobat Reader away for free, began to be used by a large number of people, and this led more people to download them not to lose the connectivity with network members (Varian et al., 2004).

If applied to migration, according to Garip and Asad, network externalities do not depend on the interaction between prior migrants and potential migrants, but are due to the fact that, when a considerable number of individuals move to another country, this phenomenon can produce important common institutionalised resources that might both reduce the cost or increase the value for those potential migrants who decide to leave their country, as described in Figure 13. Unlike the social facilitation mechanisms, the effect of this mechanism increases exponentially if the number of adopters grows, given that the maintenance of such resources depends on the size of the adopter group. Unlike the normative influence mechanism, moreover, this mechanism does not influence individuals' motivation to migrate through sanctions or incentives, but facilitates the act of migrating or helps individuals to obtain greater benefits from migration.

An example of a common resource generated through this mechanism is the organization of smugglers that helps undocumented migrants to cross the Mexico-U.S. border (Smith, 2006). Moreover, the emergence of migrant enclaves and hometown associations to support migrants' search for employment has often been associated with the presence of a large Mexican population in the country of destination (Goldring, 2004). It follows that the generations of such institutions could be seen as network externalities.

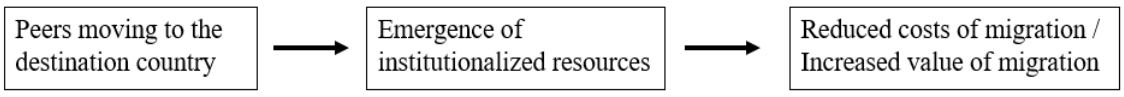

Figure 13. The network externalities mechanism.

The contextual knowledge of the situation, as well as previous empirical studies, led Garip and Asad (2016) to hypothesise, in their analysis, the combined presence of all these mechanisms. As described in the next pages, quantitative and qualitative methods were hence employed together to test this hypothesis and to shed light on the specific nature of these mechanisms.

5.3 Exploring causal mechanisms through mixed methods

5.3.1 Quantitative data

Garip and Asad started their study by analysing quantitative data coming from 124 Mexican communities surveyed by the Mexican Migration Project (http://mmp.opr.princeton.edu) between 1982 and 2008. The communities were surveyed in the winter months, the period in which typically U.S. migrants visit their families in Mexico. In each community, researchers randomly selected the households to study, and asked individuals to reconstruct the timing of their first and last trip to the U.S. and to provide demographic information. In total, researchers obtained information about 92,527 individuals.

The pieces of information collected retrospectively were organised in a panel of data. In each year, researchers reconstructed individual traits going back in time until the age of 15 years for each surveyed individual. The final individuals' variables reconstructed in such a way were: age, sex, whether the individuals are household heads, the level of education, occupation (agriculture, manufacturing, or service sector), whether they have migrated in Mexico (an experience called 'domestic migration'), household wealth (number of rooms in properties, land and business owned) and community type (rural or metropolitan) (Garip & Asad, 2016, p. 1175).

Researchers began also to look for political, socio-economic, and demographic factors showing a correlation with an individual's likelihood to migrate for the first time to the U.S. The variables collected for this purpose were: the inflation rate in Mexico, the ratio of available visas to Mexican migrants, the Mexico–U.S. trade (converted to constant US$ in 2000), the average hourly wage in the United States (in constant U.S. $ in year 2000), the number of U.S. legal residents and of U.S. migrants (non-residents) in the household, and the migration prevalence (proportion of people who have ever migrated) in the community (Garip & Asad, 2016, p. 1175).

Such variables were then analysed quantitatively through logistic regression, and a statistical model was developed to understand how the phenomenon of Mexico-U.S. migration operates generally. By controlling for education, occupation, household wealth, demographic characteristics and community type, the odds ratio of each political and socio-economic variable was measured, as illustrated in Figure 14.

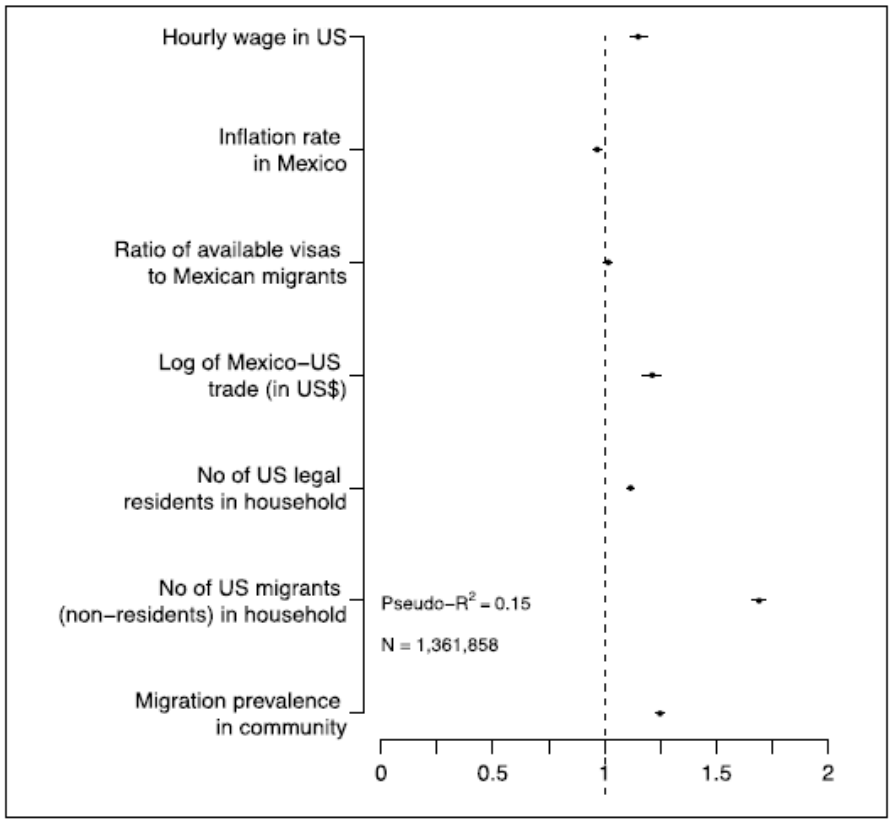

Figure 14. The odds ratio estimates. The black points represent the point estimate produced by the study, while the horizontal lines represent the 95% confidence interval (the set of values within which you can be 95% sure the true value lies). The dashed line is placed at the value (1) where there is no association between the factor (for instance, 'hourly wage in US') and the effect (i.e. migration). For instance, the probability of migration does not rise significantly when the probability of available visas to Mexican migrants rises. Furthermore, the point estimate for 'migration prevalence in community' is more likely to be the true value than the point estimate for 'hourly wage in US' (retrieved from Garip & Asad, 2016, p. 1175).

5.3.2 Qualitative data

Qualitative data were collected after the analysis of quantitative data through 138 in-depth interviews conducted in 120 households in the summer months of 2011 and 2013. The interviews were conducted in Jalisco, a state in Central-Western Mexico famous for being a major sender of migrants to the United States and to host different types of migrants (young and old migrants, female migrants, educated migrants). Information obtained from the analysis of quantitative data informed the selection, in Jalisco, of those sites containing a large concentration of different migrant types (Garip & Asad, 2016, p. 1176).

The first site selected was a rural village of around 1,000 people, whose migrants were male householders who started moving to the U.S. in the 1970s and early 1980s. The second site was a rural town with 3,000 inhabitants whose typical migrants were young men from wealthy families who migrated in the mid-1980s. The third community chosen for the study was an industrial town of 9,000 residents, from which both male and female individuals migrated (especially after the Immigration Reform and Control Act in 1986, which granted citizenship to undocumented migrants in the U.S. and allowed for family reunification). Finally, an urban neighbourhood in Guadalajara was selected as the fourth site due to the high percentage of educated male migrants who left Mexico in the mid-1990s to work in manufacturing.

For each community, around 35 households with at least one current or returned migrant have been interviewed. Overall, 166 individuals have been interviewed: 49 migrants, 49 parents of migrants, 34 spouses of migrants and 34 relatives (siblings, nieces of migrants). The interviews were semi-structured, with open-ended questions concerning the circumstances associated with the decision to migrate. After the interviews, Garip and Asad interpreted qualitatively the information obtained with the aim to shed light on those personal opinions and personal events that influenced each migrant's decision to leave Mexico to the U.S. for the first time. In such a way, researchers aimed at reconstructing the complex interrelations of factors that, together, led to the decision to migrate.

5.4 Establishing causation

Both the quantitative and qualitative analyses confirmed the salience of socio-economic factors for the decision to migrate. For many of the interviewed migrants, furthermore, the high pay in the U.S. was the main motivating factor. Almost all those interviewed

mentioned economic conditions or goals as the most important drivers of migration, but many also referred to migration as a chain of processes, where the first migrants in the community trigger others to leave to the U.S. as well. This qualitative result was confirmed by quantitative findings: like in similar quantitative studies, researchers found out that the odds of migrating were higher in communities with a high level of migration and in households with prior U.S. migrants (residents or non-residents) (Garip & Asad, 2016, p. 1177). To uncover the mechanisms responsible for such associations, qualitative data collected from interviews were crucial.

The most relevant operating mechanism, researchers discovered, was social facilitation: in 129 interviews migrants and relatives mentioned the help offered by others to cross the border or to find a job and an accommodation. Due to the risk associated with these tasks, many migrants clarified that they only relied on strong ties, and that such ties played a determinant role in their decision to migrate. For instance, a migrant's wife considered how the presence of some relatives changed her husband's perspective, making him think about the possibility of migrating:

> "An opportunity came up so that my siblings could help him—because nobody from his family was there—only my family. His cousins said they'd go with him, but they didn't, so we called my sisters, and they said he could go with them. That's why he left." (Garip & Asad, 2016, p. 1180)

From the interviews, researchers were also able to collect evidence of the presence of the normative influence mechanism. 51 interviews cast light on how communities and families often try to persuade the migrant to go by making very explicit their approval, while other respondents described the shame felt by migrants when their experiences and the expectations of their families do not match. The high expectations in the sending communities and families, furthermore, were powered by the decision made by migrants with negative experiences not to return or share their stories because of the shame felt. Some of them, even began to lie about their success, as reported by some respondents:

> "[…] they came and told stories that weren't real. Really! "No, no, over there I have," "Over there I am," and they were all lies. […] When I left, I went with friends who had [spent] years over there and noticed that they hardly had anything to eat. They've been there for years, and the ones here thought they were millionaires over there, but it was all a lie." (Garip and Asad 2016, 1182)

From the analysis of qualitative data, hence, researchers not only confirmed the presence of the hypothesised normative influence mechanism, but also realised that such a mechanism could both trigger the decision to migrate and act as a reinforcing mechanism leading to a very strong 'migration culture', as shown in Figure 15.

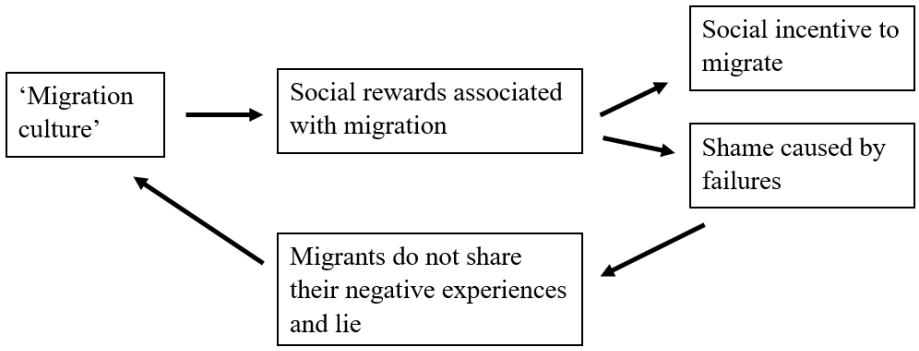

Figure 15. The normative influence mechanism can both cause migrants to leave to the U.S., and reinforce the migration culture necessary for this mechanism.

Finally, in 97 interviews researchers found evidence of the network externalities mechanism. Many migrants explained how they found a smuggler, or what kind of help they received from institutionalised resources. Due to the risks associated with trusting a smuggler, in 93 out of the 97 cases, this mechanism worked in tandem with the social facilitation mechanism, as many migrants decided to rely on their social ties to find a competent smuggler:

> "Everybody around [a nearby city] knows who the coyotes are. "Go with José. Go with that guy, look for him," and then they [coyotes] ask, "Who sent you?" "Herbierto's brother." "Oh, OK. Let me talk to him." "Luís, you sent such and such?" "Oh, OK, look after them." So, trust. […] Even for finding a coyote you need to know people." (Garip & Asad, 2016, p. 1183)

In all these cases, qualitative data were used together with quantitative data to *establish a causal relationship*. As described by the authors, quantitative data alone would not have solved the puzzle concerning the reason for the correlation between migration behaviours among peers and the decision to migrate. This, consequently, would not have ruled out the doubts concerning the presence of a spurious correlation. Only through the analysis

of qualitative data, hence, researchers were able to establish that the correlation was actually due to certain causal links.

On the other hand, however, qualitative data alone could not establish any causal link. Especially in the case under study, where the authors hypothesised more than one mechanism, it would have been possible to have masking phenomena, where more mechanisms interact and the operation of one mechanism 'masks' the operation of another mechanism, leading to a null net effect. To give an example, the normative influence mechanism might produce social incentives to migrate, while the network externalities mechanism might lead to the creation of hometown associations to support migrants' search for employment. This second mechanism might provide support to those people who would like to migrate on the one hand, and might bust the myths about the ease with which migrants earn money and get successes in the U.S. on the other hand. The awareness that migration involves economic risks and does not automatically lead to a happier life, consequently, could cancel the effect of the 'migration culture'. The overall effect of these two mechanisms, hence, might be null.

Only considering together quantitative and qualitative evidence, hence, Garip and Asad were able to establish both the presence of causal mechanisms and of a net effect on Mexico-U.S. migration. In accordance with RWT this, finally, led the authors to establish a causal link between the presence of migrants among the Mexican family and community, and the individual's decision to migrate to the U.S. from Mexico.

5.4.1 General and singular causation

In this sub-section I shall argue that the mixture of quantitative and qualitative data did not allow Garip and Asad just to establish the presence both of some mechanisms and of a correlation, but also helped them to combine different ontological categories, as described in chapter 4, section 4.4.1. My argument is that the causal relationship between the presence of migrants among the Mexican family and community, and the individual's decision to migrate to the U.S. from Mexico was established by moving from the ontological category of general causation to the ontological category of singular causation. Let us consider again the MMR study conducted by Garip and Asad.

To begin with, the first ontological categories involved in the study were both at the general level: researchers translated general hypotheses concerning the existence of network effects and of different causal mechanisms behind them into more precise (but

still general) hypotheses about the presence of such effects and mechanisms in migration flows. To give an example, the general hypothesis concerning network externalities is based on the assumption that they operate when the value of a practice is a function of the number of prior adopters (DiMaggio & Garip, 2012, p. 96). The general hypothesis can be applied to different contexts, such as communications technologies (if a software systemis used by a large number of people, more people will download it not to lose the connectivity with network members), languages (if a growing number of people speak a language, more people will decide to learn it and they will do it easily), fertility choices (choosing to have children and raising them is easier if there are friends who have kids), and divorce (the more divorced people there are in a social network, the more recommendations and suggestions are available for those who need them) (Booth et al., 1991; Church & King, 1993; DiMaggio & Garip, 2012). Garip and Asad adapted this hypothesis to the specific context of migration, claiming that such externalities were based on the generation of institutionalised resources facilitating the adoption of the behaviour to new individuals.

After this movement from general causation to a more specific general level, Garip and Asad translated the general causal hypotheses about migration into detailed general hypotheses regarding the specific context of Mexico-U.S. migration. Taking again network externalities as an example, for instance, in the specific case of Mexico-U.S. migration flow, network externalities mechanisms were hypothesised to operate through the emergence of smugglers for crossing the Mexico–U.S. border, and of migrant enclaves. Finally, to confirm the presence of such operating mechanisms in the specific context of Mexico-U.S. migration, Garip and Asad collected evidence at the local level. By means of qualitative interviews, the authors managed to establish the presence of at least one of such mechanisms in a significative number of local, singular cases.

As I have illustrated in Figure 16, the causal study conducted by Garip and Asad, consequently, combined not only methods but also ontological categories: general causation at the most general level, general causation at the migration level, general causation at the Mexico-U.S. migration level, and singular causation regarding the local cases of Mexico-U.S. migration. The movement from general to singular causation illustrated in Figure 16 was made possible by the qualitative data collected by the authors, that helped to clarify the contextual factors influencing migration decisions in some specific cases of Mexico-U.S. migration. This context, according to the authors, was

studied by many in the literature just to establish a "positive association between individuals' ties to prior migrants and their migration propensities" (Garip & Asad, 2016, p. 1185). The information collected through qualitative interviews in their study was claimed to be fundamental in order to take a step forward and uncover, in that context of Mexico-U.S. migration, the presence, in specific local cases, of the mechanisms responsible for that correlation.

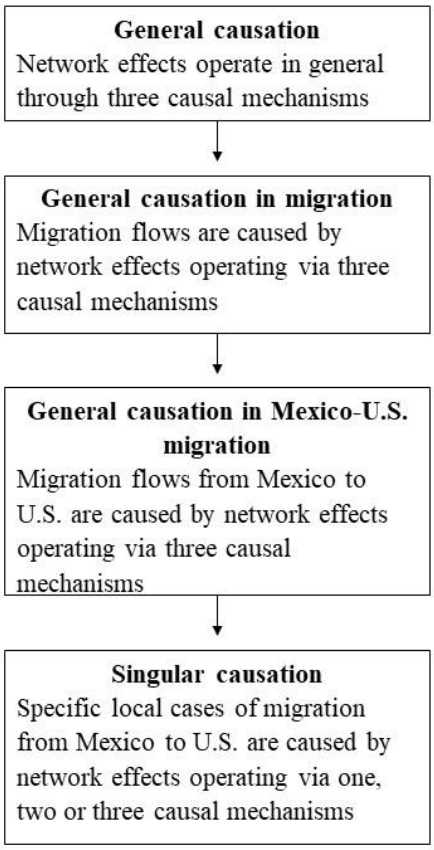

Figure 16. From general to singular causation. From a general hypothesis about the presence of network effects mechanisms, the authors moved to an intermediate general level in which they hypothesised the presence of such effects in migration flows. At this level, further details were added to the mechanistic hypotheses. Then, the authors moved to a more specified intermediate level of general causation, where the mechanistic hypotheses were adapted to the specific level of Mexico-U.S. migration. Finally, local mechanisms were hypothesised and established in singular instances of the migration flow from Mexico to the U.S.

Quantitative studies, in contrast, helped researchers to perform an opposite movement from singular causation to general causation: by analysing quantitatively the qualitative data obtained from the interviews, Garip and Asad discovered that in about 90% of the cases migrants described the operation of at least two mechanisms at the same time; and for more than 33% of the migrants interviewed, all three mechanisms worked together. Since qualitative data were collected from four very diverse Mexican communities to maximise the diversity of respondents (Garip & Asad, 2016, p. 1176), the quantitative results they obtained strengthened their confidence about the possibility of generalising the operation of the local mechanisms to further cases, not studied in the project, of Mexico-U.S. migration. In such a way, the movement from general network effects mechanisms to more contextual (but still general) network effects mechanisms in the case of Mexico-U.S. migration was confirmed by the generalisation of local mechanisms operating in singular, specific cases.

The generalisation at the Mexico-U.S. migration level, then, was used to extend the claim that more than one of the network effects mechanisms described in the study might be responsible for migration flows also in other cases of migration:

> "By delineating the sources for interdependent migration choices, we provide a deeper understanding of migration as a social process, which is crucial for anticipating future flows and policy responses." (Garip & Asad, 2016, p. 1170)

> "By focusing on the social mechanisms underlying the network effects on migration, we can anticipate whether and how these effects may decline in size, or be reversed. In particular, the mechanisms we have identified—social facilitation, normative influence, and network externalities—generate a positive feedback loop as long as migration remains a successful enterprise, or at least is perceived as such, in sending communities." (Garip & Asad, 2016, p. 1185)

In other words, first some general mechanistic hypotheses were adapted to the context of migration and then to the context of Mexico-U.S. migration, and qualitative data allowed researchers to claim that a singular, complex, instance of causation, involving three different mechanisms, exists in specific cases of Mexico-U.S. migration flow. Then, this result was used as a piece of evidence in support of the idea that such causal mechanisms

could operate generally not only in the case of Mexico-U.S. migration, but also in further cases of migration. Figure 17 represents this movement from singular causation to the different levels of general causation.

Overall, the mixture of qualitative and quantitative data helped researchers to explore causation at different ontological levels, paving the way for further studies on the causal relationships between individuals' ties to prior migrants and their decision to migrate in different cases of migration.

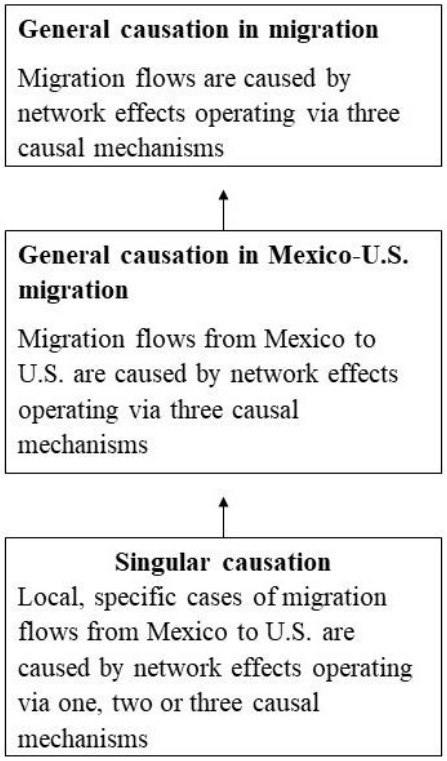

Figure 17. From singular to general causation. From some contextual pieces of evidence showing the presence of the network effects mechanisms in local cases of Mexico-U.S. migration flow, the authors moved to the general Mexico-U.S. migration level and then to the general migration level to strengthen the hypothesis that the same mechanisms can operate also in other migration contexts.

5.5 Conclusion

The Mexico-U.S. Migration mixed methods study allowed researchers to establish the causal link responsible for a phenomenon that has been discussed many times over the last century. A large sample of quantitative data from the Mexican Migration Project

helped researchers to obtain difference-making evidence in support of the correlation between having network peers who migrated to the U.S. and the propensity to migrate to the U.S.; while in-depth qualitative data allowed researchers to establish the presence of three contextual network mechanisms influencing the decision to migrate in local instances of Mexico-U.S. migration. In such a way, Garip and Asad managed to establish causation in the context of Mexico-U.S. migration.

While establishing general causation at the contextual level was the aim of the MMR study, it is worth noting that the mixture of quantitative and qualitative approaches did not help the authors just to confirm their causal hypotheses at the contextual level. On the one hand, the MMR study helped also to strengthen causal hypotheses at a more general level, where different cases of migration flows can be taken into account. In such a case, indeed, the statistical analysis of the information collected through qualitative interviews allowed for the recognition of the pervasive mixture of more network effects mechanisms together, and the prevalence of all three mechanisms. On the other hand, qualitative evidence allowed the authors to analyse singular causation and to establish the presence of local mechanisms in specific and unique instances of Mexico-U.S. migration. Overall, the case study described in this chapter offers a clear representation of a traditional piece of research aligned with RWT, and clarifies how the mixture of different methods and of different data can allow researchers to mix different ontological categories.

6 Big data in social epidemiology: the case of LIFEPATH

6.1 Introduction

LIFEPATH is an interesting project conducted by social epidemiologists and aimed at discovering the causal relationships between socio-economic factors and the development of diseases. The starting assumption is that factors external to the human body, such as socio-economic conditions, can trigger changes internal to the human body that in turn lead to the development of diseases. As a consequence, different factors both at the social and at the biological level are measured and studied within the project. In this chapter, I shall explore how LIFEPATH researchers use big data to uncover causal relationships between social and biological factors.

LIFEPATH is a social epidemiology project, consequently it is based on the combination of rigorous epidemiological methods and theoretical insights from the social sciences. The ultimate goal is not to explain a social phenomenon: researchers in LIFEPATH want to establish the *effects* that social factors, such as social structure and environment, have on individual and population health. Given this aspect, someone could argue that other projects, focused more on social phenomena and less on biological factors, would be more relevant in the social sciences. My decision to use such a case study, therefore, needs to be clarified. To justify my choice, two considerations should be taken into account.

First, LIFEPATH explores the effects of social factors, and this characteristic makes the project relevant to the social sciences. Several social scientists work on similar research questions, trying to establish the effects of social conditions on people's life. Furthermore, as I shall describe, LIFEPATH studies comply with RWT. Hence, my answer to those who argue that other projects, focused less on biological factors, would be more relevant in the social sciences, would be that the relevance of LIFEPATH in the social sciences cannot be dismissed just because the final phenomena that LIFEPATH wants to explain deal with health outcomes.

Second, what makes LIFEPATH particularly interesting if compared to other projects studying social factors, is the way in which the study is conducted. Not only are the datasets employed by LIFEPATH incredibly large (to examine the socio-economic factors that might be causally relevant to particular diseases, scientists use data from 7 child cohorts and 11 adult cohorts), but to examine the changes at the biological levels,

scientists rely on a novel approach of data analysis that allows them to identify and use pieces of information known by the name of biological markers. The identification of markers is, as I shall suggest, a ground-breaking way to obtain causal evidence from big data.

While in this chapter I shall describe how markers can be identified at the biological level, chapter 8 will suggest that similar markers can be used also at the social level. In other words, the description of LIFEPATH's use of biomarkers will pave the way both for general considerations concerning the identification of markers and for a new proposal about the possibility of using not only biomarkers, but also *socio*markers.

The chapter is structured as follows. In section 6.2 I shall present the 'ancestors' of LIFEPATH, EnviroGenomarkers and EXPOsOMICS, and the goals of the LIFEPATH project. In section 6.3, I shall examine how socio-economic data are collected and used within the project. In section 6.4, I shall focus on the use of biological data to obtain biomarkers. In particular, I shall first describe how biomarkers are used, and subsequently I shall explore how biomarkers are identified and conceptualised as intersecting signals. In section 6.5 I shall use the observations on biomarkers collected through the case study to propose a novel framework for biomarkers. I shall argue that what I will call a 'relational framework', based on Leonelli's proposal (2015), is needed for biomarkers to emphasise their nature, and I will claim that biomarkers are used both to develop new causal hypotheses and to collect evidence supporting them. It follows that the very notion of biomarkers is associated with causal reasoning, even if often this is not explicitly discussed when biomarkers are identified or used in new ways. Overall, these proposals will be used to offer a new definition of biomarkers. Section 6.6, finally, will conclude.

6.2 LIFEPATH

6.2.1 The origins of LIFEPATH

LIFEPATH came into being thanks to two previous projects, whose methods and findings contributed to its development. The first 'ancestor' of LIFEPATH was the project called EnviroGenomarkers. That project, funded by the European Union, investigated the effects of environmental exposures on the development of various diseases. The main goal was to measure the effects of environmental agents on diseases through the evolution of biomarkers. Two underlying ideas were crucial in the EnviroGenomarkers project.

To begin with, to investigate the development of diseases, scientists considered the totality of environmental exposures affecting individuals during their lifetime. Such a totality, known by the name of 'exposome'[12] (Wild, 2012), comprises both 'internal' and 'external' exposures. For instance, some metabolites contained in our biological fluids can act as 'internal' exposures, affecting inflammation, oxidative stress and various metabolic pathways. These metabolites, furthermore, can be affected by some 'external' environmental exposures, like radiation.

In addition, the project EnviroGenomarkers was founded upon the concept of biomarkers produced through '-omic' technologies. The nature of biomarkers will be discussed in detail in section 6.4, for the moment it is sufficient to know what '-omic' technologies are, and what the term 'biomarker' represents. In EnviroGenomarkers, '-omic' technologies are described as promising technologies able to find the missing links between exposome and disease. The main feature of '-omic' technologies is their capacity to analyse vast sets of biological molecules, rather than a single biological structure (like a protein or gene). The emerging methods of measuring families of cellular molecules are known by the names of 'genomics', 'transcriptomics' (gene expression profiling), 'proteomics' and 'metabolomics'.

In biology, the term 'genome' is used to refer to the whole hereditary information encoded in the DNA (or, for some viruses, in the RNA). To be more precise, the genome of an organism is the DNA sequence of one set of chromosomes. The study of an organism's genome through '-omic' technologies, aimed at understanding its complex function, is called genomics.

Transcriptomics is the application of evolving technologies to scan the fifty thousand genes that we currently know are transcribed into RNA molecules from the three-billion-letter human genome. Such mRNA transcripts in the cell reflect the genes that are actively expressed at any given time points. Measuring the expression of an organism's genes at different time points, and in diverse tissues and conditions helps scientists to understand how genes are regulated, and how certain developmental pathways such as environmental stimuli responses are triggered (Debnath et al., 2010).

[12] The concept was developed in parallel with the notion of 'genome', a term used to characterise the genetic material of an organism. The 'exposome' complements the genome by addressing both genetic and non-genetic exposures.

Proteomics indicates the analysis of gene products, proteins, on a large scale. In general, proteomics allows scientists to study protein expression profiles, protein modifications and protein networks in relation to cell function and biological processes (Macaulay et al., 2005).

Finally, metabolome refers to the complete set of small-molecule metabolites (such as metabolic intermediates, hormones and other signalling molecules, and secondary metabolites) that are found within a biological sample, such as a single organism (Debnath et al., 2010). Metabolomics, based on '-omic' technologies, aims to comprehensively study the metabolome of cells, body fluids and tissues.

In EnviroGenomarkers, the major contribution of '-omic' technologies was through the identification of relevant biomarkers. Biomarkers are measures of elements or characteristics of the environment and the organism that help to understand biological processes. In a first step, '-omic' technologies allow researchers to collect vast amounts of data on particular molecules (like DNA, metabolites) in individuals with specific health conditions. The dataset is then analysed to determine whether it is possible to obtain biomarkers, that in turn can be used in different ways: some biomarkers help to diagnose accurately the health disorder of the patient; other biomarkers are used to trace the course of a specific disease; another group of biomarkers is studied to obtain information about the risk of developing diseases. Some of these biomarkers, together, help scientists to cast light on causal processes starting with certain environmental exposures and ending in the emergence of diseases.

The more recent 'antecedent' of LIFEPATH is the project EXPOsOMICS, whose main goal and methods were very similar to those of EnviroGenomarkers. EXPOsOMICS, based on the study of biomarkers and on the concept of exposome, acted as a bridge between EnviroGenomarkers, focused solely on the effects of environmental exposures, and LIFEPATH, interested in the causal link between socio-economic conditions and diseases. It was especially during this project, indeed, that scientists began to consider socio-economic factors as causally relevant to health conditions. As an example, in the article published by Vineis et al. (2017), EXPOsOMICS researchers stated that:

> "It is generally accepted that the majority of important chronic diseases are likely to result from the combination of environmental exposures to chemical and physical stressors and human genetics. There is also evidence that the

> effects are location-specific and influenced by climatic, lifestyle and socio-economic characteristics." (Vineis, Chadeau-Hyam, et al., 2017, p. 143)

If we consider EnviroGenomarkers, EXPOsOMICS and LIFEPATH, the hypothesis that health effects are caused by socio-economic characteristics was developed when scientists were conducting EXPOsOMICS, but was not tested in that project. Scientists in EXPOsOMICS focused their attention on the causal relationships between environmental exposures and diseases. While some exposures were thought to be caused by certain socio-economic conditions, researchers in EXPOsOMICS did not use socio-economic indicators: the EXPOsOMICS datasets were made only of biological and exposure data (such as measures of chemicals in water or air). The causal relationship between socio-economic characteristics and health became then the heart of LIFEPATH, as we will see in the next section.

6.2.2 The LIFEPATH project

LIFEPATH is a big data project funded by the European Community and devoted to the investigation of the causal relationships between socio-economic factors and the development of diseases. Over the last few decades, many researchers have collected evidence that individuals from more disadvantaged socio-economic positions have higher mortality rates than people of the same age from higher socio-economic positions (Chetty et al., 2016; House et al., 2005; Mackenbach et al., 2008). Furthermore, studies have shown that such people are more likely to suffer from diseases and disabilities in their life. While the associations between socio-economic factors and health are often discussed within social epidemiology, the paths through which psychosocial and economic experiences trigger biological causes that ultimately influence health are still underdeveloped if compared to the biological pathways (Adler & Ostrove, 1999).

The aim of LIFEPATH is to fill this gap by illuminating how social, environmental and economic conditions (like wealth, education, job and housing) may lead to short and long-term health effects. In order to do this, researchers in LIFEPATH use both statistical analyses, exploring correlations between certain socio-economic factors and some diseases (see d'Errico et al., 2017), and mechanistic studies aimed at finding the intermediate mechanisms and pathways through which socio-economic conditions influence organic parameters (see Vineis, Avendano-Pabon, et al., 2017). As reconstructed by Vineis, Illari and Russo (2017, p. 5), only when researchers confirm the

presence both of a difference-making relationship and of a mechanistic relationship between certain socio-economic factors and health conditions, causation is established. To establish correlations and mechanisms, furthermore, vast amounts of data concerning socio-economic factors and health conditions are required.

Many of these data are analysed to find relevant correlations, and to identify intermediating markers that might allow researchers to recognise the processes linking socio-economic factors and biological changes. The aim of LIFEPATH is to trace the processes both at the social and biological levels, however social and biological factors are explored in different ways. On the one hand, the roles played by social factors, as we will see, are explored by using only the traditional socio-economic position indicators. On the other hand, the processes at the biological level are explored by means of biomarkers. It follows that, while the social components are analysed at a rather coarse-grained level, the biological components of the mechanisms of health are explored at a fine-grained level.

To identify biomarkers, scientists working in LIFEPATH follow the same approach used in EnviroGenomarkers and EXPOsOMICS: to begin with, they obtain new data by means of '-omic' technologies adopting a holistic view of the molecules that make up cells, tissues or organisms (Horgan & Kenny, 2011). Then, they examine whether from such data it is possible to obtain biomarkers. In the next sections of this chapter, I shall explore how such biological markers are identified, and how they might enhance our understanding of causal processes. Before answering this question, however, it is important to clarify the strategies through which socio-economic and biological data are collected.

6.3 Using socio-economic data in LIFEPATH

Research on the social determinants of health requires the use of some measures of socio-economic position. Although socio-economic position is at the heart of this kind of research, its study appears complicated by the fact that researchers have not yet agreed on a definition and on the best way to measure it.

LIFEPATH uses educational, occupational, and income data in order to measure socio-economic position and, subsequently, to study the causal relationship between socio-economic position and the development of diseases. This choice is in line with the

epidemiological practice: traditionally, epidemiological studies use all three of these measures as indicators of a person's socio-economic position.

More specifically, in LIFEPATH education is measured in every cohort in a way that makes measures comparable across countries. Three levels are identified (d'Errico et al., 2017): 1) primary and lower secondary school (during between 7 and 9 years and started immediately after the nursery), 2) higher secondary school (during 4 or 5 years and ending with a high school diploma level) and 3) tertiary education (any type of education received after high schools, like BSc, MSc, and further education). Educational data are useful for three different reasons. First, they allow researchers to classify all individuals in a society, regardless of the fact that they are active participants in the labour market (Galobardes et al., 2006). Second, such data help researchers to obtain information about childhood and adolescent socio-economic positions, given that education is in general strongly influenced by the resources of the family of origin. Third, with respect to the two other socio-economic measures, occupation and income, the level of education has the advantage of being unlikely to be caused by health conditions occurred in adulthood.

The fact that educational data provide information about childhood and adolescent socio-economic positions is considered particularly relevant because one of the hypotheses linking socio-economic conditions and health is that bad conditions in early life can affect the current health situation of a person and, at the same time, can increase the possibility of long-term effects in adulthood. The collection of educational data helps researchers to obtain information about people's childhood, and such knowledge can be then used to verify whether there are strong correlations between low socio-economic conditions during childhood and diseases manifested many years later, especially in the cases in which the socio-economic conditions improved from childhood to adulthood.

The third advantage of measuring education is based on the awareness that occupation and income might both affect health conditions and be influenced by poor health during adulthood. The use of occupational and income data, hence, might lead to a causal loop in which health conditions influence occupation and income, and occupation and income influence health conditions. Education, on the contrary, is affected by health conditions only in very specific cases and only in childhood. Consequently, statistical correlations to educational level can be analysed without the worry that health conditions in adulthood, rather than being an effect of education, are some of the causes influencing it.

As regards the occupation measurements, the conceptual reference of LIFEPATH is the European Socio-economic Classification (ESeC), that classifies occupations in ten ordinal categories, as shown in Table 6. Such categories are formed by grouping occupations with similar opportunities and 'life chances': as an example, one of the strategies used by researchers is to consider the similarity of employment relations and conditions. Occupational data are used in LIFEPATH not only to assess the socio-economic conditions of the working individuals, but also to recognise the socio-economic conditions of children. In that case, parental occupation is assumed to influence children's conditions through access to material resources (that determine material living standards), work privileges (i.e. social security), and toxic exposures (Galobardes et al., 2006).

	ESeC class	*Common term*	*Employment regulation*
1	Large employers, higher grade professional, administrative and managerial occupations	Higher salariat	Service Relationship
2	Lower grade professional, administrative and managerial occupations and higher grade technician and supervisory occupations	Lower salariat	Service Relationship (modified)
3	Intermediate occupations	Higher grade white collar workers	Mixed
4	Small employer and self employed occupations (exc agriculture etc)	Petit bourgeoisie or independents	Not applicable
5	Self employed occupations (agriculture etc)	Petit bourgeoisie or independents	Not applicable
6	Lower supervisory and lower technician occupations	Higher grade blue collar workers	Mixed
7	Lower services, sales and clerical occupations	Lower grade white collar workers	Labour Contract (modified)
8	Lower technical occupations[1]	Skilled workers	Labour Contract (modified)
9	Routine occupations[1]	Semi- and non-skilled workers	Labour Contract
10	Never worked and long-term unemployed	Unemployed	Not applicable

Table 6. The ten occupation categories of the European Socio-economic Classification (ESeC), retrieved from Rose and Harrison (2007, p. 464).

Finally, personal income data are collected only in three cohorts, and in different ways. Consequently, before using income data, scientists need to make measures from different cohort comparable though a process of harmonisation. These data are collected having in mind a specific hypothesis, according to which income might influence health through

two paths: allowing people to obtain material resources for living (such as food, clothing and housing) and access to health care; and giving people the opportunity to participate in activities such as sport and cultural events.

In the epidemiological research tradition, for a long time education, occupation and income measures have been used as if they were interchangeable, with the assumption that they represent exactly the same phenomenon. Only in the last 10 years this practice has been questioned and tested. A few studies such as that of Geyer et al. (2006), have recognised that education, income and occupation are only moderately correlated and have different strengths of relationship with different health outcomes.

The consideration that such measures might represent different aspects of the socio-economic position and that, therefore, the correlations between them might vary, led scientists to examine the relationships between the socio-economic indicators used in LIFEPATH. One of the first papers published by d'Errico et al. (2017), was specifically aimed at presenting the three socio-economic indicators employed in LIFEPATH and to compare their correlations to mortality. By analysing data from 518,061 participants, grouped in 18 cohorts (7 child cohorts and 11 adult cohorts), the authors obtained two relevant results. First, they discovered that, among men, mortality was not associated with the occupational class of the father. Second, they also considered that, among women, only education and occupation showed an association with mortality. Individual income and mortality in women were not associated, as there were cases in which women characterised by very low personal income lived longer than those women with a middle income. From such a result, researchers concluded that, for women, the use of their sole income data without considering those of their partners might cause misclassification of their social position. Women with low income might not need to work given their privileged position or the high income of their partners. More generally, furthermore, this result cast light on the difficulties associated with the choice of the most suitable socio-economic indicators in specific contexts.

6.4 Using biological data in LIFEPATH: the novel biomarker approach

To identify correlations between socio-economic measures and health conditions, researchers obtain disease and mortality data from administrative information about each

cohort. In addition, LIFEPATH researchers collect vast amounts of biological data that are used to obtain biomarkers.

In section 6.2.1, it has been said that biomarkers are measures of elements or characteristics of the organism. An illustrating example where biomarkers are successfully used to predict the incidence of certain diseases is the case of exposure to aflatoxin. Aflatoxin is a poisonous carcinogen produced by certain fungi which grow in soil, decaying vegetation, hay, and grains. Such fungi can contaminate some types of foods, such as rice and peanuts. If people eat these contaminated foods, they are exposed to aflatoxin. In the liver, aflatoxin is metabolised by CYP450 isoforms to form a reactive intermediate, AFB1-8,9-epoxide. This intermediate, furthermore, can form DNA adducts via covalent binding to the N7 atom of guanine. In this case, aflatoxin-N7-guanine adducts in serum are used as biomarkers of exposure to aflatoxin, while aflatoxin-N7-guanine adducts in the liver are used as biomarkers of early biological effects (Egner et al., 2001).

The biomarker approach is novel for three different reasons. First, traditionally epidemiology establishes correlations between population-level variables: those associated with the environment and those describing diseases. Biomarkers, on the contrary, allow researchers to go much smaller: not only the population-level factors themselves, such as environmental exposures, can be studied at the chemical level, but also their effects are now measured at the molecular level.

Second, the aflatoxin example represents just one of the possible situations in which biomarkers are identified and used. The term ‘biomarker’, indeed, can be attributed to measures of different bodily entities and features. While some biomarkers are identified by measuring existing entities which are either parts of the body (like pancreas cells, biological fluids, insulin receptors) or found inside the body (as in the case of aflatoxin-albumin adducts in serum), it is often the case that biomarkers are obtained measuring qualities of a bodily component. An illustrating example is urine’s colour, that can be measured through a Urine Color Chart (that attaches a number to a particular colour) and can be used as a quality biomarker of dehydration. In other circumstances, measures of bodily processes such as reduced insulin production can count as biomarkers.

In all these cases, it is essential that the biomarker is objectively measurable. Thus, when researchers analyse qualities and processes, they have to establish how to attach numbers

to them: for instance, scientists might decide that blood glucose is observable through blood samples and might establish the best strategy to represent it numerically. Blood samples, hence, might be the means through which a specific quality is made observable and measurable. Moreover, there are cases where the entity, quality or process is so multifaceted and complex that, in order to measure it, researchers use a cluster of variables that need to have particular values. An example is the case of allostatic load, a process of 'wear and tear' experienced by the body during stressful situation (more details will be discussed in section 6.4.3): in this case, more than 25 biomarkers can be used to measure some aspects of the process, and quite often they are combined to obtain a single number associated with the process (Juster et al., 2010, pp. 10–11).

It is important to note that not all the bodily qualities that can be measured qualify as biomarkers: as an example, the measure of a colour of a bodily component could be a biomarker of certain diseases, while the measure of the size and temperature of that component might not be relevant, therefore they would not be used as biomarkers.

The third reason why the biomarker approach can be considered a novelty, is that biomarkers have to be linked to the causal process under study. Biomarkers are always biomarkers *of something*, in relation to a specific process, as I shall suggest in section 6.5.1. The link between biomarkers and causal processes, however, can vary, as shown in Figure 18. Biomarkers can be directly involved in the causal process causing a specific event, as shown in Figure 18(a): we have seen, for instance, that measures of aflatoxin-albumin adducts in serum are used as biomarkers of exposure to aflatoxin, and that such a presence is caused by exposure, and can produce causal effects in the body.

In addition, biomarkers can be causally linked to (but not acting parts of) such a process, as illustrated in Figure 18(b): as an example, Vineis, Illari et al. (2017) considered that measures of tobacco stain on fingers might be used as a biomarker of lung cancer given that such a bodily quality is causally linked to smoking, a cause of lung cancer. A more realistic case is the use of the measure of HPRT hypoxanthine-guanine phosphoribosyltransferase (HPRT) gene mutation as a surrogate for other target genes (such as mutated p53 gene) that are directly involved in the development of cancer. In this case, the biomarker is used to substitute another event or factor that is unmeasurable but directly involved as an acting part in the causal process, like in Figure 18(c). In other circumstances, the biomarker might be caused by an unmeasurable part of the process,

like in Figure 18(d): for instance, the process from an infectious agent to chronic inflammatory disease in general involves an immune response, which in turns produces inflammatory cytokines such as IL-17, IL-1 and IL-6. IL-6 also exerts an effect on the liver, where it induces the production of the C-reactive protein (CRP). In numerous cases, CRP measures are used as biomarkers of the inflammatory process leading to chronic inflammatory disease, even though CRP is just correlated to such a process because it is caused by IL-6.

Finally, biomarkers can also be used as predictors of events when they cause the factors that trigger the causal processes, like in Figure 18(e): a case in point is the use of the human chorionic gonadotropin measure (the biomarker of pregnancy) to predict the possibility of the Ballantyne syndrome, a rare disorder affecting pregnant women. In such a case, the biomarker would be considered a background condition: a condition required for the cause to exert its effect.

These possibilities help to distinguish between simple proxies and biomarkers. In order to have a proxy, it is sufficient to find a correlation between one unmeasurable factor and one measurable factor (the proxy). In order to have a biomarker, instead, it is necessary to consider the causal links between the biomarker and the other measurable and unmeasurable factors. Let us consider again the case of yellow fingers. The fingers' colour might be measured and used as a proxy for smoking or for lung cancer without knowing the causal relationships between such factors. Moreover, yellow fingers might be used as a biomarker linked to (but not part of) the process from smoking to lung cancer once we find out that yellow fingers are caused by smoking (like in Figure 18(b)). Finally, it might be possible to use yellow fingers as a biomarker also in the case we discover that tar deposition itself, the cause of yellow fingers, is part of the process from smoking to lung cancer (like in Figure 18 (d)).

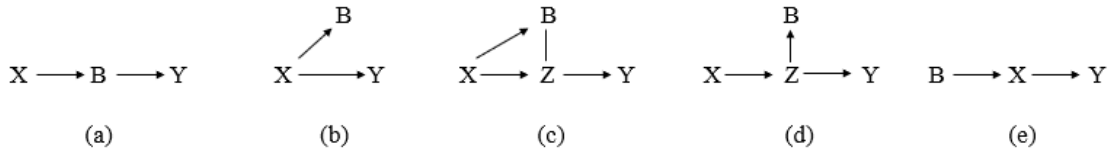

Figure 18. Possible DAGs showing the position of biomarkers. (a) represents the situation in which the biomarkers B is part of the causal process linking the cause X and the effect Y. In (b), the biomarker B and the effect Y are both caused by X. In (c), the biomarker B and Z,

that is part of the process but cannot be measured, are both caused by X. In (d), the biomarker B is caused by the unmeasurable part of the process Z. In (e), B is a background condition of the process from X to Y.

6.4.1 Defining biomarkers

Two recent definitions of biomarkers have tried to describe what biomarkers are. In 2001, the International Programme on Chemical Safety, a joint venture established by the World Health Organization (WHO), the United Nations and the International Labor Organization, defined a biomarker as:

> "any *substance, structure, or process* that can be measured in the body or its products and *influence or predict* the incidence of outcome or disease" (International Programme on Chemical Safety, 2001, emphasis added)

In the same period, the National Institute of Health Biomarkers Definition Working Group proposed a definition according to which a biomarker is:

> "a *characteristic* that is objectively measured and evaluated as an *indicator* of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention." (Biomarkers Definitions Working Group, 2001, emphasis added)

If we consider the novelties characterising biomarkers described above, it can be concluded that neither of these two definitions was able to fully address the nature and the function of biomarkers. While the International Programme on Chemical Safety's definition emphasised that biomarkers can be substances, structures, or processes, without considering the possibility of using biological properties as biomarkers; the Biomarkers Definitions Working Group only recognised that any observable *characteristic* can be used as a biomarker.

Moreover, such definitions did not address precisely the relationship between the biomarker and the effect produced by the causal process. Although the International Programme on Chemical Safety's definition claimed that biomarkers influence or predict health outcomes, neither that definition nor the one proposed by the Biomarkers Definitions Working Group made explicit that diverse biomarkers can be related in different ways to the effect. Let us take Figure 18 illustrated above: in all those cases

biomarkers can predict the health outcome under study[13], in some of them biomarkers can also influence it. If we consider a biomarker directly involved in the process, as in the case of Figure 18(a), such a biomarker would influence the effect and could be a perfect predictor of the outcome. In addition, a factor correlated to the outcome due to the presence of a common cause, as in Figure 18(b), or a marker correlated to an unmeasured part of the process like in Figure 18(c) and 18(d), would allow researchers to predict (but not to influence) the effect. Finally, if we consider a background condition, as illustrated in Figure 18(d), it is obvious that such a condition would influence the outcome (its presence is necessary to cause the effect). This condition, furthermore, would be correlated to the causal effect, therefore it could be used as a predictor.

My proposal is that, in the biomarker definition, it is very useful to make explicit the possible ways in which biomarkers can be linked to the effect of the causal process under study; and if a definition omits this aspect, the reasons for the omission should be clarified. These different possibilities were not included in the two official biomarkers definitions, with the consequence that it is not easy to disambiguate the possible relationships between biomarkers and health effects. This omission, moreover, was not discussed or justified.

As for the use of biomarkers as indicators of processes, in addition, the term 'indicator' is an umbrella term that alone does not clarify the nature of biomarkers. Although, generally, it might be argued that indicators are used to infer conclusions on the phenomenon of interest (the indicandum), how such conclusions are inferred is not specified in the definition[14].

Notably, furthermore, nothing was said about the *use* of biomarkers. It is true that some biomarkers are used to predict certain health outcomes, as recognised by the International Programme on Chemical Safety's definition, however in other situations biomarkers are used to understand specific causal processes, like in the example described above of exposure to aflatoxin. While the Biomarkers Definitions Working Group's claim according to which biomarkers act as indicators of biological processes appears to be based on the assumption that biomarkers might provide information about such processes, the functions that biomarkers could play was not really made explicit. To address these

[13] For this reason biomarkers are often understood merely as predictive tools.
[14] In chapter 8 I will propose a more detailed discussion on the concept of indicators in the social sciences.

problems, the next sections will discuss in detail the process of biomarkers identification and how biomarkers are used. These discussions will then inform section 6.5, in which it will be proposed a new framework for biomarkers.

6.4.2 From data collection to biomarkers: the process of biomarkers identification

Thanks to new methods of data collection such as '-omic' technologies, large amounts of biological data are now available to scientists. Among them, researchers have to identify the right biomarkers. The process of biomarkers identification is made complex by the fact that biomarkers can be obtained from numerous types of data: for instance, it has been said that biomarkers can be obtained from the study of gene expression profiling (transcriptomics), or by analysing the metabolites in a specified biological sample (metabolomics). Multiple biological samples, in addition, can be studied by scientists to obtain information on metabolites, for instance blood and urine are often examined when looking for biomarkers, while in certain circumstances tissue blocks are obtained and analysed for that purpose.

Given that collecting biological data is often time-consuming and expensive, researchers need to narrow the list of biological data that might be used to obtain biomarkers in their specific study. It is for this reason that the search for biomarkers is generally informed by causal hypotheses and background knowledge. This consideration has been already explicated in some discussions about biomarkers. As an example, Jain (2015) suggested that:

> "For biomarkers to assume their rightful role, greater understanding of the mechanism of disease progression and therapeutic intervention is required." (Jain, 2015, p. 87)

A case in point illustrating the importance of causal hypotheses in the identification of biomarkers is the role played by Warburg's hypothesis (1956) in the search for cancer's biomarkers. Otto Warburg in the 1950s developed the causal hypothesis according to which, in contrast to normal differentiated cells (which rely on mitochondrial oxidative phosphorylation as the source for energy), cancer cells rely on anaerobic metabolism to generate the energy needed for cellular processes. Warburg's hypothesis became the central idea guiding the search for cancer biomarkers: as an example, biomarkers such as altered redox balance and intra-tumoral metabolic interactions have been identified and

studied by examining extensively data on cancer central carbon metabolism[15] in accordance with such a hypothesis.

Causal knowledge and hypotheses might play a crucial role from the very first stage of the process of biomarkers identification, that is the selection of the level at which data should be collected to identify possible biomarkers. Once that choice has been made, some data are produced by certain '-omic' technologies (like the mass spectrometry machines) and are cross-checked. For instance, in a serum profiling experiment conducted to reveal potential biomarkers of colorectal cancer, electrospray ionisation (ESI) tandem mass spectrometry was employed to quantitatively profile the concentrations of 26 amino acids that could potentially be used as biomarkers (Armitage & Barbas, 2014).

It is important to clarify that '-omic' data and biomarkers are not the same things: biomarkers can be obtained *from* '-omic' data. To understand better this distinction, let us consider Figure 19, representing a mass spectrometry datum for an individual in the sample. The horizontal axis represents the m/z-value, where m stands for ion's mass and z stands for the number of electrons taken from molecules to create single charged ions. The vertical axis represents the intensity or abundance of ion (the occurrence of a certain ion in relation to the occurrence of the most abundant ion). The peaks represented in Figure 19 stand for the charged ions originating from proteins in the sample. Such peaks allow researchers to detect the mass of the proteins, and from the mass researchers try to identify the specific protein molecules causing the peaks. For instance, Figure 19 shows a peak at a particular m/z value, 8000. Researchers then expect that there is a protein associated with the ion corresponding to that m/z value. If such a peak is found in many subjects (in other words, in a relatively large set of data), it is then possible to hypothesise that the peak could be used as a biomarker.

In such a way, scientists' set of data might be used to obtain a new biomarker. As already said above, in general, when a biomarker is identified, scientists need to obtain from their data a measure of an entity, quality or event.

At the following stage of the process leading to the identification of biomarkers, scientists examine the correlations between the putative biomarkers obtained from their data, the

[15] Central carbon metabolism employs a complex series of enzymatic steps to transform sugars into metabolic precursors (Noor et al., 2010).

measure concerning disease manifestations, and, in some cases, the measures of the putative cause producing the phenomenon (in the case of LIFEPATH, socio-economic indicators). In such a way, they start to select the biomarkers able to give information about the disease process.

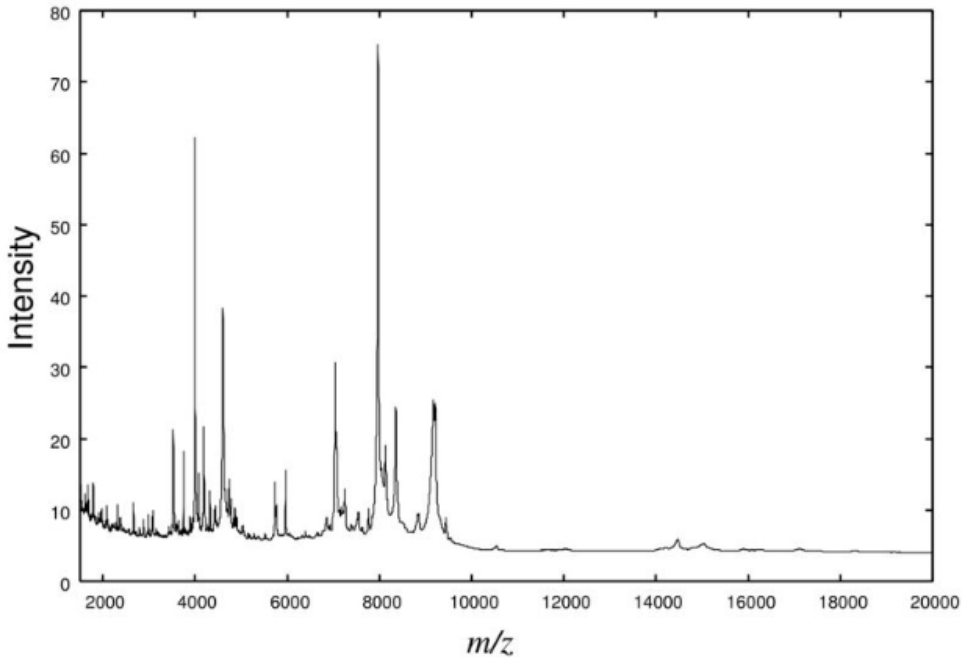

Figure 19: a mass spectrometry datum for an individual (Fushiki et al., 2006, p. 359).

Finally, scientists must select, among the list, those measures that might be used as markers of the phenomenon they want to study. Even at this stage, background knowledge as well as further evidence can guide scientists' choice. For example, in the case of biomarkers of colorectal cancer, 11 of the 26 amino acids that were measured were initially selected through statistical analysis, and among them 10 had already been reported as potential colorectal cancer biomarkers. This background knowledge strengthened the idea that the measures of such amino acids could actually be used as colorectal cancer biomarkers (Armitage & Barbas, 2014).

It is important to consider that the identification of a biomarker does not coincide with its validation. The criteria for validation are established according to the question that the study is intended to address and the level of certainty that is required for the result. Once the biomarker is established to be valid for the study, researchers examine whether the

biomarker is fit for the intended general use, and establish if the biomarker is ‘fit-for-purpose’ (Hunter et al., 2010, p. 4).

6.4.3 Biomarkers as intersecting signals

The way in which biomarkers are identified casts light on an important difference between socio-economic indicators and biomarkers. Socio-economic indicators are generally used to illustrate a specific condition: they are meant to offer information about an individual's access to social and economic resources. Figuratively, they are like portraits representing the social condition of some individuals: they represent how wealthy a person is, the conditions in which the person works, and the social relationships that such a person has (I shall discuss this point in detail in chapter 8).

On the contrary, biomarkers are, as the word says, just ‘markers’. Schulte and Perera (1993, pp. 13–14) described biomarkers in terms of ‘events’ in the continuum from exposure to disease; Chadeau-Hyam et al. (2011, p. 85), defined them as ‘possible intersecting signals’ representing parts of a process. It follows that more than one biomarker is generally required to gain a full understanding of the causal process under study, and that the role played by biomarkers is not to represent a situation, like in the case of socio-economic indicators, but to show the pathway taken by the process. An example can help us to understand this point. One of the main causal hypotheses in LIFEPATH is that experiencing stress and low socio-economic conditions over life might contribute to physiological dysregulations, that can be subsequently translated into certain diseases by means of prolonged activations of stress response systems (McEwen & Stellar, 1993).

To study this causal process starting from socio-economic experiences and leading to diseases, scientists in LIFEPATH first examined the process through which the body adapts to stressful circumstances in order to maintain physiological stability, known by the name of allostasis. When compensatory physiological mechanisms are repeatedly activated, the body is likely to experience a physiological wear-and-tear called allostatic load. To investigate the phenomenon of allostatic load, researchers selected 14 markers concerning four physiological systems: some markers linked to the neuroendocrine system (like the salivary cortisol), some markers concerning the immune and inflammatory system (such as the insulin-like growth factor), certain markers associated with the metabolic system (as low-density lipoprotein) and some markers from the

cardiovascular and respiratory system (such as blood pressure) (Barboza Solís et al., 2016).

On the one hand, researchers examined how these markers were linked to low socio-economic conditions during childhood by considering statistical correlations to parental occupation, income and education. Biomarkers were hence taken as signals representing the biological effects of stressful situations.

On the other hand, scientists in LIFEPATH relied on a growing literature supporting the claims that such biomarkers contribute to the development of certain forms of diseases. For instance, measures linked to inflammatory and immune stress responses had been used to discover certain pathways leading to diseases such as cancer (Castagné et al., 2016). Similarly, it is has been confirmed that people with a certain level of blood pressure are at risk of progressing to overt hypertension and cardiovascular diseases (Wang et al., 2007).

Biomarkers were hence used as signals of some mediating events (such as the secretion of cortisol) between the experience of chronic stress and the development of disease (like in Figure 18(a)). From the biological point of view, the use of biomarkers as intersecting signals was conceptualised by Vineis and Perera (2007), who provided the idea of a 'meet-in-the-middle' approach. According to that original notion, the discovery that biomarkers, already known to be related to particular exposures (like stress in our example), are also involved in the development of certain diseases would support the causal claims linking exposures and diseases.

From the philosophical point of view, the fact that biomarkers show the causal link connecting factors at different scales (from population-level factors such socio-economic experiences to individual-level factors like biological changes) has led some philosophers to claim that the concept of productive causality (i.e. how causes and effects are linked) requires reconceptualization (Illari & Russo, 2016). The classic approach to productive causality assumes that causality entails a physical connection between the cause and the effect. In general, it is assumed that a causal process is a continuum in spacetime allowing for the transmission or propagation of physical quantities (see also section 3.2.1). Reichenbach and Salmon developed the idea of productive causality and proposed that a process is causal if, and only if, was it marked, the mark would be transmitted to later stages of the same process (Reichenbach, 1958; Salmon, 1984).

Such a philosophical idea, however, is specifically tailored to physics and cannot clarify the role of biomarkers in causal studies. The process from socio-economic experiences to disease explored by means of biomarkers, indeed, involves non-physical aspects such as psychological events and factors. What follows is that it is not always possible to find a physical connection between the cause (socio-economic conditions) and the effect (disease). Furthermore, in biomarkers studies no marks are introduced to establish the causal nature of a process.

By considering these limitations, Illari and Russo suggest that the detection of signals should be reconceptualised to accept that a marker might be *something*[16] that is already transmitted by a causal process, rather than a 'physics' mark that should be introduced. This would appear in line with what is done in LIFEPATH, where researchers follow a longitudinal thread of marks (observing, for instance, that aflatoxin causes the formation of adducts, which in turn induce p53 mutations that are the same found in liver cancer). Biomarkers, in this sense, would be conceptualised as signals intersecting what is propagated from the cause to the effect.

6.5 Learning from LIFEPATH

The analysis of LIFEPATH as a case study has led me to investigate in detail the use of biomarkers in social epidemiology. In section 6.4.1, the two main definitions of biomarkers proposed by the International Programme on Chemical Safety and the National Institute of Health Biomarkers Definition Working Group have been discussed. I have considered that both these definitions only partly address the nature of biomarkers as measures of entities, qualities and events. Moreover, terms such as 'predictors' or 'indicators' are vague enough to allow for different relationships between the health effects and the biomarkers, and do not specify the nature of biomarkers as 'intersecting signals'. In sections 6.4.2, I have argued that causal knowledge and causal hypotheses are often crucial in order both to decide what biological data to collect and to identify biomarkers. In section 6.4.3, finally, I have used some LIFEPATH studies to clarify why biomarkers are conceptualised as intersecting signals. Taking into account all these

[16] Illari and Russo have spelt out the idea of mark transmission in terms of 'information transmission'. They have considered that in genome studies there is some possibility of a precise definition of information, given that genes are more clearly defined than '-omic' measures. In biomarkers studies, on the contrary, information is still not well-defined.

considerations, this section will be devoted to the development of some original proposals concerning how biomarkers are used and should be defined. First, I shall argue that a biomarker can be considered as such only if it is possible to articulate the relationship between the biomarker and the causal process under study. Then, I shall claim that biomarkers are used both to develop hypotheses about certain causal processes and to establish the presence of some causal processes.

6.5.1 Developing a relational framework for biomarkers

In this sub-section, I shall propose a relational framework capable of casting light on the particularity of the function that biomarkers play in research. My intuition is built on Leonelli's data relational framework (2015). Investigating the nature of data, Leonelli proposed:

> "[…] to view data as any product of research activities, ranging from artifacts such as photographs to symbols such as letters or numbers, which is collected, stored, and disseminated in order to be used as evidence for knowledge claims. […] Hence, any object can be considered as a datum as long as (1) it is treated as potential evidence for one or more claims about phenomena and (2) it is possible to circulate it [the object] among individuals […]. Within this relational framework, it is meaningless to ask what objects count as data in the abstract, because data are defined in terms of their function within specific processes of inquiry, rather than in terms of intrinsic properties. The question 'what is data?' can only be answered with reference to concrete research situations, in which investigators make specific decisions about what can be used as evidence for which claims." (Leonelli, 2015, pp. 817–818)

According to Leonelli's framework, the definition of a datum is always associated with a particular purpose: any object produced by human interactions with research instruments can be considered a datum, but only if it is examined within specific research circumstances. The same consideration, I suggest, can be applied to biomarkers. Not all measures count as biomarkers in every particular study, as well as not all products of research activities count as data in a specific situation.

This interpretation is in line with the process of biomarkers identification described in section 6.4.1. First, data produced through several data-collection techniques are collected (in general relying on causal knowledge and hypotheses); then, scientists have to apply

measurement procedures if the data obtained are not numeric, and next they have to understand whether such measures count as biomarkers or not. The only way to establish if biological measures count as biomarkers is by considering the specific research circumstances.

To give an example, in a recent paper Vineis et al. (2013, p. 465) described how biomarkers were selected in a study of breast cancer. Researchers examined 96 cases of breast cancer and the genome-wide methylation profile associated with each patient: considering casual hypotheses and statistical analyses, they collected data about 10 CpG sites (regions of DNA where a cytosine nucleotide and a guanine nucleotide appear consecutively on the same strand of nucleic acid) and measured the methylation fractions, that were already known to be strongly associated with breast cancer onset. Next, they observed the associations between such methylation fractions and hundreds of diverse exposure items, looking for candidates exhibiting statistical associations only with reproductive factors (in particular, ever having breastfed and the age at menopause, two well-known risk factors). Their decision was motivated by the hypothesis according to which such types of exposure can trigger processes that, through the methylation fractions of some sites, lead to breast cancer. Only the methylation fractions of 4 of the 10 CpG sites were finally identified as possible biomarkers linking reproductive factors and the development of breast cancer. It should be considered that, from this study, it does not follow that the methylation fractions of the other 6 CpG sites cannot be used to find biomarkers in different situations. Furthermore, it might be plausible that the biomarkers selected by Vineis et al. are used also as biomarkers of other health phenomena, such as other forms of cancer.

My proposal is that a measure can act as a biomarker of one causal phenomenon or of a variety of causal phenomena; and that anything that counts as a biomarker requires someone able to articulate the relationship between that biomarker and the causal process responsible for the phenomenon under study. To clarify this point, let us consider again Figure 18: my argument is that a measure can be said to be a biomarker of a particular phenomenon only if scientists can identify, among the five possibilities illustrated in Figure 18, which one properly represents the link between the biomarker and the phenomenon. In other words, biological measures become biomarkers when researchers develop causal hypotheses that allow them to locate the markers in specific positions in

relation to the causal processes responsible for the phenomena[17]. As recognised by Henderson et al. (1989), indeed:

> "The term 'biomarker' refers to the use made of a piece of information, rather than to a specific type of information" (Henderson et al., 1989, p. 65)

For instance, in the study conducted by Vineis et al. (2013, p. 465), the methylation fractions of the 4 CpG sites selected were hypothesised to be the right biomarkers directly linking reproductive factors and the development of breast cancer, in a situation similar to Figure 18(a).

The discussion on this feature can help in the development of an exhaustive definition of biomarkers. My first suggestion is that we can call a 'biomarker' any measure of a biological entity, quality or event that is used to obtain information about a specific biological process linking the cause A to the causal effect B. Any biomarker, in order to be used as such, must have the following characteristics: 1) can be objectively measurable, and 2) is linked to the causal process under study either because is directly involved in the process, or because is caused by the same causal factor A, or is caused by another unmeasurable biomarker involved in the process or is a background condition of the causal process (as illustrated in Figure 18).

This definition would solve some of the limitations characterising the two official definitions of biomarkers currently used that have been discussed in section 6.4.1. Indeed, the definition clarifies what can count as biomarkers, and makes explicit the relationship the can exist between a causal process and the biomarker associated with it. It remains to understand, however, what kind of information biomarkers can provide. In the next section I shall address this question.

6.5.2 Tracing causal processes

If biomarkers work as 'possible intersecting signals' representing parts of an active process (Chadeau-Hyam et al., 2011), the identification of several biomarkers linked to the same phenomenon can help to *trace* the causal process of disease formation.

At the methodological level, such an approach is in general called *process tracing*. This has been used as an umbrella term encompassing a variety of techniques (Gerring, 2006,

[17] If it were enough to establish a correlation between the biomarker and one part of the process, there would be no difference between a proxy for the process and a biomarker.

p. 173), however they all share the core idea of concatenation: causal events are linked and form a causal chain connecting some variables to the causal effect under study. This idea is strongly associated with the philosophical interpretation of mechanisms as processes through time and space along which a signal is propagated (Dowe, 2007; Reichenbach, 1958; Salmon, 1997), as described in section 3.2.1. This means that, if researchers hypothesise that the factor C causes the effect E, they assume that C causes E through a set of causal steps, as described in Figure 20:

$$C \longrightarrow M_1 \longrightarrow M_2 \longrightarrow M_3 \longrightarrow E$$

Figure 20. The assumption behind process tracing is that the causal path from C to E is mediated through certain factors such as M_1, M_2 and M_3. In the figure, the arrows represent causal links.

It is easy to find a similarity with the way in which biomarkers are used: it has already been said that Schulte and Perera (1993, pp. 13–14) described biomarkers in terms of 'events' in the continuum from exposure to disease; and that similarly the National Institute of Health Biomarkers Definition Working Group claimed that biomarkers are indicators of *biological processes*. The analogy between biomarkers and mediating factors M_1, M_2 and M_3, furthermore, appears even more intuitive once we consider again Figure 18, where biomarkers can be used to get insights into the intermediate steps of the process from the putative cause to the effect, like in Figure 18(a), 18(c) and 18(d).

There is, however, one aspect that needs to be clarified. So far, it has been claimed that biomarkers are used to 'trace', 'discover', 'cast light on' causal processes. These verbs, however, do not clarify if biomarkers are employed to *develop hypotheses* about such causal processes, or if they actually *count as evidence for* causal hypotheses. In the former case, biomarkers would help to identify where and how to collect evidence of such processes; in the latter case, biomarkers would provide evidence that might allow scientists to establish the presence of causal processes.

The distinction between building and confirming causal hypotheses has been proposed within the process tracing literature: according to Bennett and Checkel, for instance, process tracing can be both 'bottom-up', aimed at developing causal hypotheses, and 'top-down', aimed at testing hypotheses (Bennett & Checkel, 2015, p. 7). Similarly, Beach and Perderson (2013) recognised that different variants of process tracing can be

distinguished considering whether their goal is to develop or test theorised causal mechanisms. The same division, furthermore, is familiar in philosophy, where logical empiricists drew the distinction between the context of discovery and the context of justification (Reichenbach, 1949, pp. 433–434).

If researchers have little prior knowledge of the phenomenon in question, process tracing is claimed to be performed through inductive studies, with the aim to build new causal hypotheses (Beach & Pedersen, 2013). In such a situation, researchers act as detectives: they start analysing events backward through time, like detectives try to piece together the last few hours of the victim's life. Due to the lack of prior knowledge, a massive amount of information is gathered: these pieces of information may or may not become components of the causal hypotheses. As for the situations where causal hypotheses already exist, process tracing proceeds more deductively: researchers try to translate theoretical expectations into observable implications or signals that should be found by exploring the process.

Is it possible to establish whether biomarkers are used to develop or to test causal hypotheses? If we consider what I have argued in section 6.5.1, biomarkers can be used as such only if researchers can at least hypothesise the relationships between such biomarkers and the phenomenon. In this sense, almost all biomarkers studies can be characterised as process tracing studies aimed at building new causal hypotheses.

The study conducted by Vineis et al. (2013, p. 465) described in section 6.5.1, for instance, was aimed at building new causal hypotheses about the process starting from certain types of exposure and leading to breast cancer. The identification of the methylation fractions of the 4 CpG sites, according to the authors, ‘lent credibility’ to the causal hypothesis according to which exposure to reproductive factors (in particular, ever having breastfed and the age at menopause) influences such methylation fractions, that in turn lead to the development of breast cancer. To confirm the presence of such a causal process, however, scientists would need further studies, which might in turn be based on biomarkers. Biomarkers, indeed, might offer useful evidence to test causal hypotheses.

In the example described in section 6.4.2, for instance, the starting point was an existing causal hypothesis: experiencing stressful and low socio-economic conditions over life contributes to physiological dysregulations, that can be subsequently translated into disease by means of prolonged activations of stress response systems (McEwen & Stellar,

1993). To test the hypothesis, Barboza Solís et al. (2016) identified 14 signals that were hypothesised to be involved in the process. It was already known that such factors were linked to disease, but researchers expected that they were also correlated to low socio-economic conditions. They tested the hypothesis examining the statistical correlations between these measures and parental occupation, income and education, and the final evidence supported their hypothesis.

What follows from these examples is that biomarkers are used both to develop causal hypotheses concerning the causal pathways leading to health outcomes and to provide evidence for them. Of course, there are situations where biomarkers are just used to predict specific outcomes, and scientists have no interest in hypothesising or testing causal processes. Nevertheless, even in such cases researchers rely on the development and test of causal hypotheses concerning at least part of the causal processes.

In some cases, for example, biomarkers are used as substitutes of clinical endpoints. When markers have this function, they are called 'surrogate markers', and are expected to predict how a patient feels, functions or survives (surrogate biomarkers are in general used to establish the causal effects triggered by clinical trials). As recognised by some authors (Fleming & DeMets, 1996; Katz, 2004), the correlation of the marker with the clinical enpoint is a necessary but not sufficient condition to confer the 'surrogate' status to a measure (a famous expression proposed by Fleming and DeMets (1996, p. 605) on this point is "a correlate does not a surrogate make"). Before using a marker as a surrogate, scientists need to:

> "[…] understand the (biological) relationship between the surrogate and the clinical outcome. In other words, we need to know if and how the surrogate and the clinical presentation are intrinsically biologically related to one another […] there may be many pathophysiologic pathways that contribute to disease manifestations, and the surrogate may be "in" some, any, or none of these pathways" (Katz, 2004, p. 193)

In order to use surrogate markers, consequently, the part of the process linked to the biomarkers and leading to the endpoint needs to be at least hypothesised. This aspect has already been discussed more than twenty years ago by Fleming and DeMets, who argued that, before using a surrogate biomarker, researchers need to have proper understanding both of the causal pathways of the disease and of the intervention's mechanism (1996, p.

611). They identified, as illustrated in Figure 21, different possible relationships between the surrogate biomarkers and the causal process. In these terms, also the selection of surrogate biomarkers is based on the development or test of causal hypotheses.

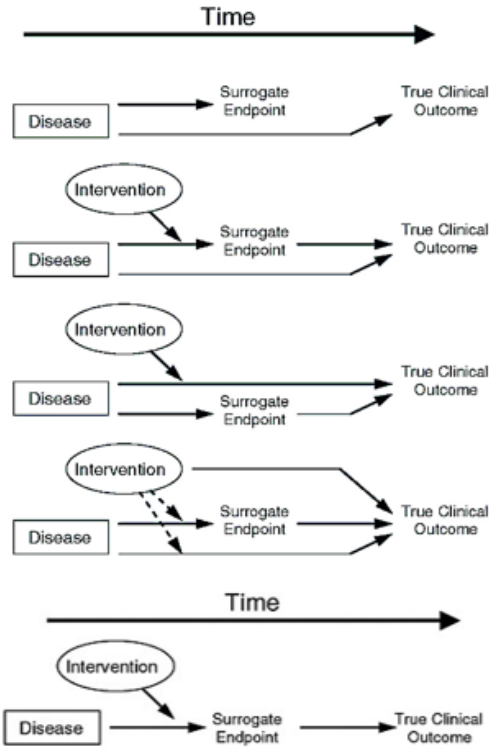

Figure 21. Possible relationships between the surrogate biomarker and the process causing the clinical outcome. Figure retrieved from Fleming and DeMets (1996, p. 606).

In section 6.5.1, I have argued that any measure of a biological entity, quality and event that is used to obtain information about a specific biological process linking the cause A to the causal effect B can count as a biomarker.

Overall, putting together all the considerations discussed in this chapter, I can now complete the definition of biomarkers in the following way: a 'biomarker' is any measure of a biological entity, quality or event that helps researchers to develop a causal hypothesis or to provide evidence for a hypothesised causal process linking the cause A to the causal effect B. Such knowledge can be used to enhance our understanding of the causal process, or to predict an effect according to what we know of the process. Furthermore, any biomarker must have the following characteristics: 1) can be objectively measurable, and 2) is linked to the causal process under study either because is directly involved in the process, or because is caused by the same causal factor A, is caused by

another unmeasurable factor involved in the process or is a background condition of the causal process.

6.6 Conclusion

In this chapter, I have described the LIFEPATH project to cast light on a novel approach through which biological data are used. I have explained where LIFEPATH comes from, and what is the main purpose of this project. Next, I have examined the data used within the project: the social data used to obtain measures that indicate people's socio-economic position, and the biological data used to identify biomarkers. In the rest of the chapter, I have focused my attention on the concept and use of biomarkers. I have explained why biomarkers can be considered a novelty in epidemiology, and how they are identified and conceptualised. I have argued that the process of biomarkers identification requires causal knowledge and hypotheses, and that biomarkers are in general understood as the intersecting signals between exposure and health effects. In section 6.5, I have used the information collected by studying LIFEPATH to develop a more precise account of biomarkers. By following Leonelli's data relational framework, in section 6.5.1 I have argued that every measure can act as a biomarker of one phenomenon or of a variety of phenomena, and that anything that counts as a biomarker requires someone able to conceptualise the relationship between that biomarker and the causal process under study. In section 6.5.2, finally, I have claimed that biomarkers are used both to hypothesise the presence of some causal processes, and to test them. In conclusion, putting together all these considerations, I have developed a new definition of biomarkers illuminating all the characteristics that biological measures need to have in order to count as biomarkers. According to this definition, causal reasoning is the key aspect differentiating biomarkers from usual indicators and proxies: when biomarkers are identified and used, researchers cannot avoid answering causal questions about the phenomenon to be explained and the link between it and the selected biomarkers.

Part III

Big data for causal discoveries

The notion of mechanism belongs to the everyday causal vocabulary of all social scientists. To give an example, social psychologists discuss multiple social mechanisms through which individuals acquire normative beliefs to conform to the expectations of their peers so as to obtain social approval and avoid not being accepted. Similarly, political scientists focus on institutional and voting mechanisms leading to the creation of legislative outcomes. Economists, furthermore, identify market-related mechanisms causing resource allocation, price discovery, assignment of property rights and many other things besides.

As discussed in chapter 3, there are different ways in which social scientists can conceptualise causal mechanisms in scientific research. Researchers might understand mechanisms in terms of complex systems made of organised entities and activities, but they could also consider mechanisms as causal processes linking the cause and the effect through time and space. This distinction is partly aligned with another distinction made in the philosophical literature on causal mechanisms, according to which mechanisms can be thought either to 'constitute' the phenomena-to-be-explained or to 'produce' the outcomes. In the former case, mechanisms are called constitutive and are claimed to *constitute* the phenomena, in the latter case, mechanisms are called etiological and are claimed to *cause* the phenomena under examination.

In the social sciences' debates, the distinction between these types of mechanisms is often not explored in detail. Nevertheless, depending on whether mechanisms relative to particular phenomena are understood as constitutive or etiological, researchers can search for different types of evidence in order to establish their presence. In the case of constitutive mechanisms, researchers need to collect enough evidence to support the claim that the constitutive relevant parts of a mechanism actualise the capacity that makes possible the phenomenon-to-be-explained. Regarding etiological mechanisms, researchers need to find evidence of the process producing the effect under study.

This part of the thesis has a twofold aim. On the one hand, by providing a clear discussion about the different ways in which constitutive and etiological mechanisms are

conceptualised, this part will clarify how social scientists can study different types of mechanisms. On the other hand, the distinction between these forms of mechanisms will be used to develop possible solutions to the third challenge guiding this thesis, namely how big data can be used, in the social sciences, to obtain evidence of causality that goes beyond correlations. In particular, in these chapters I will study how to obtain evidence of mechanisms from big data.

The next two chapters are characterised by different approaches. In chapter 7, I shall use the existing literature on constitutive mechanisms to discuss how constitutive mechanisms are conceptualised and how big data might help to study them. The existing discussions on constitutive mechanisms are concerned with metaphysical questions and make use of very complex formalisations. I shall offer a less formalised account that attempts to bridge the gap between such metaphysical discussions and the questions of how to obtain evidence of mechanisms. The account I will propose, however, will only partially succeed in dealing with the methodological challenges social scientists in general have to overcome when they collect evidence of mechanisms. In the last sections of chapter 7, I shall examine this problem. I shall argue that, in order to contribute to the scientific debate on big data and causality, it is important to provide philosophical accounts that are well-informed about scientific procedures and problems. These observations will lead me to conclude that, although social scientists might be able, under specific circumstances, to identify constitutive relevant parts of social mechanisms, the philosophical account of constitutive mechanism can only moderately contribute to clarifying how big data can be used in scientific research to collect evidence of causality that goes beyond correlations.

Chapter 8 will be based on a different approach. The starting point will be the detailed observation of how data have been used in scientific research to trace etiological mechanisms. Combining the discussion about how social scientists use social data, with the analysis of biomarkers of chapter 6, I shall develop a scientifically-informed philosophical discussion on the use of markers to trace etiological mechanisms. This account, I shall argue, is not only consistent with some scientific procedures, but might also improve the way in which big data are used in causal studies.

7 Uncovering constitutive mechanisms with big data

7.1 Introduction

A constitutive mechanism is in general understood as a mechanism made of certain relevant parts that make the phenomenon-to-be-explained possible. How to account for constitutive mechanisms is a current topic in philosophical debates, however little has been said about how to link this debate to social scientists' mechanistic discussions. My aim in this chapter is to fill this gap by providing a tentative account of constitutive relevance that works with simple examples of social mechanisms, and to use such an account to provide a solution on how to use big data to obtain evidence of constitutive mechanisms.

This chapter is motivated by three considerations. First, there are social phenomena that have the typical characteristics of constitutive mechanisms. The economic phenomenon known by the name of rebound effect, for instance, is in general described as a constitutive mechanism whereby increased energy efficiency leads to growing energy consumption and offsets any energy saving (Herrmann-Pillath, 2015). Such a phenomenon is made possible by the activities of its constitutive parts: the individuals who, thanks to the increased energy efficiency, increase their demand for energy (with a corresponding change in demand at the macro level) and the individuals who decide to spend the money saved thanks to the energy efficiency on other goods that require, in turn, energy (with a corresponding increased energy consumption at the macro level). As I shall discuss in this chapter, further examples might be the activities performed by a social organisation, Schelling's mechanistic model of segregation, and firms' behaviours.

Second, the philosophical debate on constitutive relevance has become very complex over time, and the use of formalisations makes it difficult for non-expert readers to follow the discussion. While, in order to discuss the existing accounts of constitutive relevance, I shall need to use such formalisations, in my account I will try to limit them to make my proposal as comprehensible as possible.

Third, the way in which constitutive mechanisms are understood in the social sciences can have important moral implications. A simple example is the case of a clandestine cell system like a terror cell: in such a case, an accurate constitutive relevance approach is essential in order to identify *all and only* the constitutive parts of the cell.

The idea according to which constitutive mechanisms have constitutive parts that make the phenomenon-to-be-explained possible entails that such parts *do not cause* the phenomenon. The general assumption behind the notion of constitutive mechanisms, indeed, is that there are no causal relationships between the constitutive parts and the phenomenon performed by the mechanism. This assumption is justified by three different observations.

First, the behaviour of constitutive parts and that of the whole phenomenon are overlapping, whereas in the case of causation the cause and the effects are distinct and non-overlapping. Second, the behaviour of components occurs during the behaviour of the mechanism as a whole, while it is in general recognised that causes precede their effects. Third, causation in general implies asymmetry (causes produce their effects, not vice versa), but the relationship between a component and the behaviour of a mechanism as a whole is always symmetrical: one cannot intervene to change the value of the whole without changing the value of at least one component.

To illustrate these points, let us consider a social constitutive mechanism. Perhaps the most obvious example is a social organisation consisting of an organised group of individuals linked through a structured set of relationships to each other. As a result of its members interacting with each other, a social organisation can have an impact on the world that its members alone would not have if they were not parts of such an organisation. A good example is an orchestra that has the power to produce Ravel's Boléro because its components—the conductor and the players with their instruments—inter-relate in an organised way. If these components were not organised into an orchestra, the conductor and the players with their instruments would not have the causal power to produce that harmonious music. Furthermore, the playing orchestra *does not cause* the execution of Ravel's Boléro: the execution is not the final stage of a causal sequence, but a phenomenon constituted by the whole activities of the conductor and the players with their instruments.

More specifically, if we consider the relationship between the orchestra and one of its components, the trumpet player X, it can be claimed that this relationship does not appear causal because it is impossible to isolate the behaviour of its trumpet player X from the behaviour of the orchestra: such behaviours overlap. Similarly, the orchestra and the trumpet player X act at the same time: if the trumpet player X begins to play at the right

time, it is impossible to say that he started playing his part of Ravel's Boléro, and only later the orchestra began to play the trumpet part of Ravel's Boléro. Finally, the only way to change the behaviour of the orchestra (i.e. the act of the orchestra playing Ravel's Boléro) is to change either the behaviour of one of its players, or the behaviour of the conductor. It is not possible to change the activity of the orchestra playing Ravel's Boléro without intervening on one of its components.

If social scientists want to collect evidence about the mechanism producing a phenomenon such as the execution of Ravel's Boléro, they have to ask what kinds of more basic entities and behaviours can, taken together, possibly result in the phenomenon-to-be-explained. Next, they have to verify whether the mechanism is actually made of such entities and behaviours. The first part of the chapter will explore in detail *how*, according to the literature, the relevant constitutive entities and behaviours of a mechanism can be recognised. This analysis will lead me to discuss some of the limitations associated with the major accounts of constitutive relevance, and to provide a novel account. After having discussed my proposal, I shall argue that, especially when the phenomenon-to-be-explained is not a causal capacity but an actual phenomenon, big data can improve our ability to collect evidence about the constitutive entities responsible for it. This possibility, however, is limited to the situations where it is possible to use the constitutive relevance account. In the social sciences, indeed, there are several cases where such an account is not feasible or helpful to uncover the operations of causal mechanisms. Overall, moreover, the formalisations used in the general discussions on constitutive relevance and the distance between such discussions and scientific practices might pose some challenges if the aim is to clarify how big data studies on constitutive mechanisms are or can be conducted.

The chapter is structured as follows: in section 7.2 I shall describe Craver's mutual manipulability account and I shall discuss the problems associated with such an account and some of the solutions that have been proposed in the literature. In section 7.3 I shall analyse in detail three recent proposals for constitutive relevance: Baumgartner and Casini's abductive theory of constitution, Krickel's causation-based account for constitutive relevance, and Baumgartner, Casini and Krickel's horizontal surgicality account. The description of these proposals will be accompanied by a thorough study of their methodological and conceptual limitations. In section 7.4 a new constitutive relevance account, called the extended horizontal surgicality account, will be proposed,

and in 7.5 this account will be used to explore whether big data can enhance our ability to collect evidence of mechanisms. In section 7.6 I will explore under which conditions the extended horizontal surgicality account can be used to collect evidence of social constitutive mechanisms. Finally, section 7.7 will provide a general consideration about the usefulness of the discussions on constitutive relevance to overcome the third challenge guiding this thesis, namely how big data can be used to collect, in the social sciences, evidence of causality that goes beyond correlations. Section 7.8 will conclude the chapter.

7.2 Craver's mutual manipulability approach

The most prominent account of how to identify the constitutive parts of a mechanism is Carl Craver's mutual manipulability approach to constitutive relevance (2007). According to Craver, if we want to identify the relevant parts that constitute the phenomenon, the first step is to exclude what acts just as a background condition and what has only sterile effects.

To illustrate what a background condition is, we can consider the presence of oxygen in the theatre in which the orchestra plays Ravel's Boléro: without oxygen, certainly the orchestra could not play, nevertheless the presence of oxygen is in general not considered a constitutive part of the causal capacity of the orchestra. It is important to note that the identification of background conditions is pragmatic and generally based on the assumption that such conditions (like the presence of oxygen in a room) are in general stable (Franklin-Hall, 2015).

The concept of sterile effects is more complex: let us imagine that the orchestra playing Ravel's Boléro is composed of players playing their instruments, and that the activity of playing an instrument causes to the trumpet player the immediate reddening of his face. This reddening does not produce any change in the other components of the orchestra, and even if it causes an effect on the other players (perhaps some players could smile because of that), the changes it produces make no difference to the orchestra playing Ravel's Boléro. Consequently, the reddening is considered a sterile effect.

The illustration of this possible scenario is proposed in Figure 22. Let us take S to be the mechanism as a whole (the orchestra); Ψ the behaviour of S that is to be explained (the activity of the orchestra playing Ravel's Boléro), X one of the entities composing the mechanism (the trumpet player), and Φ_1 and Φ_2 the activities exhibited by X (playing the

trumpet and reddening the face). The expression X's Φ_1-ing means that the entity X is exhibiting the behaviour Φ_1, while the expression S's Ψ-ing means that the mechanism S is exhibiting the behaviour Ψ. The dotted line represents the constitutive relationship, while the arrow represents a causal link. X's Φ_2-ing (the reddening face) is not constitutive of S's Ψ-ing. It is just a sterile effect.

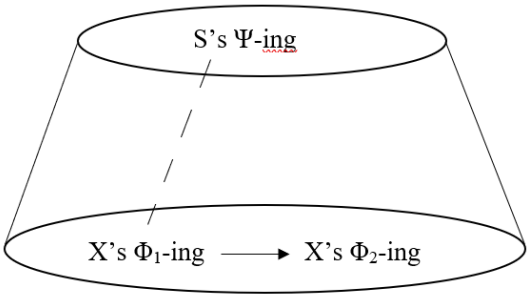

Figure 22. The representation of a constitutive mechanism. While X's Φ_1-ing (the player's activity of playing the trumpet) is constitutive of S's Ψ-ing, the effect of X's Φ_1-ing, X's Φ_2-ing is not constitutive of S's Ψ-ing. X's Φ_2-ing is just a sterile effect.

After having ruled out parts acting as background conditions or producing only sterile effects, in Craver's view, scientists should test whether the component and the phenomenon are *mutually manipulable* through an *ideal intervention*. The notion of ideal interventions was developed by Woodward and Hitchcock (2003, pp. 12–13) and has been widely used in the literature on constitutive mechanisms. According to this notion, an intervention I on an acting entity represented by the variable A is defined with respect to a second acting entity represented by the variable B, and is ideal if it meets the following conditions:

1) I causes A;

2) I is not causally relevant to B through a route that excludes A. That is, any directed causal path from I to B goes through A: I does not directly cause B and does not cause any cause of B except for A and other possible causes built into the connection I − A – B;

3) I is not correlated to any variable Z that causes B and that is on a directed path that does not go through A;

4) I acts as a switch for all other variables that cause A. That is, certain values of I are such that when I attains those values, A ceases to depend on the values of other variables that cause A and instead depends only on the value taken by I.

Imagine we want to intervene on A, the trumpet player playing in the orchestra, with respect to B, the orchestra playing Ravel's Boléro. We intervene by making the trumpet player stop playing by tying his arms. Such an intervention is ideal because:

1) our intervention causes A, a change in the trumpet player's behaviour;
2) our intervention is not causally relevant to the orchestra playing Ravel's Boléro through any route that excludes A, the trumpet player who plays Ravel's Boléro. In other words, our intervention does not change the behaviour of the orchestra through alternative routes involving other parts, such as the behaviour of the conductor or the behaviour of other players;
3) if, due to our intervention, the behaviour of the orchestra is changed by any other part Z (for instance by another player) this change goes through A, the trumpet player who plays Ravel's Boléro;
4) our intervention makes A, the trumpet player playing Ravel's Boléro, depend only on the fact that we have tied his arms: his behaviour no longer depends on his ability to read the music, to blow or to move his fingers on the trumpet to reproduce the notes written on the music sheet.

If we take S to be the mechanism as a whole (in the example, the orchestra); Ψ the behaviour of S that is to be explained (the execution of Ravel's Boléro), X one of the entities composing the mechanism (the trumpet player with the trumpet), and Φ the activity performed by X (the execution of the trumpet part of Ravel's Boléro), we are now able to understand Craver's mutual manipulability account, consisting in two manipulative steps:

"(*Mutual Manipulability*) X's Φ-ing and S's Ψ-ing are mutually manipulable iff:

1) there is an ideal intervention on X's Φ-ing with respect to S's Ψ-ing that changes S's Ψ-ing;

2) there is an ideal intervention on S's Ψ-ing with respect to X's Φ-ing that changes X's Φ-ing." (Craver 2007b, 153).

It is possible to illustrate this account by using again the example of the orchestra. The trumpet player and the orchestra are mutually manipulable iff:

1) it is possible to ideally intervene on the trumpet player with respect to the orchestra playing Ravel's Boléro, for instance making him stop playing, and to cause also a change in the phenomenon of the orchestra playing Ravel's Boléro (namely, the trumpet's part will not be played, therefore the Boléro will be partly played);

2) there is an ideal intervention on the orchestra playing Ravel's Boléro that makes the trumpet player modify his behaviour, for instance such an intervention could make him stop playing.

Note that the mutual manipulability account entails ideal interventions, where an intervention I on X's Φ-ing with respect to S's Ψ-ing is ideal if and only if X's Φ-ing changes only due to the influence of I, S's Ψ-ing is not changed by I directly and all the other variables are held fixed.

The first justification behind the notion of mutual manipulability is that bottom-up ideal interventions (interventions on X's Φ-ing with respect to S's Ψ-ing) cannot rule out the possibility that X's Φ-ing is just a background condition. Furthermore, ideally intervening on S's Ψ-ing with respect to X's Φ-ing does not exclude the possibility that X's Φ-ing is just correlated to the intervention: Craver (2007, p. 151), for instance, observed that in brain studies an intervention on the mechanism performing a cognitive task can increase blood flow even if the increase in blood flow is not a constitutive part of the mechanism performing such a task. As a consequence, only if X's Φ-ing and S's Ψ-ing are mutually manipulable, scientists can establish that X's Φ-ing is constitutive of S's Ψ-ing.

7.2.1 Two problems for the mutual manipulability approach

Recently, Craver's mutual manipulability approach has been subject to criticism (Baumgartner & Gebharter, 2016; Leuridan, 2012). Since the notion of ideal interventions is a tool developed to uncover causal relationships, while constitutive relevance is said to be a non-causal relation, it has been argued that Craver's use of such a notion is not appropriate. Craver himself supported the idea that constitutive relations are non-causal,

nevertheless his appeal to ideal interventions, under Woodward's assumptions, inevitably entails causal relationships.

In the interventionist account, a variable A can be considered a (direct) cause of a variable B if it is possible to satisfy some requirements about the intervention I on A with respect to B. The requirements listed, however, are the same used by Craver in his mutual manipulability account, namely: 1) A changes only due to changes in I; 2) I does not change B directly; 3) there is not a further variable Z that is not on the causal path between I and A that is changed by I and influences B; and 4) I is not probabilistically correlated to a further variable Z that causes B over a path not going through A (Woodward, 2005; Woodward & Hitchcock, 2003).

This consideration has been mentioned by certain researchers (Baumgartner & Gebharter, 2016; Leuridan, 2012; Romero, 2015) to show the problem characterising Craver's proposal: while the notion of an ideal intervention is described as an intervention on A with respect to B to characterise what it is for A to cause B, the use of such a notion in Craver's mutual manipulability account is aimed at showing that there is a non-causal constitutive relevance relationship between A and B.

There is another criticism, moreover, of the mutual manipulability account: even accepting that ideal interventions can be used to identify non-causal relationships, it appears impossible to ideally intervene on a constitutive mechanism. According to interventionism, if there is a correlation between two variables A and B when an intervention is performed, the correlation is always due either to the fact that one of the variables causes the other, or to the fact that the intervention is a common cause of them. Now, if A is a constitutive part of B, it should be impossible for one variable to cause the other. As a consequence, the only remaining option is that the intervention is a common cause of A and B. This solution, however, is clearly in contrast with the definition of an ideal intervention, because an ideal intervention on A with respect to B cannot cause B.

To address these issues, diverse adjustments and solutions have been proposed. First, according to some (Harinen, 2014; Leuridan, 2012), given that constitutive relationships satisfy the interventionist criteria for causation, the problem can be solved by accepting that constitutive relevance turns out to be a causal relation. While this solution might allow for the use of the notion of ideal interventions to describe constitutive relevant parts, it leads to a further problem. In the introduction it has been observed that i) the behaviour

of the constitutive parts and that of the phenomenon are always overlapping, ii) the behaviour of such parts occurs during the behaviour of the mechanism as a whole, and iii) the relationship between these parts' behaviours and the behaviour of a mechanism as a whole is always symmetrical. All these observations have been used by other researchers to reject the claim that constitutive relevance is a causal relation. In line with these authors, this chapter is focused on the solutions based on the assumption that constitutive relevance is not a causal relationship.

Certain authors have proposed adjustments to save the non-causal nature of constitutive relevance. For instance, Woodward himself (2014, pp. 333–334) has offered a modified account of ideal interventions, often called ideal* interventions, that can be spelt out in these terms:

> I is an ideal* intervention on A with respect to B if, and only if, I satisfies the following conditions:
>
> 1) I causes A;
>
> 2) any directed path from I to B goes through A or through a variable A* which is related to A in terms of constitutive relevance;
>
> 3) I is (statistically) independent of every cause of B which is neither located on a path through A nor on a path through a variable A* which is related to A in terms of constitutive relevance;
>
> 4) I acts as a switch for all other variables that cause A. That is, certain values of I are such that when I attains those values, A ceases to depend on the values of other variables that cause A and instead depends only on the value taken by I.

In other words, as shown in Figure 23, an intervention I on A with respect to B is ideal* if the intervention changes B, while all other variables that are not on the causal path between I, A and B are kept fixed, except for those variables that I, A and B non-causally depend on (like A* in Figure 23), that can change (Woodward, 2014).

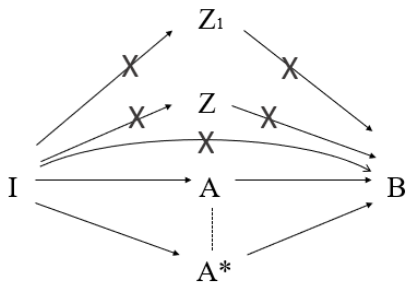

Figure 23. The ideal* intervention I can be a common cause of A and A* because they are related by a non-causal dependency relation.

This notion of ideal* intervention, however, was implicitly linked by Woodward to another modified notion, according to which A is a cause of B only if there is a possible ideal* intervention on A with respect to B. This modified notion, although was not presented by Woodward in an explicit formulation, was highlighted and made explicit by Baumgartner and Gebharter, who drew the following conclusion from Woodward's assumptions:

> "A is a cause of B with respect to the variable set **V** if and only if there possibly exists an ideal* intervention on A with respect to B, when all variables in **V** that are not on a causal path from A to B and are not related in terms of supervenience to A or B are held fixed, such that the value or the probability distribution of B changes." (Baumgartner & Gebharter, 2016, p. 745)

Although, apparently, the notion of ideal* intervention accommodates constitutive relevance relationships, the problem about the nature of the relationship between a component and the mechanism's phenomenon-to-be-explained remains unsolved both within the ideal and ideal* intervention frameworks. Indeed, the way in which ideal* interventions are defined leads to the conclusion that, if an ideal* intervention is performed on a mechanism's phenomenon with respect to its acting component, and the acting component is changed because of that, the acting component and the mechanism's phenomenon should be considered causally related. The problem emerges because in such a case we are not intervening on a variable with respect to another variable, but we are intervening on a mechanism's phenomenon with respect to its acting component[18]. Figure

[18] According to Woodward and Hitchcock (2003, pp. 12–13) an ideal or ideal* intervention on an acting entity is always defined with respect to a second acting entity.

24 clarifies the difference: in Figure 24(a) we intervene on the acting entity A with respect to the acting entity B, the intervention changes also A* because A* is related to A by a non-causal dependency relation. We can hence conclude that there is a causal relationship between A and B, and there is a non-causal relationship between A and A*. In Figure 24(b), we intervene on A, the mechanism exhibiting the phenomenon, with respect to its component A*. As a consequence, the relationship between A and A* appears both constitutive and causal (Baumgartner, 2013).

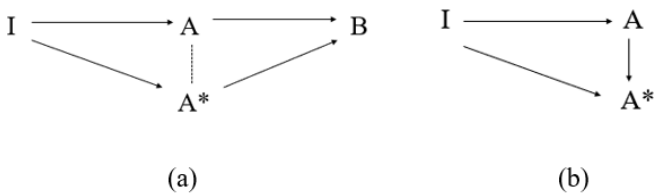

Figure 24. In ideal* interventions, the relationship between A and its acting component A* appears both constitutive and causal. In (a) the ideal* intervention on A with respect to B can be a common cause of A and A* because they are related by a non-causal dependency relation; in (b) the ideal* intervention is on A with respect to its acting component A*. Even if A* is constitutive of A, if the ideal* intervention changes A* it follows that there is a causal relationship between A and A*.

Another solution has been proposed by Romero (2015), who has defined constitutive relevance not in terms of ideal interventions but in terms of fat-handed[19] interventions (i.e. common causes of two or more variables): every intervention on the phenomenon is a fat-handed intervention on the phenomenon and at least one component, every intervention on a component is a fat-handed intervention on the component and the phenomenon. Similarly, Baumgartner and Gebharter (2015) have claimed that, in order to establish constitutive relevance, scientists should find out that every intervention on the phenomenon is a fat-handed intervention on the phenomenon and at least one of its components.

Finally, three recent proposals have been discussed by Baumgartner and Casini (2017), Krickel (2018), and by these three authors together (Baumgartner et al., 2018). To begin

[19] The expression 'fat-handed' is used to denote something that 'grabs' several things at the same time. A fat-handed intervention cannot change one variable without necessarily modifying other variables.

with, Baumgartner and Casini further develop the concept of fat-handed interventions. Krickel's account, instead, is based on the idea that only if we distinguish between parts acting in different spatio-temporal regions and between different temporal sequences involved in the phenomenon-to-be-explained, we will be able to identify what acts as a relevant constitutive part. Their latest proposal, finally, partly combines such ideas in a new horizontal surgicality account. In the next section, these three proposals will be described and analysed in detail.

7.3 Three recent proposals

7.3.1 Baumgartner and Casini's abductive theory of constitution

Baumgartner and Casini (2017) have developed their account from the consideration that all interventions on a constitutive mechanism should be fat-handed. In other words, rather than considering, as Craver, the possibility of top-down and bottom-up interventions on mechanisms as the main feature of constitution, Baumgartner and Casini claim that *the impossibility of such interventions* is the key feature of constitutive relevance.

The authors suggest that there is a difference between the so-called top-down and bottom-up interventions. Every intervention on the mechanism's behaviour S's Ψ-ing entails a change in one behaviour Φ of one of its constituents X (that means, every cause of S's Ψ-ing necessarily is a common cause of S's Ψ-ing and at least one constituent's behaviour X's Φ-ing). However, not every intervention on the behaviour of a constitutive part X's Φ-ing entails a change in the phenomenon S's Ψ-ing. In other words, it is impossible to conduct top-down interventions on S's Ψ-ing without changing X's Φ-ing, but it is possible to intervene with a bottom-up intervention on X's Φ-ing without changing S's Ψ-ing.

Bottom-up interventions are possible for two reasons. To begin with, it may happen that the change of the behaviour Φ of X is not relevant to S's Ψ-ing, because there are two values X's Φ_1-ing and X's Φ_2-ing that constitute the same value of S's Ψ-ing. A case in point is an intervention on the behaviour Φ of a constitutive part X that replaces one of its relevant aspects with respect to S's Ψ-ing with another aspect that can constitute the same value of S's Ψ-ing and X's Φ-ing. Moreover, another possibility is that the behaviour we are changing is not actually constitutive of the behaviour under study. Φ might be a sterile effect performed by the constitutive part X. If we consider the orchestra

example, it has been said above that the activity of playing an instrument might cause in the trumpet player the immediate reddening of his face, and that this reddening can be considered a sterile effect because it does not produce any relevant change in the behaviours of the other components of the orchestra and in the orchestra playing Ravel's Boléro. This means that we might intervene on this sterile effect, for instance by reducing the reddening, without producing any change on the phenomenon-to-be-explained, the orchestra playing Ravel's Boléro.

By considering these options, a first caveat proposed by Baumgartner and Casini (2017, p. 226) is the elaboration of a redundancy-free requirement, that is spelt out in the following way:

> "1) Every cause of S's Ψ-ing is a common cause of S's Ψ-ing and at least one X's Φ_i-ing[20] in the set of the constitutive parts **X's Φ-ing**.
>
> 2) No proper subset **X's Φ-ing** \ { X's Φ_i-ing}is common-cause coupled with S's Ψ-ing—more precisely, for any X's Φ_i-ing in **X's Φ-ing**, there exists at least one cause of S's Ψ-ing, which is not a common cause of S's Ψ-ing and any X's Φ_i-ing in **X's Φ-ing** \ { X's Φ_i-ing}". (Baumgartner & Casini, 2017, p. 226)

Imagine we intervene on the orchestra playing Ravel's Boléro by making the trumpet player stop playing: our intervention causes both a change in the orchestra playing Ravel's Boléro (the orchestral piece will be just partially played), and a change in the trumpet player's behaviour, X's Φ_i-ing (he will not play the trumpet's part). Now, imagine we have a subset made of the behaviours of the constitutive parts except for the behaviour Φ_i-ing of the trumpet player X. This subset would contain the other behaviours exhibited by the trumpet player, the conductor and the other players. The redundancy-free requirement is met if there is not a component of this subset (that is, a different behaviour exhibited by the same player, another player, or the conductor) that is always changed when we intervene on the orchestra by changing the trumpet player. In other words, there is at least one occasion in which our intervention on the orchestra changes the trumpet player's behaviour without changing any other behaviour exhibited by the same component and by the other components of the orchestra. In such a way, it is possible to

[20] The variable Φ_i is taken to represent the specific behaviour Φ_i exhibited by one entity X. We can then find expressions such as X's Φ_i -ing, meaning that the entity X is exhibiting the behaviour Φ_i.

rule out the possibility that we are considering a sterile effect as a constitutive part, because a sterile effect is always changed together with another behaviour of the subset containing all the constitutive parts' behaviours.

Although the redundancy-free requirements are considered a first step towards the identification of constitutive relevant parts, such conditions allow effects of a mere common cause to be considered linked by a constitutive relevance relation. It appears difficult to distinguish between an intervention that simply causes a change in two factors not linked by a constitutive relationship, and an intervention on one constitutive part of the phenomenon and the phenomenon itself. These two situations are easily distinguishable by observing Figure 25. Figure 25(a) represents two interventions, I_Ψ and I_Φ, on the phenomenon S's Ψ-ing and the behaviour Φ_1 of one of its constitutive parts, X; while Figure 25(b) illustrates two interventions, I_Ψ and I_Φ, that cause a change in two factors, S's Ψ-ing and X's Φ_1-ing, whose only links are the common causes I_Ψ and I_Φ.

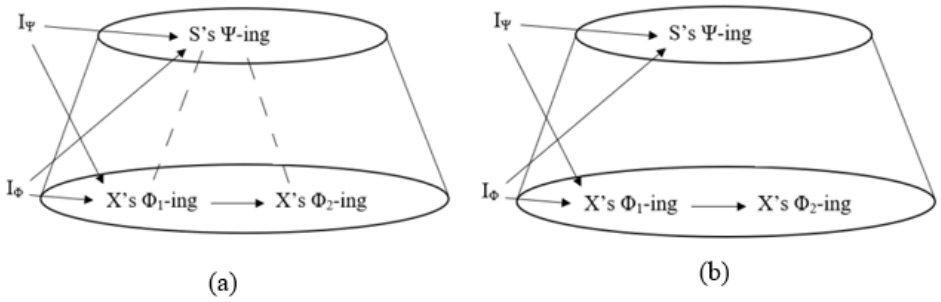

Figure 25. Possible scenarios with the redundancy-free requirements. (a) represents two interventions, I_Ψ and I_Φ, on the phenomenon S's Ψ-ing and the behaviour Φ_1-ing of one of its constitutive parts, X; while (b) illustrates two interventions, I_Ψ and I_Φ, that cause a change in two factors, S's Ψ-ing and X's Φ_1-ing, whose only links are the common causes I_Ψ and I_Φ.

For this reason, Baumgartner and Casini propose what they call the unbreakability condition. Let us consider the variables set containing the possible interventions on S's Ψ-ing. The expansion of this set would lead coincidental common-cause couplings to be broken due to the recognition that is possible to surgically intervene on S's Ψ-ing. As illustrated in Figure 26, while variables linked through a constitutive relevance relation would persist across all variable set expansions, in the case of coincidental common-cause

couplings, such expansions sooner or later would introduce a new intervention $I_{n\Psi}$ that would allow for surgical changes on S's Ψ-ing.

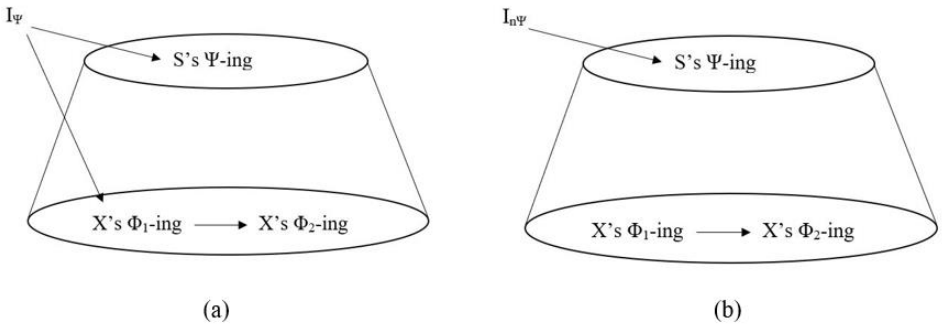

Figure 26. The unbreakability condition. (a) represents the situation in which S's Ψ-ing and X's Φ_1-ing are both changed by the intervention I_Ψ, even if there is not a constitutive relationship between them. (b) illustrates the moment in which the intervention $I_{n\Psi}$, added to extend the set of the interventions on S's Ψ-ing, breaks this coupling by causing a change in S's Ψ-ing and not in X's Φ_1-ing.

By considering all these requirements together, the authors offer their 'abductive theory of constitution' (2017, pp. 228–229). Their theory can be formalised in the following way:

X's Φ-ing is constitutively relevant to S's Ψ-ing if, and only if, there exists a variable set **V** containing S's Ψ-ing and a proper subset of **V**, **X's Φ-ing** made of all the behaviours of the components, such that all of the following conditions are satisfied:

> "1) **Parthood**. For every complex instance of **X's Φ-ing**, there is an instance of S's Ψ-ing such that the former is part of the latter.
>
> 2) **Coupling**.
>
> i) Every cause of S's Ψ-ing in **V** is a common cause of S's Ψ-ing and at least one X's Φ_i-ing in **X's Φ-ing.**
>
> ii) for no X_i's Φ_i-ing in **X's Φ-ing** does **X's Φ-ing** \ { X_i's Φ_i-ing }comply with 2.i).

3) **No De-Coupling**. The Coupling of **X's Φ-ing** and S's Ψ-ing cannot be broken (invalidated) by extending **V** adding new interventions on S's Ψ-ing[21]." (Baumgartner & Casini, 2017, pp. 228–229)

To clarify this formalisation, we can consider again the example of the orchestra playing Ravel's Boléro. In that situation, we can say that the trumpet player's behaviour is constitutively relevant to the orchestra's behaviour if there is a variable set **V** containing the variable representing the orchestra's behaviour, the variables representing the possible interventions on the orchestra's behaviour, and a subset of the constitutive parts' behaviours containing the variables representing all the players' and the conductor's behaviours such that:

1) For every specific behaviour of a constitutive part, there is at least one instance of the phenomenon of the orchestra playing Ravel's Boléro such that the constitutive part exhibiting the behaviour is part of the orchestra playing Ravel's Boléro.

2i) Every intervention on the orchestra playing Ravel's Boléro causes both a change in the orchestra's behaviour and a change in at least one trumpet player's behaviour belonging to the subset of the constitutive parts' behaviours;

2ii) If we consider the subset containing all the behaviours of all the parts except for that specific trumpet player's behaviour, there is not a component's behaviour belonging to this subset that is always changed when an intervention changes both the orchestra playing Ravel's Boléro and that trumpet player's behaviours.

3) Even if we extend the variables in **V** by considering further interventions on the orchestra playing Ravel's Boléro, the coupling between the trumpet player's behaviour and the orchestra's behaviour cannot be broken: it will still

[21] In the formulation of the No De-Coupling (NDC) requirement for constitutive relevance, Baumgartner and Casini (2017, pp. 228–229) do not define the variable set **V**, but claim that the No De-Coupling condition is met if the coupling of X's Φ-ing and S's Ψ-ing cannot be broken (invalidated) by extending **V**, that means, by adding new components X's Φ-ing. Given that the addition of new constituents would not lead to the finding of surgical interventions on S's Ψ-ing, I have taken **V** to contain also the possible interventions on S's Ψ-ing. My decision is supported by the fact that, before presenting their formalised NDC condition, Baumgartner and Casini discuss a figure in which the variable set **V** contains S's Ψ-ing, the components X's Φ-ing, and the possible interventions on S's Ψ-ing (Baumgartner & Casini, 2017, p. 227).

be impossible to surgically intervene on the phenomenon of the orchestra playing Ravel's Boléro in the same way (with respect to the trumpet player's behaviour) without changing the trumpet player's behaviour.

In order to find constitutive relevance relationships, the abductive theory of constitution requires testing for unbreakable common-cause coupling. This requirement is very challenging: the ways of intervening on a phenomenon S's Ψ-ing and on the behaviours of the potentially constitutive relevant parts of the constitutive set **X's Φ-ing** must be systematically changed to expand **V** and check whether, outside **V**, there exist surgical causes that break the common-cause coupling of **X's Φ-ing** and S's Ψ-ing.

Baumgartner and Casini consider that, of course, it is not feasible to conduct all the possible interventions on S's Ψ-ing because this would require an infinite expansion of the variable set. As a consequence, they claim that, after many attempts, the constitutive relevance relationship can only be abductively inferred, after having inductively inferred the satisfaction of the Coupling and No De-Coupling requirements.

Although the main limitation of the abductive theory of constitution has been claimed to be the unfeasibility of such requirements and the necessary use of induction and abduction, in the next section I will focus on another limitation, the fact that, in particular situations, the "fat-handed" nature of interventions on constitutive mechanisms makes really difficult to address cases of multiple realizations.

7.3.1.1 Multiple realizations and bottom-up intervention

The main reason that has led Baumgartner and Casini to propose their account is that, since every intervention on a constitutive mechanism is fat-handed, it is impossible to distinguish between bottom-up and top-down interventions. In the case of constitutive mechanisms, indeed, such a distinction is far from being clear.

Let us go back to the example of the orchestra playing Ravel's Boléro: changing the mechanism's behaviour S's Ψ-ing means changing the behaviour of the orchestra, that is the activity of playing Ravel's Boléro. To change that, the only possibility is to change the behaviour of at least one of the components of the orchestra, for instance by stopping the activity of the trumpet player X_1. In such a way, we are changing both the mechanism's behaviour S's Ψ-ing and the behaviour Φ_1 exhibited by its component X_1. Now, let us assume that in the orchestra there is a trumpet player X_2 ready to start

exhibiting the behaviour Φ_1 if and only if the first trumpet player X_1 is unable to play (that means, X_1 is unable to perform the behaviour Φ_1). This situation is a case of multiple realization: the change of one constitutive part does not lead to a change of the whole phenomenon because other parts can compensate that change. In other words, even a relevant change in the behaviour Φ_1 of X_1 would not affect S's Ψ-ing because another part X_2 exhibiting the behaviour Φ_1 would take the place of the relevant part we have changed.

If the trumpet player X_2 immediately takes the place of the trumpet player X_1 when we make X_1 stop playing, we will have an ambiguous situation in which, even though we wanted to change S's Ψ-ing (i.e. the behaviour or the causal capacity of the orchestra), we just changed X_1's Φ_1-ing. In other words, apparently, we performed a bottom-up intervention rather than a fat-handed intervention. When there is X_2 ready to start exhibiting the behaviour Φ_1 if X_1 is stopped, it is actually impossible to perform fat-handed interventions on S's Ψ-ing and X_1's Φ_1-ing, even though X_1's Φ_1-ing is constitutive of S's Ψ-ing.

It is important here to recall what we have said in section 7.3.2: we cannot intervene on S's Ψ-ing without changing also at least the behaviour of one constitutive part, X's Φ-ing, but there is the possibility that we intervene on X's Φ-ing without changing S's Ψ-ing. Indeed, surgical interventions on X's Φ-ing are possible because we could make the constitutive entity X change its behaviour Φ in a way that does not affect S's Ψ-ing, or we could change a behaviour that is not actually constitutive of the phenomenon under study (X's Φ-ing might indeed be a sterile effect always coupled with another behaviour of the same or of a different part).

According to Baumgartner and Casini's account, in a case of multiple realization we would arrive at the following conclusions:

1) Since the top-down intervention on S's Ψ-ing changes only X_1's Φ_1-ing, while S's Ψ-ing remains unchanged, the intervention fails the Coupling condition and should be interpreted as a bottom-up intervention on X_1's Φ_1-ing. The consequence is that X_1's Φ_1-ing should not be considered a constituent of S's Ψ-ing.

2) Since the behaviour Φ_1 of the trumpet player X_2 would only be caused by interventions on Ψ that change also X_1's Φ_1-ing, X_2's Φ_1-ing does not satisfy the redundancy-free requirements and should be interpreted as a sterile effect.

In cases of multiple realizations, hence, Baumgartner and Casini's account does not allow us to claim that either X_1's Φ_1-ing before the intervention or X_2's Φ_1-ing after the intervention can be considered constitutive of S's Ψ-ing.

7.3.2 Krickel's account: the causation-based constitutive relevance

Recently, Krickel has developed a new solution to the problems related to constitutive relevance. Her starting point is the assumption that constitutive mechanisms entail both entities and activities. She then considers that the use of just one variable to represent the phenomenon is not consistent with the fact that phenomena often consist of changes over time. For this reason, she proposes to distinguish between *spatial entity-involving occurrents* and *temporal entity-involving occurrents*, where both kinds of parts are themselves acting entities. Temporal entity-involving occurrents and spatial entity-involving are defined as follows:

> "An acting entity E1 is a **temporal entity-involving occurrent** of another acting entity E2 iff:
>
> 1) the entity involved in E1 is identical with the entity involved in E2;
>
> 2) the activity involved in E1 begins later and ends earlier than the activity involved in E2, or the former begins simultaneously with the latter and ends earlier than the latter, or the former begins later than the latter and ends simultaneously with the latter." (Krickel, 2018, p. 64)
>
> "An acting entity E1 is a **spatial entity-involving occurrent** of another acting entity E2 iff:
>
> 1) the entity involved in E1 occupies a proper sub-region of the spatiotemporal region occupied by the entity involved in E2;
>
> 2) the activity involved in E1 occurs during the activity involved in E2." (Krickel, 2018, p. 64)

To give an example, Krickel considers the phenomenon consisting of the mouse navigating the Morris water maze: a mouse put in the water maze has to escape from

water by successfully navigating and thereby locating a hidden platform. If we have to recognise spatial entity-involving occurrents, we can consider the mouse's muscles moving, or the hippocampus generating spatial maps while the mouse is navigating. They are spatial entity-involving occurrents constituting the acting-entity under study (the mouse navigating the Morris water maze). Furthermore, the mouse being put into the pool in a particular area of the water maze, the mouse swimming for a while in one direction in another area of the water maze, swimming in another direction in another area of the water maze, and finally finding the platform, can be described as temporal entity-involving occurrents of the phenomenon.

Figure 27 illustrates this distinction: one the one hand, a phenomenon is constituted by spatial entity-involving occurrents that occupy sub-regions of the acting entity, like a); on the other hand, a phenomenon is also constituted by temporal entity-involving occurrents that involve the same acting entity of the phenomenon itself, but that occupy limited time intervals, like b).

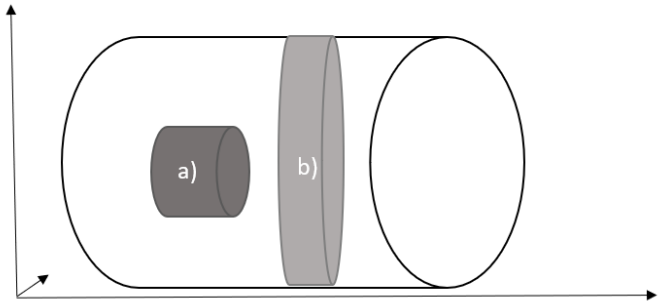

Figure 27. Spatial entity-involving occurrents and temporal entity-involving occurrents. a) represents a spatial entity-involving occurrent that occupies a sub-region of the acting entity; b) represents a temporal entity-involving occurrent that involves the same acting entity of the phenomenon but that occupies a limited time interval.

The idea proposed by Krickel is that spatial entity-involving occurrents and temporal entity-involving occurrents of the same acting entity, if occurring in different spatio-temporal regions, can causally interact. The justification for their causal interactions is that entity-involving occurrents that occur in different spatio-temporal regions can be wholly distinct events.

On the basis of this observation, Krickel has developed a new account of constitutive relevance, called causation-based constitutive relevance, based on the notion of ideal* interventions and defined as follows:

> "X's Φ-ing is *constitutively relevant* for S's Ψ-ing iff:
>
> 1) X's Φ-ing is a spatial entity-involving occurrent of S's Ψ-ing,
>
> 2) there is a temporal entity-involving occurrent of S's Ψ-ing that is a cause of X's Φ-ing, and
>
> 3) there is a temporal entity-involving occurrent of S's Ψ-ing that is an effect of X's Φ-ing." (Krickel, 2018, p. 64)

To formulate this causation-based constitutive relevance in terms of interventionism, Krickel needs to represent the different acting entities by variables: she represents X's Φ-ing by the variable Φi, and the temporal entity-involving occurrents of S's Ψ-ing by Ψ1 and Ψ2. If Woodward's interventionist account is then applied to her notion, it follows that condition 2) is satisfied if and only if there is a variable Ψ1=ψ1 for which it is true that there is an ideal or ideal* intervention on Ψ1 with respect to Φi=φi that changes Φi while all other variables not on the causal path between Ψ1 and Φi are kept fixed except for variables that Ψ1 and Φi non-causally depend on. Condition 3) is satisfied if and only if there is a variable Ψ2=ψ2 for which it is true that there is an ideal or ideal* intervention on Φi=φi that changes Ψ2 while all other variables not on the causal path between Φi and Ψ2 are kept fixed except for variables that Ψ2 and Φi non-causally depend on.

Let us consider again the mouse navigating the Morris water maze for a period of time going from t_1 to t_4. Krickel focuses on the question of whether the hippocampus's activity at t_3 is constitutively relevant to that phenomenon (2018, p. 65). To answer that question, it should be verified at first that there is a temporal entity-involving occurrent of the mouse's navigation behaviour for which it is true that, had there been an ideal* intervention on that temporal entity-involving occurrent with respect to the hippocampus's activity at t3, then the hippocampus's activity at t3 would have been different. Second, it should be verified the existence of a temporal entity-involving occurrent of the mouse's navigation behaviour for which it is true that, had there been an ideal* intervention on the hippocampus's activity with respect to that temporal part, the temporal entity-involving occurrent would have been different.

According to Krickel, condition 2) might be satisfied if one could imagine that had there been an ideal* intervention on the entering of the mouse into the maze at t_1 (e.g. a change in the location of where the mouse was put into the maze), then the hippocampus's activity would have been different at t_3 (e.g. different neural representations in the hippocampus would have been active). In addition, condition 3) could be satisfied if there was an ideal* intervention on the hippocampus's activity at t_3 so that the mouse's finding the platform at t_4 would have been different (e.g. it would have found the platform later). While Krickel claims that her proposal can save the mutual manipulability approach, in the next pages I will discuss four limitations of her account.

7.3.2.1 Causation between entity-involving occurrents requires parts occupying different temporal regions

Let us consider some possible scenarios based on Krickel's account. Intuitively, the general case described by Krickel is the situation where we ideally or ideally* intervene on the temporal entity-involving occurrent of S's Ψ-ing at t_1, this causes a change in the spatial entity-involving occurrent X's Φ-ing at t_2, which consequently leads to a change in the temporal entity-involving occurrent of S's Ψ-ing at t_3, like in Figure 28. In this situation, we would not have problems because spatial entity-involving occurrents and temporal entity-involving occurrents of the same acting entity, if occurring in different spatio-temporal regions, are wholly distinct events and can causally interact.

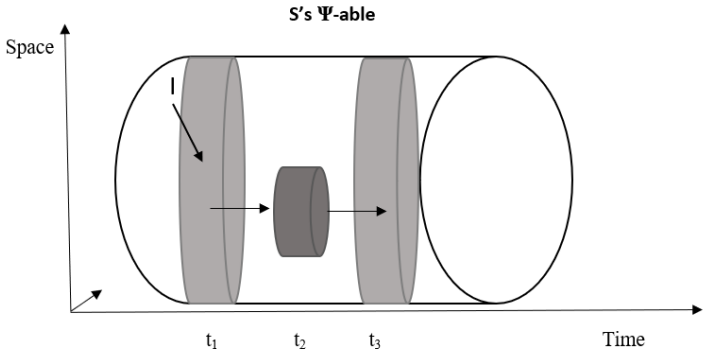

Figure 28. The general case described by Krickel. A temporal entity-involving occurrent of S's Ψ-ing at t_1 is a cause of X's Φ-ing at t_2, and there is a temporal entity-involving occurrent of S's Ψ-ing at t_3 that it is an effect of X's Φ-ing at t_2.

There are further scenarios, however, that pose a problem to Krickel's account. Although in her paper Krickel clarifies in different occasions that, in order to be causally dependent, entity-involving occurrents must occur in different spatio-temporal regions, her definition of causation-based constitutive relevance does not rule out that a temporal region can be occupied by both spatial and temporal entity-involving occurrents. More specifically, the definition of spatial entity-involving occurrents entails that such parts occupy just sub-regions of the spatiotemporal region, while the definition of temporal entity-involving occurrents entails that they occupy the whole spatial region and occur at a sub-region of the temporal region. Consequently, while a spatial entity-involving occurrent and a temporal entity-involving occurrent will never occupy the same spatial region, it might happen that they occupy the same temporal region.

This observation causes a counterexample if applied to the conditions for constitutive relevance. On the one hand, X's Φ-ing is claimed to be constitutively relevant to S's Ψ-ing iff: (i) X's Φ-ing is a spatial entity-involving occurrent of S's Ψ-ing, (ii) there is a temporal entity-involving occurrent of S's Ψ-ing that is a cause of X's Φ-ing, and (iii) there is a temporal entity-involving occurrent of S's Ψ-ing that is an effect of X's Φ-ing. On the other hand, the temporal regions of such parts are not specified, with the consequence that it might happen that the same temporal region is shared by the spatial entity-involving occurrent X's Φ-ing and its putative cause, the temporal entity-involving occurrent of S's Ψ-ing (as in Figure 29); or by the spatial entity-involving occurrent X's Φ-ing and its putative effect, the temporal entity-involving occurrent of S's Ψ-ing (as illustrated in Figure 30).

As described above, Krickel uses the notion of ideal or ideal* interventions to establish causation: considering Figure 29, for instance, we can establish that S's Ψ-ing at t_1 is a cause of X's Φ-ing at t_1 iff there is an ideal or ideal* intervention on S's Ψ-ing at t_1 with respect to X's Φ-ing at t_1 that changes the value of X's Φ-ing at t_1 given that all other causes of X's Φ-ing at t_1 that are not on the causal path between I, S's Ψ-ing at t_1, and X's Φ-ing at t_1 are kept fixed at their actual values (except for variables that I, S's Ψ-ing at t_1, and X's Φ-ing at t_1 non-causally depend on).

Since, as argued in section 7.2, ideal and ideal* interventions allow for causal relationships between one entity and its constitutive parts, the ideal or ideal* intervention on S's Ψ-ing at t_1 with respect to X's Φ-ing at t_1 would not be problematic in

interventionist terms. My argument is that such a conclusion, however, would be in contrast with the idea that causation cannot be established if the spatial entity-involving occurrent X's Φ-ing and the temporal entity-involving occurrent of S's Ψ-ing share the same temporal region because in that situation they are not wholly distinct events.

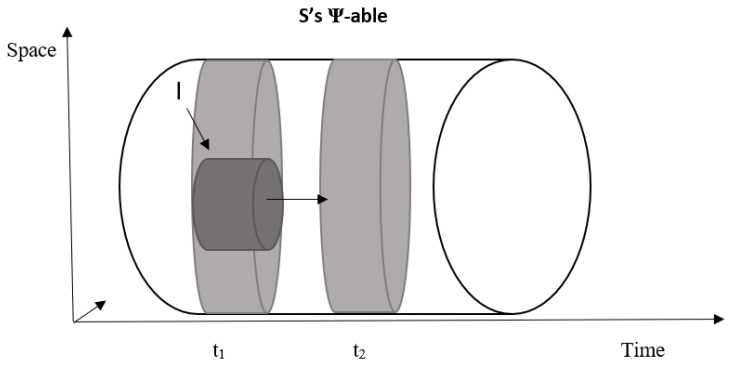

Figure 29. A temporal entity-involving occurrent of S's Ψ-ing at t_1 is the cause of X's Φ-ing at t_1.

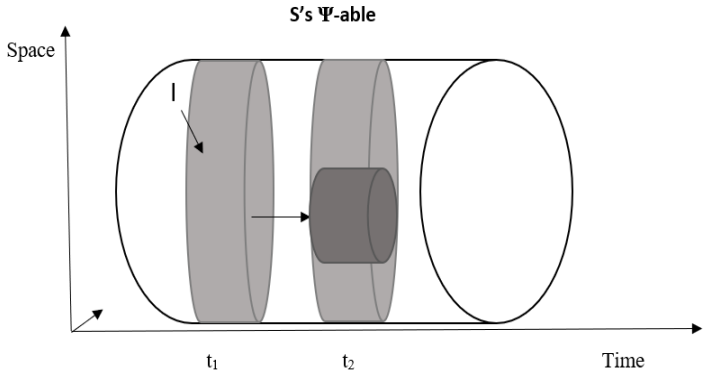

Figure 30. A temporal entity-involving occurrent of S's Ψ-ing at t_2 is the effect of X's Φ-ing at t_2.

7.3.2.2 Do we need temporal entity-involving occurrents?

The observations proposed in section 7.3.2.1 can be used to identify another problem of Krickel's account. Since a spatial entity-involving occurrent and a temporal entity-involving occurrent can share the same temporal region, it is likely that each temporal entity-involving occurrent (that is characterised by the same entity of the whole mechanism) is constituted by spatial entity-involving occurrents. Figure 31 helps to clarify this situation: the temporal entity-involving occurrents at t_1, t_2, and t_3 are all constituted by spatial entity-involving occurrents sharing the same temporal regions.

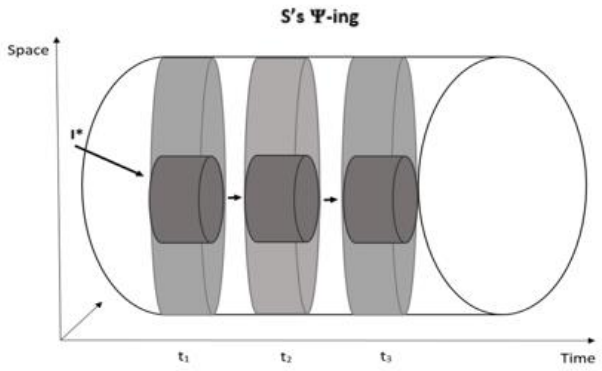

Figure 31. The temporal entity-involving occurrents at t_1, t_2, and t_3 are all constituted by spatial entity-involving occurrents sharing the same temporal regions.

If we intervene with an ideal* intervention on the temporal entity-involving occurrent at t_1 with respect to the spatial entity-involving occurrent at t_2, our intervention will always change both the temporal entity-involving occurrent at t_1 and at least one of its constitutive spatial entity-involving occurrents at t_1. Similarly, if we intervene on the spatial entity-involving occurrent at t_2 with respect to the temporal entity-involving occurrents at t_3, our intervention will always change also one spatial entity-involving occurrent at t_3. The problem is that, with ideal* intervention, it is impossible to establish whether we are intervening on the temporal entity-involving occurrent at t_1 with respect to the spatial entity-involving occurrent at t_2, or whether we are intervening on the spatial entity-involving occurrent at t_1 with respect to the spatial entity-involving occurrent at t_2. In other words, due to the fat-handed nature of an ideal* intervention, we do not know where to locate it. It is hence questionable whether Krickel's account needs temporal entity-involving occurrents, or whether we could spell out her proposal just in terms of spatial entity-involving occurrents in different time regions of the phenomenon-to-be-explained.

7.3.2.3 Multiple realizations and the absence of ideal interventions

As described in section 7.3.1.1, the account proposed by Baumgartner and Casini might cause problems when constitutive mechanisms are characterised by multiple realizations. A similar problem can be found also in Krickel's account of causation-based constitutive relevance. If we applied Krickel's proposal, it would follow that those parts characterised by multiple realizations should not be considered as constitutive parts. The reason is that, in order to establish constitutive relevance, it should be verified that there is an ideal or ideal* intervention on X's Φ-ing at t_n that changes S's Ψ-ing at t_{n+1} while all other

variables not on the causal path between X's Φ-ing at t_n and S's Ψ-ing at t_{n+1} are kept fixed except for variables that X's Φ-ing at t_n and S's Ψ-ing at t_{n+1} non-causally depend on. An intervention on X's Φ-ing at t_n that causes another spatial entity-involving occurrent Y's Φ-ing to replace X's Φ-ing, as happens in the case of multiple realizations, would maintain S's Ψ-ing at t_{n+1} unchanged, with the consequence that it would not be considered an ideal or ideal* intervention. The direct consequence would be that a possible causal relationship between X's Φ-ing at t_n and S's Ψ-ing at t_{n+1} would be excluded. Furthermore, it would be established that X's Φ-ing does not meet the requirement for constitutive relevance.

7.3.2.4 Constitutive mechanisms without activities

In section 7.3.2 I have observed that Krickel's starting point is the assumption that constitutive mechanisms entail both entities and activities. According to this position, mechanisms merely characterised by a disposition cannot be described as mechanisms. There are different counterexamples, however, that might lead to the conclusion that there are systems characterised only by causal capacities that should be considered mechanisms.

A clear case in point that illustrates a situation in which we have a causal disposition is an unexploded bomb. The bomb is not exhibiting any behaviour, therefore the phenomenon Ψ (that is, the bomb exploding) is just dispositional. Additionally, the bomb's components are not engaged in any activity. Overall, however, we can claim that the components' behaviours are potential, and that this is enough to claim that the bomb is a mechanism, given that it has the causal capacity to explode.

A similar example is the case of an unwound watch S. It could be claimed that the watch is a mechanism because it has the dispositional causal capacity Ψ to indicate the right hour, and that its components X_1, X_2, X_3, X_4 (for instance the gears and the hands) are constitutive parts that potentially can exhibit the constitutive behaviours Φ_1, Φ_2, Φ_3 and Φ_4.

Another example from the social sciences is the componential system whereby the central bank can influence the money supply through the potential activities of some financial entities (for instance, certain entities can buy or sell government securities; influence interest rates; or change the reserve requirement). This complex system of entities and

potential activities can be considered a mechanism even if the central bank is not influencing in any way, in a specific moment, the money supply.

Similarly, we could claim that the orchestra S ready to start playing Ravel's Boléro is a mechanism because it has the disposition to do Ψ (that is, to play) and that, even though its components (the players and the conductor) are not acting, potentially they can exhibit the activities required to constitute the phenomenon Ψ.

Overall, all these examples suggest that constitutive mechanisms do not necessarily require actualised behaviours to be defined as such. If a constitutive mechanism is not engaged in any activity, however, it is difficult to distinguish between different time regions. Consequently, it might be argued that Krickel's proposal is not adequate to include constitutive mechanisms characterised by dispositional behaviours.

7.3.3 Baumgartner, Casini and Krickel's latest proposal: horizontal surgicality

Very recently, Baumgartner, Casini and Krickel have developed together another proposal that combines some of their previous ideas and tackles some of the problems I have discussed above. Starting from Baumgartner and Casini's account, the authors recognise that only when an intervention causes *simultaneous changes* both in the phenomenon-to-be-explained and in one of its parts, such changes can be considered as revealing of constitution[22].

In the horizontal surgicality approach proposed by Baumgartner, Casini and Krickel, this observation is associated with the assumption that it is possible to find constitutive parts of a phenomenon on every possible lower level without gaps between levels. By considering the example of the orchestra playing Ravel's Boléro, such an assumption would require that an intervention on the orchestra changes not only the behaviour of one of its players, but also one of the constitutive parts of the player (such as his respiratory system), and other constitutive parts in lower levels (such as the pulmonary circulation).

Together, such ideas are used to develop the principle of universal constitution (Baumgartner et al., 2018, p. 5), that can be spelt out in this way:

[22] Simultaneously changes in parts and phenomena are claimed by the authors to be non-causal because causes always temporally precede their effects.

> **Universal constitution**. Every (change in a) state of any (non-fundamental) phenomenon S's Ψ-ing is necessarily and simultaneously realised by (a change in) the state of at least one constituent X's Φ-ing of S's Ψ-ing on every lower level, such that S's Ψ-ing ≠ X's Φ-ing.

The requirement of simultaneous changes rules out the possibility of recognising sterile effects as constitutive parts since producing such effects would require time. Furthermore, producing simultaneous changes in the phenomenon-to-be-explained and one of its parts through an intervention would entail either that the intervention can change both the phenomenon and the part along a single causal path (I→ S's Ψ-ing → X's Φ-ing), or that the intervention changes them through two different causal pathways, and it is consequently a common cause of the phenomenon-to-be-explained and one of its parts (S's Ψ-ing ← I → X's Φ-ing). Since the possibility of a single causal path is ruled out by the fact that there cannot be causal relationships between a phenomenon and its constitutive parts, the only possible explanation of simultaneous changes is that the responsible interventions are fat-handed (Baumgartner et al., 2018, p. 5).

Baumgartner, Casini and Krickel add a further requirement for constitutive relevance: every fat-handed intervention is a direct cause of at most one behaviour on every level lower than S's Ψ-ing. In other words, it is impossible that an intervention on S's Ψ-ing changes simultaneously both X_1's Φ_1-ing and X_2's Φ_2-ing, or X_1's Φ_1-ing and X_1's Φ_2-ing that can be found at the same level (for this reason, interventions are called *surgical*). This requirement rules out the possibility that the intervention is just a common cause changing the phenomenon-to-be-explained and an acting part that is not constitutively relevant. Indeed, every time we intervene on a constitutive mechanism we necessarily change both the phenomenon and the behaviour of a constitutive acting part of the mechanism. Therefore, if the intervention does not change only the acting part under study X_1's Φ_1-ing, but also X_1's Φ_2-ing or another acting entity at the same level, X_2, we can establish that X_1's Φ_1-ing is not constitutive of the phenomenon.

Together, such requirements form the 'horizontal surgicality account', according to which I is a horizontally surgical intervention variable on a part X's Φ-ing with respect to S's Ψ-ing iff:

1) I is a cause of X's Φ-ing;

2) If I causes changes in both X's Φ-ing and S's Ψ-ing, these changes occur simultaneously;
3) I is a direct cause of at most one behaviour on every level lower than S's Ψ-ing.

The authors, finally, claim that X's Φ-ing is a constitutive part of S's Ψ-ing iff[23]:

1) The instances of X's Φ-ing are spatiotemporal parts of instances of S's Ψ-ing;
2) There exists a (possible) horizontally surgical intervention I on X's Φ-ing with respect to S's Ψ-ing that causes changes in both X's Φ-ing and S's Ψ-ing.

7.3.3.1 A solution to the problem of multiple realizations

Even if the authors did not discuss this aspect, the recognition that constitutive relations, unlike causal relations, do not require time, is the key point that helps us to solve the problem of multiple realizations. Let us consider again the example of the orchestra playing Ravel's Boléro. If during the concert we change the trumpet player's behaviour by tying his arms, we will change simultaneously (even if just for a very short time) the behaviour of the orchestra. After a very short moment, the substitute will start playing to replace the behaviour of the tied trumpet player, and the orchestra's behaviour will return to the condition it had before the intervention. In such a situation, the causal relationship between the behaviour of the first trumpet player and the substitute will take time.

Figure 32 helps to clarify this point: if we use Krickel's style of representing constitutive mechanisms, it is possible to distinguish between different temporal intervals (t_1 and t_2) and between different spatiotemporal parts of the phenomenon. We can identify the spatiotemporal part X_1 characterised by the behaviour Φ_1, and the spatiotemporal part X_2 characterised by the behaviour Φ_1 that could replace X_1's Φ_1-ing. Let us imagine we intervene on X_1's Φ_1-ing at t_1 with respect to S's Ψ-ing, immediately X_1 and S will change their behaviours Φ_1 and Ψ. After that, at t_2, X_2 will change (activate) its behaviour Φ_1 to replace the (identical) behaviour of X_1 at t_1.

What follows is that, if we require that the intervention I causes simultaneous changes in the phenomenon-to-be-explained and the constitutive part, we will be able to identify constitutive parts even in cases of multiple realizations.

[23] Baumgartner, Casini and Krickel formalised their proposal differently, for further detail see Baumgartner, Casini and Krickel, 2018, p. 9-12.

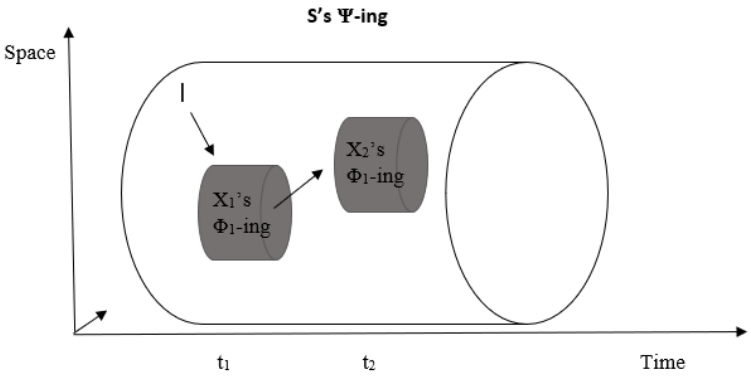

Figure 32. A case of multiple realization. If we consider the mechanism S, we can distinguish between the spatiotemporal part X_1 that simultaneously changes behaviour due to the intervention I at t_1, and the spatiotemporal part X_2 whose activity Φ_1 is caused by X_1's Φ_1-ing at t_2.

7.3.3.2 Two limitations: parthood and dispositional phenomena

There are two conceptual limitations in the account proposed by Baumgartner, Casini and Krickel.

The first limitation that can be found is that, like in their previous proposals, the horizontal surgicality account requires that the constitutively relevant entities of the mechanism are *physical parts* of the mechanism. There are at least two scenarios, however, in which this requirement cannot be met even if the entities under study are constitutive of the phenomenon-to-be-explained.

In the first scenario there are clear physical entities that can be considered parts of the mechanism exhibiting the phenomenon-to-be-explained, but some constitutive entities are external to the mechanism. An example is the action potential mechanism exhibiting the phenomenon of muscle contraction, that comprises also an external neuronal signal (see Kaiser & Krickel, 2016). This situation is not new in the philosophical debate: Craver himself recognised that mechanisms "frequently transgress compartmental [physical] boundaries" (Craver, 2007, p. 141). In such cases, in order to claim that these external entities are parts of the mechanism, we have to modify the ways in which the physical or spatial boundaries of the constitutive mechanism are drawn.

The second scenario is more complex and involves a phenomenon exhibited by a mechanism that is not a physical object. Above it has been suggested that clear examples of social constitutive mechanisms can be found in the discussions on social organisations,

in general defined as the articulations "of different parts which perform various functions" (Ogburn & Nimkoff, 1964, p. 9). Such articulations are not necessarily physical objects: in several situations social organisations are abstract concepts that can be understood only by determining which entities should be considered as components.

An example often used to clarify this observation is the social organisation that we call 'family'. Although there are some rules to establish whether someone is a family member (we can examine whether the persons are linked by blood, marriage or adoption), the idea about what individuals constitute a family varies across the world. There are societies in which the concept of family is nuclear: only parents and children are members of a family. Other societies extend the boundaries and consider also other relatives as members of the family. Since a family is not a physical object, the boundaries are invisible and contextual: if a person is asked to mention the individuals who are parts of her family, the answer will depend on the concept of family she has. The direct consequence of this observation is that, in some cases, researchers are required to establish by definition (analytically) the boundaries of the mechanism, and their decisions are likely to be influenced by cultural and subjective considerations (this point will be discussed in detail in section 7.6). In such situations, it is impossible to establish definitively whether an entity is or is not a mechanism's part.

Overall, hence, the parthood condition can cause conceptual and methodological problems when used for constitutive relevance because:

1) When the mechanism at the macro level is a physical mechanism, it could happen that external entities, not belonging to that physical mechanism, exhibit activities constitutive of the phenomenon-to-be-explained, as for the action potential mechanism.
2) In the social sciences it is likely to find situations in which the mechanism exhibiting the phenomenon at the macro level is not a physical object with well-defined boundaries, but an intangible mechanism whose boundaries are determined analytically (by definition), like in the case of the family.

There is a further limitation of the horizontal surgicality account: like in the previous proposals, this account does not allow for constitutive mechanisms with a dispositional behaviour (a detailed discussion has been provided in section 7.3.2.4).

7.4 An extended horizontal surgicality account

In this section, I shall propose some changes to the horizontal surgicality account both to tackle the conceptual limitations described above and to offer a philosophical proposal with as few formalisations as possible. To begin with, the parthood problem can be easily addressed by removing the first requirement for constitutive relevance, namely that the instances of X's Φ-ing are spatiotemporal parts of instances of S's Ψ-ing, and by talking about constitutive relevant *entities*.

Furthermore, to make explicit that we are considering both actualised and dispositional phenomena, it is possible to modify the formalisation proposed by Baumgartner, Casini and Krickel in their various account by using 'X's Φ-able' to refer to the entity X characterised by either an actualised or dispositional causal capacity Φ, and 'S's Ψ-able' to symbolise the potential or actualised phenomenon Ψ characterising S.

When the phenomena-to-be-explained are dispositional, a new problem arises. Let us imagine that we have the orchestra ready to play Ravel's Boléro, with the trumpet player and his substitute. In a first scenario, illustrated in Figure 33(a), we intervene on the trumpet player X_1's Φ_1-able by tying his arms, and we do not realise that there is a second trumpet player X_2's Φ_1-able ready to transform his potential behaviour Φ_1 into an actual behaviour when needed. In such a situation, we might conclude that our intervention does change simultaneously both the *potential* behaviour of the trumpet player X_1's Φ_1-able and the *potential* behaviour of the orchestra S's Ψ-able (in other words, the trumpet player and the orchestra would lose their casual capacity). In a second scenario, described in Figure 33(b) we are aware of the presence of the second trumpet player X_2's Φ_1-able. We intervene on the potential behaviour of the first trumpet player X_1's Φ_1-able but we realise that this intervention does not change simultaneously also the causal capacity of the orchestra because the substitute X_2's Φ_1-able can transform his potential behaviour Φ_1 into an actual behaviour when needed (when the orchestra starts performing Ravel's Boléro). In this second case, the intervention looks like a bottom-up intervention.

In the former scenario, we would arrive at the right conclusion (X_1's Φ_1-able is constitutive of S's Ψ-able) ignoring the presence of X_2's Φ_1-able. In the second scenario, we would conclude that X_1's Φ_1-able is not constitutive of S's Ψ-able. In both cases, however, the situation would be very different if we intervened once the behaviour Ψ was actualised. Indeed, regardless of whether we are aware of the presence of X_2's Φ_1-able or

not, an intervention on the acting mechanism S's Ψ-able would change simultaneously both the actualised X_1's Φ_1-able and S's Ψ-able, as illustrated in Figure 32 above, while the 'activation' of X_2's Φ_1-able would take time. In other words, when the phenomenon-to-be-explained is dispositional, the problem of multiple realizations can put our conclusion about the constitutive relevance of an entity on the wrong track. When the phenomenon-to-be-explained is actualised, instead, we do not have to worry about the problem of multiple realizations.

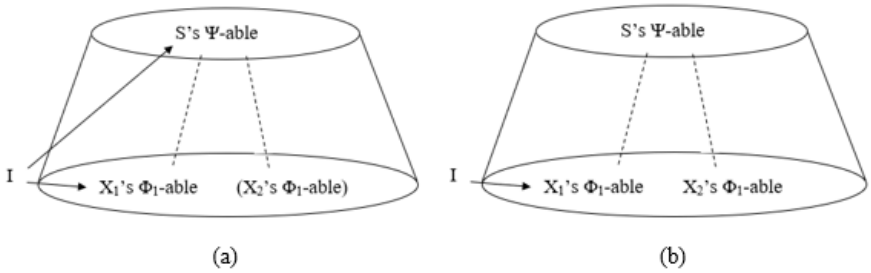

Figure 33. An extended horizontal surgicality account. In (a) the S is just a dispositional behaviour and we do not see X_2's Φ_1-able. When we intervene on X_1's Φ_1-able, consequently, we think that we change simultaneously also the potential behaviour Ψ of the orchestra S. In (b), we know that there is X_2's Φ_1-able, consequently we realise that our intervention on X_1's Φ_1-able is just a bottom-up intervention because the potential behaviour Ψ of the orchestra S is not changed.

Consequently, this issue can be solved by changing the second requirement in the following way: *if the behaviour Ψ of S were actualised, there would exist a (possible) horizontally surgical intervention I on X's Φ-able with respect to S's Ψ-able that would cause changes in both X's Φ-able and S's Ψ-able*.

This solution allows researchers to clearly determine if an entity is constitutive of a mechanism when the mechanism is *exhibiting* the phenomenon-to-be-explained, while it can be used as a thought experiment with mechanisms characterised by dispositional behaviours. Overall, the changes I have proposed so far lead to the following extended horizontal surgicality account:

I is a horizontally surgical intervention variable on X's Φ-able with respect to S's Ψ-able iff:

1) I is a cause of X's Φ-able;

2) If I causes changes in both X's Φ-able and S's Ψ-able, these changes occur simultaneously;
3) I is a direct cause of at most one behaviour on every level lower than S's Ψ-able.

X's Φ-able is constitutive of S's Ψ-able iff:

1) X is an entity with the dispositional or actualised behaviour Φ;
2) If the behaviour Ψ of S were actualised, there would exist a (possible) horizontally surgical intervention I on X's Φ-able with respect to S's Ψ-able that would cause changes in both X's Φ-able and S's Ψ-able.

7.5 Big data and constitutive mechanistic evidence

Now that I have established how to identify constitutive relevant entities, it is time to explore whether big data can be used to find them and to collect constitutive mechanistic evidence. From the big data's description provided in the introduction of the thesis, it emerges that the time interval between a data collection and another, thanks to new sophisticated devices, can often be reduced to a few milliseconds; and that it is now possible to collect information regarding a huge amount of entities. It follows that big data might help researchers to monitor every small change both in the phenomenon-to-be-explained and in the behaviours of all the possible constitutive entities. Such capacities are crucial in order to verify if changes in the phenomenon and in the constitutive entity happen simultaneously and if there are changes in further behaviours at the same level as the constitutive entity we are studying.

The use of big data to find constitutive relevant entities through the extended horizontal surgicality account might provide useful mechanistic evidence especially when the phenomenon-to-be-explained is actualised. Let us consider again the example of the orchestra playing Ravel's Boléro. Through big data, researchers can measure all the behaviours of the putative components belonging to a specific level. For instance, researchers can monitor all the behaviours of the components and of the conductor of the orchestra. If the behaviour of the orchestra is actualised, furthermore, big data can help to detect changes in how Ravel's Boléro is played. In such a way, it would be possible to establish if an intervention on the trumpet player simultaneously changes also the behaviour exhibited by the orchestra while all the other behaviours are maintained unchanged.

The example of the orchestra used so far, however, runs the risk of being too simplistic. Social constitutive mechanisms are often made of numerous entities and are much more complex than an orchestra with just one trumpet player. Let us consider a real traditional symphonic orchestra: this in general is not made of one, but of eight trumpet players. In a traditional orchestra playing Ravel's Boléro, hence, if we made one of the trumpet players stop playing, there would be seven other trumpet players still playing their parts. In this situation, should we say that our intervention on the behaviour Φ_1 of the trumpet player X_1 changes the phenomenon Ψ under examination, namely the orchestra playing Ravel's Boléro?

Let us assume the other members of the orchestra do not notice what happens to the trumpet player and keep playing. Due to the presence of the other trumpet players, the change in the phenomenon might be imperceptible to the human ear. However, big data could help us to monitor every micro-change in the phenomenon-to-be-explained. In the case of an orchestra with eight trumpet players, for instance, we might collect sound intensity data and, during the intervention on the trumpet player, observe a reduction in the decibel level. It might be argued, hence, that big data could improve our ability to determine whether an intervention on the behaviour of a (putative) constitutive entity changes simultaneously also the behaviour of the phenomenon-to-be-explained.

Another example can illustrate this possibility: the constitutive mechanism exhibiting the phenomenon of segregation. In chapter 3, section 3.2.2, I have described Schelling's mechanistic model of segregation, where the individuals of a population can be understood as the relevant constitutive entities of the mechanism. By interacting with each other, each free individual will look for the neighbourhood with the preferred composition. Since if an individual abandons his neighbourhood, his activity changes the composition of both his past and future neighbourhoods, other individuals might react to such changes and decide to change their neighbourhood as well. Due to the continuous activities of the agents, everyone will end up living in segregated neighbourhoods. Even in this case, if we changed the behaviour of one of the individuals (for instance, if we made him move to another neighbourhood), we would simultaneously change also the phenomenon of segregation. Consequently, it would be possible to establish that the individual is constitutive of the phenomenon-to-be-explained. It is important to note that a change in the phenomenon-to-be-explained might be almost insignificant for the phenomenon itself, which could continue to exist despite having been modified. There

might be, indeed, two values S's Ψ_1-ing and S's Ψ_2-ing whose differences are so minor that both are considered to be the same phenomenon. For the phenomenon of segregation, for instance, the change of an individual's behaviour might not drastically change the phenomenon itself, that most likely will continue to exist. Observing such minor changes, however, helps us to identify the constituents of the phenomenon-to-be-explained.

The situation is much more complicated when we look for constitutive entities in a mechanism with a dispositional phenomenon. In such a case, the extended horizontal surgicality account requires the use of a thought experiment to imagine what would happen if the phenomenon were actualised. The possibility of thought experiments, in general, is not improved by the availability of big data. There is, nevertheless, a particular situation in which big data might help researchers to conduct such an experiment. If the phenomenon in question has been actualised in the past and data on the activity have been collected, the availability of such data could provide useful insights on the dispositional phenomenon-to-be-explained. Let us imagine again we are studying the orchestra ready to play Ravel's Boléro, before it starts playing its concert. Big data would help us to monitor the entities that we think might be constitutive of the dispositional phenomenon, but such data would not measure changes in the causal capacity of such entities. Similarly, we would not be able to measure anything about the dispositional phenomenon-to-be-explained.

In this condition, mechanistic background knowledge is crucial for reasoning on disposition, and often is the only source available to think of mechanisms with dispositional phenomena. However, let us imagine we have a video of the same orchestra playing Ravel's Boléro some days before the concert we are going to listen. We can verify that in the present we are observing the same entities that were recorded in the past. The data we can collect from the video, hence, can help us to conduct our thought experiment and to answer the question about constitutive relevance.

Overall, the extended horizontal surgicality account might be used by social scientists to identify constitutive relevant entities of mechanisms manifesting actualised behaviours. In very rare situations, moreover, the account could be used to find the constituents whose dispositional behaviours could make the phenomenon-to-be-explained possible. There are further aspects, however, that should be discussed when considering how to use big data to collect evidence of constitutive mechanisms. To begin with, it is questionable

whether social scientists could really use such an account when studying social phenomena. In addition, the practicality of this approach seems to be weakened by the distance between the approach and existing scientific research. These aspects are those that I shall examine next.

7.6 On the usefulness and feasibility of this account

In section 7.5 I have argued that there might be some situations where researchers can use the extended horizontal surgicality account and, through big data, identify the constitutive entities of a mechanism. There are, however, some observations that need to be discussed in relation to the feasibility of such an approach in the social sciences. I argue that there are different problems that might emerge when we use the extended horizontal surgicality account to study social mechanisms.

First, in section 7.3.3.2 it has been observed that often social mechanisms (such as social organisations) do not have physical boundaries and that, consequently, their borders are established by definition, analytically, according to cultural beliefs or norms. The lack of material boundaries in social mechanisms can lead to an important consequence: researchers can refer social phenomena to mechanisms that are not taken to be unified wholes, but composite systems.

To clarify the difference between these two possibilities, let us consider some examples. The phenomenon of pain can be referred both to the person (in psychological terms) and to the composite system made of neural activities. Similarly, the behaviours of the cells can be referred both to the cells as proper entities, and to the operations of their components, the cell organelles. On the contrary, in several cases social phenomena can only be referred to the activities of their constituents. If we consider a social phenomenon such as a riot, for instance, its definition necessarily requires information about the interrelations between individuals and their motivations. As recognised by Ramström (2018, pp. 372–373), there might be cases when we consider only the macro-level characteristics of the phenomenon: if we looked at a protesting group of Buddhist monks from a helicopter, we would just observe a square-shaped, slow-moving, and orange acting entity. Nevertheless, if we did not have information about the individuals constituting such an entity, their activities and their intentions, we would not be able to establish that such a phenomenon is a political protest, or even that the acting entity is performing a social phenomenon at all.

In this and similar cases, the usefulness and feasibility of any constitutive account, including the extended horizontal surgicality account, is questionable. A strategy to decompose and localise the different entities that make possible a phenomenon can provide useful insights only if we are not sure about the role played by such entities. The question motivating the use of the extended horizontal surgicality account should be "is the potential or actualised behaviour of this entity constitutive of the phenomenon-to-be-explained?". When the phenomenon is, by definition, composed of the behaviours of such entities, researchers using the extended horizontal surgicality account can only confirm what has already been established analytically. In these situations, consequently, monitoring an intervention on a constitutive entity (either with or without big data) does not advance our understanding of the phenomenon-to-be-explained.

Second, in order to use the extended horizontal surgicality account, social scientists need to decide how to distinguish between the different organisational levels of the mechanism, and unfortunately there are not precise rules on how to decompose this type of mechanism.

This issue has been recognised by different philosophers: Levins (1970), Kauffman (1971) and Wimsatt (2007) in several occasions argued that complex mechanisms in the social sciences (and, in some cases, also in the life sciences) can be decomposed in a number of different ways and it is often impossible to establish which one is the best. This problem, they claimed, is due to the fact that the division between levels is not guided by ontological considerations, but by explanatory reasons. In the social sciences this issue is made very evident when we consider non-physical entities. Let us consider the example of the central bank, treated as an entity with the capacity to influence the money supply. Such a phenomenon is related to the activities of some of its components (open money markets, auctions and regulative legislation), that cannot be spatially located. As a consequence, it is arguable whether they should all be considered at the same level or not. In other cases, when dealing with individuals' activities, different researchers might use different ways to locate individuals' actions, individuals' beliefs and social norms. The absence of precise ways to decompose constitutive mechanisms poses a challenge to the extended horizontal surgicality account because it requires that the intervention I is a direct cause *of at most one behaviour on every level* lower than the level of the phenomenon-to-be-explained. If researchers do not agree on the level on which different

entities should be located, it is not possible to establish whether this requirement is satisfied or not.

Third, even if researchers study a phenomenon generated by a constitutive mechanism understood as a whole, and there is an agreement about where to locate all the entities that might compose the mechanism, there is a further challenge that they need to overcome in order to apply the Extended Horizontal Surgicality: they have to find a way to surgically intervene on the mechanism and its constitutive entity.

While experimental studies might allow, in specific (and, arguably, rare) circumstances to intervene on biological phenomena, in the social sciences it is very difficult to surgically intervene on the phenomena-to-be-explained. Simple cases like the orchestra's example lend themselves well to this kind of approach, but most of the phenomena studied by social researchers are too complex to allow for surgical interventions, especially when the mechanisms might be constituted by non-physical acting entities such as social norms and individual's beliefs. When the intervention is not feasible, this account cannot be used to collect real evidence of the mechanism, but can nevertheless provide a heuristic (understood as a research guide or rule of thumb) for how to uncover some mechanistic information.

By considering these three observations together, it can be argued that, for social phenomena, the extended horizontal surgicality account can improve our understanding of the mechanism or of the phenomenon-to-be-explained only when:

1) constitutive mechanisms are taken to be *unified wholes*;

2) there is a general agreement on the level at which constitutive entities should be located (in this way, it is possible to establish if the intervention I is a direct cause of at most one behaviour on every level lower than the level of the phenomenon-to-be-explained);

3) it is possible to conduct a surgical intervention on the phenomenon-to-be-explained (otherwise, the Extended Horizontal Surgicality account can only be used as a heuristic to think about the mechanism and the phenomenon under study).

There are particular cases where these requirements can be met. A case in point is the analysis of a firm's behaviour as the output of a mechanism made of routines. As recognised by several evolutionary economists (Becker et al., 2005; Nelson & Winter, 1982; Vromen, 2011), firms are in general conceptualised as *wholes* whose behaviours

are made possible by the operation of several routines such as business and advertising strategies, through procedures for hiring and firing, and policies about investment, research and development. Such routines are understood as multi-person phenomena themselves, involving the actions and interactions of different individuals, and there is a strong agreement in the literature about the difference between such organisational levels (the level of the firm's phenomenon-to-be-explained, the routines' level and the individuals' level). Not only there is a general agreement on the level at which constitutive entities should be located, but routines can also be changed by means of interventions. A firm's hiring routine, for instance, might be changed in order to accomplish the same task (hiring people) with less effort and in less time (Feldman, 2000). A change in such a routine could maintain the other routines unaltered (thus, would be horizontally surgical) and might cause simultaneously a change in the firm's behaviour[24]. Intervening on an organisational routine to investigate its role in a firm's phenomenon-to-be-explained, consequently, is feasible and might allow researchers to collect useful evidence of the mechanism under study.

In numerous other circumstances, however, such requirements cannot be met, and the extended horizontal surgicality account can only be used to think mechanistically about social phenomena.

7.7 General consideration

There is a final observation that, I argue, is important to discuss. Although Craver's original account of constitutive mechanisms was informed by real neuroscientific studies and was aimed at providing a practical set of norms to identify the constituents of a mechanism, the philosophical discussions on constitutive mechanisms have, with time, turned away from scientific practices, developing accounts focused on metaphysical questions that could hardly be applied.

Such philosophical accounts are meant to provide a discussion concerning what interventions on constitutive mechanisms are. Their aim is not to discuss the main methodological challenges that social scientists have to solve in order to collect evidence of constitutive mechanisms. The use of scientific examples has been replaced by the use

[24] As recognised by Becker et al. (2005, p. 1): "a central proposition of routine theory is that organizations change what they are doing and how they are doing it by changing their routines".

of increasingly complex formalisations, which have little to do with how researchers study constitutive mechanisms nowadays.

Attempts to improve the current accounts of constitutive mechanisms to provide a useful approach to scientific practitioners, consequently, face one main challenge: the distance between such discussions and scientific practices weakens the possibility of using these discussions to clarify how scientific studies on constitutive mechanisms are or should be conducted. This, in turn, makes difficult to overcome the third challenge guiding this thesis, namely how to use big data, in the social sciences, to obtain evidence of causality that goes beyond correlations (in this case, evidence of constitutive mechanisms).

While the emergence of new philosophical accounts, over the last decades, has helped to cast light on the importance of establishing the presence of causal mechanisms, how to establish such a presence in the social science is still partly unexplored in the philosophy of the social sciences. To fill this gap, I argue, is important to develop philosophical accounts that dynamically interact with scientific research. In the next chapter, I shall try to do it by developing a new approach to trace etiological mechanisms.

7.8 Conclusion

In this chapter, I have investigated whether big data can be used to collect evidence of social constitutive mechanisms. The first part of the chapter has described and analysed the existing proposals to uncover constitutive relevant entities of a mechanism. To begin with, I have shown that Craver's mutual manipulability account, despite having the merit of having paved the way for more detailed discussions on constitutive mechanisms, has two conceptual limitations that Craver has not managed to solve. After having described some of the solutions to the mutual manipulability account's problems proposed in the literature, I have analysed in detail three recent proposals for constitutive relevance.

I have observed that the first proposal, Baumgartner and Casini's abductive theory of constitution, is not able to account for the problem of multiple realizations. Next, I have argued that the second proposal, Krickel's causation-based account for constitutive relevance, is characterised by several conceptual and methodological limitations associated with the notion of temporal entity-involving occurrents, the phenomenon of multiple realizations, and the assumption that constitutive mechanisms always entail activities. Finally, I have described the most recent approach for constitutive relevance

proposed by Baumgartner, Casini and Krickel: the horizontal surgicality account. I have shown that this account, thanks to the observation that constitution, unlike causation, does not take time, can solve the problem of multiple realizations. I have nevertheless argued that two limitations still pose some problems to this account: on the one hand, constitutive entities are required to be parts of the mechanism, but there are cases where this requirement is hardly met; on the other hand, the account does not allow for dispositional constitutive mechanisms.

By taking into account these problems, in section 7.4 I have developed an extended horizontal surgicality account, which does not have a parthood requirement and allows for dispositional constitutive mechanisms. I have then used such an account to investigate how to use big data to collect evidence of mechanisms. My analysis has led me to conclude that big data can be very helpful especially when the phenomenon-to-be-explained is actualised, and that in some circumstances (when data on the same phenomenon have been collected in the past) big data can improve our studies also when the phenomenon-to-be-explained is dispositional.

Next, I have observed that, in the social sciences, there are many restrictions regarding the feasibility and usefulness of this account. The extended horizontal surgicality account can help researchers to collect useful evidence only when constitutive mechanisms are taken to be *unified wholes*; there is a general agreement on the level at which constitutive entities should be located; and it is possible to conduct a surgical intervention on the phenomenon-to-be-explained. While in some cases, like for firms' routines, such requirements can be met, there are numerous social mechanisms that, due to such restrictions, cannot be studied through the extended horizontal surgicality account.

Finally, I have discussed a general observation: there is a distance between the existing discussions on constitutive relevance and the scientific practices aimed at uncovering causal mechanisms. Due to this distance, it is questionable if any of the existing constitutive relevance accounts can be used to clarify how scientific studies on constitutive mechanisms are or should be conducted (and, consequently, how big data might help to collect evidence of constitutive mechanisms). More scientifically-informed accounts, I have suggested, might provide more useful insights into the ways in which big data can be used to collect evidence of mechanisms.

8 Tracing etiological mechanisms: the role of sociomarkers

8.1 Introduction

Most of the phenomena studied in the social sciences evolve over time and can be analysed in terms of causal processes: phenomena such as economic crises, the emergences of civic associations, and the outbreaks of wars can be understood as the outcomes produced by etiological mechanisms. On this perspective, collecting evidence of mechanisms entails tracing the processes[25] *causing* the phenomena-to-be-explained.

As I observed in chapter 6, section 6.5.2, at the methodological level this approach is known by the name of process tracing and is associated with the idea of identifying the causal chain connecting the events that, ultimately, lead to the phenomenon under study. Process tracing has been often used and discussed by social scientists and philosophers of the social sciences (Beach & Pedersen, 2013; Bennett & Checkel, 2015; Ruzzene, 2014). In political science, for instance, process tracing is in general considered an approach casting light on the way in which a process is triggered in an actual situation. In psychology, it has been associated with experimental techniques aimed at reconstructing the pathways from information processing approaches to individuals' choices (Ford et al., 1989). In anthropology and sociology, Malinowski's study (1935) has been discussed as an exemplary case (Goldthorpe, 2016). Malinowski tested a hypothesised process linking the number of wives and wealth among Trobriand chiefs: according to such a process, the primary form of wealth in Trobriand society consisted of yams, and every man was required to offer an annual contribution of yams to the husbands of his married sisters. Consequently, in Trobriand society the more wives a man had, the wealthier he was[26].

In chapter 6 I have argued that process tracing can be used to trace also biological processes and that the use of biomarkers in causal studies, especially when aligned with the 'meet-in-the-middle' approach, can be conceptualised in these terms. By analysing

[25] In this chapter, etiological mechanisms and processes will be used as synonyms to represent causal chains linking causes and effects through time and space.

[26] Although this study was the first example of process tracing proposed by Steel (2004), who argued that Malinowski found a chain of causation from one variable to another, it might be argued that this mechanism can be better conceptualised as a constitutive mechanism, rather than as an etiological mechanism.

some biomarkers studies, indeed, I have shown that biomarkers are used to find the intermediate events between exposure and disease.

In this chapter, I shall start from the analysis of LIFEPATH to propose a novel way to trace social processes based on the use of markers at the social level[27]. As I will explain, the novelty resides in the way in which data are used to obtain evidence to establish the presence of a particular causal process. Rather than using data to *represent* social phenomena, such pieces of information can be used to identify *markers* to trace the etiological mechanisms linking the putative causes to the putative effects.

My analysis will start by considering how social data are used at the moment in causal studies. In section 8.2 I shall claim that social phenomena are in general studied by means of indicators and proxies, and I will examine the differences between such measures. Next, in section 8.3 I shall argue that the use of indicators and proxies does not allow researchers to uncover the mechanisms responsible for statistical associations with social factors, and I will argue that social measures and biomarkers are used in very different ways. This discussion will lead me to develop a general notion of markers and, in section 8.4, to adjust this notion to provide a notion of sociomarkers. Sections 8.5 and 8.6 will then provide two examples of sociomarkers. In section 8.5, I shall propose an example of sociomarkers that might help to understand how the social can 'get under the skin' causing biological changes. In section 8.6, I will consider how data from sociometric badges might be used as sociomarkers to trace social mechanisms. In section 8.7 I shall discuss how markers at the individual level can be used to collect evidence of mechanisms operating at the population level, while in section 8.8 I will argue that big data might be used to obtain sociomarkers. Section 8.9, finally, will conclude.

8.2 Measuring the social

It is evident that any causal study involving factors at the social level requires the use of data to *measure* them. How to measure social factors, however, is a matter of debate, especially considering that not all social factors are conceptualised in the same way.

[27] The material discussed in sections 8.2, 8.3, 8.4, 8.5 and 8.7 has been partly published in Ghiara and Russo (2019). Section 8.2, 8.3, 8.4 and 8.5 were originally developed by me and then discussed with Federica Russo, while section 8.7 is the result of a joint work with Federica Russo. The material originally developed by Federica Russo and published in Ghiara and Russo (2019) is not used in this thesis.

Some concepts refer to certain qualitative or quantitative characteristics that things have, such as the average age at marriage in a specific country, the number of minors in a neighbourhood, and the average female pay rate. Other concepts categorise things according to several loosely interconnected features, as in the case of well-being, poverty and socio-economic position. To give an example, the gross domestic product (GDP) per capita has been used to represent the individual standard of living in a country. As Martha Nussbaum highlighted, however, GDP is not able to represent all the relevant aspects that should be considered when measuring such a standard. An emblematic case is the case of South Africa under Apartheid, which, despite having a high GDP per capita, was characterised by serious inequalities that threatened the welfare of a considerable group of citizens (Nussbaum, 2013, p. 47).

When concepts cover specific features, the selection of the methods used to generate accurate and precise measures is more straightforward and such measures are generally direct. There is no need to develop indirect measurement procedures: as an example, in order to know how many women live in a neighbourhood, researchers can directly count them. On the other hand, when a concept is multifaceted, the choice of the measurement procedures is controversial and is likely to entail the sacrifice of some of its aspects. The GDP example illustrates this point, but another example might be the concept of socio-economic position, which is thought to refer both to material and social resources and assets, and to individuals' rank or status in a social hierarchy (Krieger, 2001). The measurement procedure is in general indirect and requires the selection of indicators. In this case, researchers have two options: they can either develop/choose various indicators to measure all the dimensions characterising such a concept, or they can select more precisely only some of the features constituting the concept and measure them, accepting to ignore the other dimensions characterising the concept. The selected features, then, can be measured by means of aggregate variables, developed by aggregating several indicators.

It should be noted that this distinction between simple and multifaceted concepts does not correspond to the distinction between observable and unobservable things. Although certain simple concepts can be directly observed (like a person's sex[28]), there are cases where observation cannot be performed with precision (as in the case of age). As for sex,

[28] In rare situations, determining sex might be difficult because some individuals have a naturally occurring intersex variation and show both female and male biological characteristics.

any doctor is able to assign biological sex to a person by considering observable bodily characteristics. Age, instead, is not always easily observable. In the case of young individuals applying for international protection without proper birth documentation, for instance, physicians use measures obtained through wrist radiographs and clavicle computerised tomography to estimate age (Sauer et al., 2016).

The fact that some concepts are not easily observable has a great impact on the social sciences, as measuring them appears challenging. Researchers tackle this problem through the use of proxies. Unlike indicators, proxies are not supposed to *represent* concepts. A proxy is a measure that *correlates to the unobservable or unmeasurable concept* under study sufficiently highly to be used as its predictor (Sapsford, 2006, p. 145).

Some examples can clarify this point: let us consider the way in which researchers study socio-economic position. In low-income countries, given the shortcomings of available data, the presence or absence of night-time lights visible from space is increasingly used as a proxy for the concept of socio-economic position (Chen & Nordhaus, 2011). In the UK, moreover, this notion is often examined through the use of the Free School Meal Eligibility proxy (Hobbs et al., 2007) and the council tax band (Howe et al., 2012). The number attached to the presence of night-time lights in a specific zone, the Free School Meal Eligibility and the council tax band are not supposed to represent the concept of socio-economic position. Rather, given the absence of available measures representing it, researchers use alternative data to obtain information about that concept.

The adequateness of indicators and proxies entails, hence, different questions. As for direct and indirect measures used as indicators, researchers have to establish if the measures correctly *represent* the phenomena. In other words, they have to examine the adequateness of the relation between the concept and the numerical measurement.

The idea that social indicators represent social phenomena has been proposed by many researchers. Reading discussions on social indicators, it is easy to find claims such as:

> "The purpose of indicator-building is to represent the complex reality – the state of a society, its social, economic and ecological connections, and their development – by means of an informative indicator" (Radermacher, 2005, p. 163)

> "Indicators represent—select, depict, condense, and summarize—conditions, processes, relationships, structures, and other phenomena" (Woodson, 2007, p. 135)

> "Indicators differ substantially with regard to their ability to numerically represent the phenomenon they are said to represent" (Gerrets, 2015, p. 155)

In the case of a proxy, adequateness requires the *correlation* between the proxy and the phenomenon to be sufficiently high. These distinctions are illustrated in Table 7.

Terminology of concepts	Meaning	Example	Measures	Purpose
Simple concepts/Complex concepts	Concepts cover specific features/Multifaced concepts	Number of women in a workplace/Poverty	Direct measures/Set of indicators	Representation
Observable/Unobservable	Concepts that can be directly observed/Concepts that cannot be directly observed	Sex/Age	Direct measures or set of indicators/Proxies	Predict the presence of the concept under study

Table 7. A recap of the terms used in the literature, where they apply and their purpose.

Recently, some researchers have argued that big data may offer both reliable proxies of social indicators and new reliable indicators (di Bella et al., 2018). According to this view, big data are useful because they provide new information that can, in turn, be used to measure traditional indicators and proxies. The use of night-time satellite images can help to clarify this point: without such data, it might be very difficult to obtain reliable measures on people's socio-economic condition in developing countries. Traditional indicators, indeed, require survey data that are often not available in such countries. In this way, satellite data allow to measure a proxy that is likely to be the only measure associated with people's socio-economic condition in such countries. This measure, nevertheless, is just a new proxy that has the same characteristics of a traditional proxy (like the council tax band proxy in UK): the correlations between them and the direct measure of poverty appear to be sufficiently high and stable. In the rest of this chapter, I

shall investigate whether big data can be used to obtain not just indicators or proxies, but a new type of evidence of etiological mechanisms.

8.3 More than correlations: learning from LIFEPATH

When researchers conduct causal studies, in general social phenomena are analysed through social indicators and proxies, which are then used to establish correlations. Some examples of this approach can be found in the case studies discussed in chapter 5 and in chapter 6. When Garip and Asad (2016) studied the phenomenon of Mexico-U.S. migration, they used indicators and proxies measuring political, socio-economic, and demographic factors to find significant correlations to individuals' likelihood to migrate.

Similarly, in LIFEPATH researchers study the social determinants of health by measuring three aspects of socio-economic position (education, occupation, and income) and using such measures as indicators. Such indicators are then used to find statistical associations with biological factors. In such a way, researchers manage to accumulate a large and compelling body of evidence demonstrating that socio-economic factors are correlated to a wide range of health outcomes. Unfortunately, however, often the use of socio-economic indicators and proxies does not help to uncover the causal mechanisms behind such correlations.

The case of LIFEPATH is particularly interesting because it is possible to compare the way in which researchers use socio-economic indicators with the way in which researchers study biomarkers. As I have argued in chapter 6, biomarkers can be used to trace the biological mechanisms leading to the formation of diseases. Socio-economic indicators, nevertheless, can only be correlated to such mechanisms. Considering this difference, it might be questioned whether, like in the case of biological data, also social data can be used as 'picking up signals' to trace etiological mechanisms at the social level.

8.3.1 Towards a general notion of markers

In chapter 6, I have argued that, traditionally, epidemiology established correlations between population-level variables: those associated with the environment and those describing diseases. The identification of biomarkers, instead, allows researchers to go much smaller. Not only the population-level factors themselves, such as air pollution, can be studied at the chemical level, but also their effects are now measured at the molecular

level. By considering all these factors together, researchers can finally reconstruct the process leading to disease and establish causal links between exposure and disease.

The analysis of markers at the biological level has led me to observe, in section 6.5.1, that: i) a biomarker can be any measure of a biological entity, quality or event; ii) a biomarker is objectively measurable; iii) a biomarker is linked to the causal process under study either because is directly involved in the process, or because is caused by the same causal factor A, is caused by another unmeasurable factor involved in the process or is a background condition of the causal process (as illustrated in Figure 18); iv) a biomarker helps researchers to develop a causal hypothesis or to provide evidence for a hypothesised causal process linking the cause A to the causal effect B.

I can now try to adjust these features to develop a general notion of markers, that can be applied to different domains. If we consider scientific research in general, it can be said that it is aimed at developing causal explanations of phenomena: researchers do not only aim to claim that some factors are correlated, they want to establish that some factors are *causally linked* and want to explain *how* something can cause something else. While the recognition that there is a causal link can be obtained through different methods; typically, the 'how' question is answered by considering the etiological mechanism linking the putative causal factor to the effect. To do so, researchers have to detect *something* that can help them to identify the link.

When the causal process appears to be particularly complex, researchers might trace it by means of some *markers*. These markers are clues, signals to detect in order to trace the chain linking the putative cause to the final outcome (Illari and Russo 2016). In many cases markers are not just measures of 'objects out there': they can be measures of actual entities, but can also be just measures of specific characteristics or events offering some insights into the causal process under study. In all these cases, when something is used as a marker, it should be possible to objectively measure it. Similarly, in order to use something as a marker, researchers should be able at least to hypothesise the reason for the correlation between the marker and the phenomenon under study: in general, such a correlation is due to one of the situations represented in Figure 18 of chapter 6. Markers, finally, are always defined by the function they have in a particular process of inquiry: nothing is a marker *per se*, and any measure of an entity, quality or event can become a

marker if researchers think that it can help to develop a causal hypothesis or to provide evidence for a hypothesised causal process.

While biological markers are often used as signals to trace processes that happen at the biological level, the use of data at the social level is still based on the traditional category of direct measurements, indicators and proxies. My argument is that, even at the social level, it is possible to pick up signals of causal processes. I call these signals sociomarkers.

8.4 Sociomarkers

Sociomarkers are markers, signals at the social level that allow for the recognition of a causal process. Sociomarkers can hence be used to trace both entire processes at the social level, and mixed processes consisting of both social and biological factors. As I shall discuss, a sociomarker might be for example the measure of a specific type of emotional and physical abuse in childhood that helps researchers to reconstruct the causal process from social factors at the population-level, such as the family's socio-economic position, to health outcomes. Sociomarkers, furthermore, might be the measures obtained from behavioural data in the workplace that allow for the reconstruction of the causal process from gender diversity in the workplace to the discrepancy between men and women in senior positions.

Above, I have argued that the key characteristic of markers is that they are always markers of something. In the social sciences, this feature allows for the distinction between sociomarkers, indicators and proxies. When something is measured to be used as an indicator, it is supposed to represent the phenomenon under study. When something is measured to be used as a proxy, furthermore, it is sufficient to know that it is correlated to the phenomenon under study. On the contrary, when researchers want to use a measure as a sociomarker, it is not required that the sociomarker represents a phenomenon, and it is not sufficient to establish a correlation to the phenomenon.

It is important to note that, to shift from the notion of social indicators and proxies to the concept of sociomarkers, there is no need to obtain new measures: those already used as indicators and proxies can indeed be used as markers. What is required is a change in the way in which the measures are employed. The aim, when using sociomarkers, is to pick up signals to reconstruct the process from the putative cause to the effect, in analogy with how biomarkers help to pick up signals from exposure to disease at the biological level.

Like biomarkers, there are five general situations in which we can use sociomarkers. Let us consider again Figure 18 of chapter 6. First, a sociomarker can be part of the process from the cause to the effect. For instance, in chapter 5 I have described the mechanism whereby a society characterised by a 'migration culture' can offer compensation to the members who decide to migrate, and this can cause more people to migrate. In such a case, a sociomarker measuring how such a migration culture is translated into behavioural factors might help social scientists to trace the causal process leading to the phenomenon of migration, as illustrated in Figure 18(a). Second, a sociomarker and the outcome under study might be caused by the same cause, as showed in Figure 18(b). In such a situation, the sociomarker is not part of the process, but it is one of the effects caused by the cause producing also the outcome under study. For example, let us imagine researchers want to understand if relational disorders in adulthood are caused by household dysfunctions in childhood. They might know that the putative cause (household dysfunctions in childhood) is likely to produce, among other things, low scores at school in childhood. Even if researchers do not think that low scores at school are directly involved in the process from household dysfunctions in childhood to relational disorders in adulthood, they could still use score measures to get insights into the hypothesised causal process (for instance, such a sociomarker could strengthen the hypothesis concerning the presence of household dysfunctions). Furthermore, this marker might be very helpful in the case described by Figure 18(c), where an event or a factor that is part of the process cannot be observed, but the correlation between this and another measurable event or factor allows for the use of such an event or factor, not directly involved in the process, as a sociomarker. For instance, measures of low school scores in childhood could be used as a sociomarker to replace another unmeasurable event playing an active role in the causal process. Fourth, there might be an unmeasurable factor in the process that influences something (a putative marker) outside the process, like in Figure 18(d). Let us consider the 'hikikomori syndrome', a form of severe social withdrawal characterised by adolescents and young adults who decide to live as recluses in their parents' homes, without working or studying for months or years. In such a case, researchers might be interested in the process that, from performance anxiety, leads to the 'hikikomori syndrome'. Researchers might hypothesise that bullying experiences at school could act as a mediating sociomarker, but might not be able to measure them. However, they could also hypothesise that such experiences would influence academic performance (that is not thought to be part of the process) and might hence use the measure of changes in academic

performance as a sociomarker. Fifth, the sociomarker might be the background condition, or the cause of the putative cause leading to the observed effect, as illustrated in Figure 18(e).

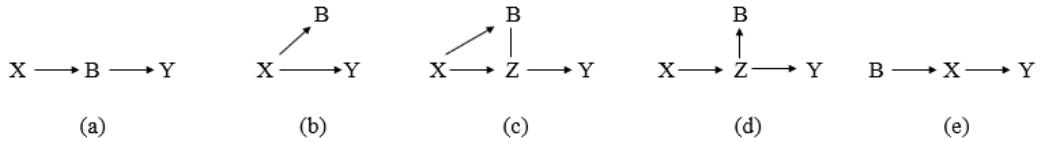

Figure 18, retrieved from chapter 6, section 6.4. X and Y represent the putative cause and effect, B represents the sociomarker.

8.5 Sociomarkers in health mechanisms

8.5.1 Exploring socio-biological processes

It is a common causal claim that certain diseases arise from a set of risks identified at the behavioural and social level, such as people's smoking, dietary habits and social norms (Galea & Link, 2013). Lung cancer, for instance, is caused by smoking, but smoking is influenced by social networks, advertising and social norms (Christakis & Fowler, 2008). It is also generally accepted that the health of populations is shaped by certain 'social determinants' like social ties, poverty, and low education (Braveman & Gottlieb, 2014; Wilkinson et al., 2003). Despite this broad core of agreement, however, social and behavioural factors are seldom brought into the etiological discussion of mechanisms of health.

This is partly because social factors are thought to provide merely a classification of the parts of the population more at risk. Their relevance is associated with categorization, not explanations. Furthermore, even when it is accepted that social factors play an active role in the development of disease, how to integrate socio-behavioural and biological variables to develop all-encompassing models of disease causation remains a matter of debate. For a long time, researchers have considered that biological and social factors are difficult to be integrated due to a conceptual barrier. This assumption is still evident in the linguistic use of terms such as "proximal" and "distal" causes, or biological and wider determinants of health (Kelly et al., 2014, p. 309), and in some discussions where the relevance of social factors is reduced to people's risky behaviours that may affect health (Kelly & Russo, 2017).

In the last few decades, many have appealed to a new integrated model of disease causation, where biological and social factors are considered equally relevant to the causal understanding of health (Kelly et al., 2014; Russo, 2009; Vineis & Kriebel, 2006). In an editorial accompanying the reprint of their famed article "Actual Causes of Death in the United States" (McGinnis & Foege, 1993) the authors themselves considered that:

> "[…] it is also important to better capture and apply evidence about the centrality of social circumstances to health status and outcomes" (McGinnis, 2004, p. 1264)

This and similar positions offered a raison d'être for those social epidemiological studies, like LIFEPATH, putting at the heart of their investigations how social factors might influence health. Nevertheless, attempts to enhance our understanding of the social component of health have been often downgraded by the increasing gap between the progress in biological research and the progress in social research. Indeed, while from the biological point of view the striking technological improvements have contributed to enhancing the mechanistic understanding of disease; at the social level the progress has not been as fast (Kelly et al., 2014).

This brings us to the present, when the biological component of the mechanisms of health is explored at a fine-grained level, but the social component is analysed at a rather coarse-grained level[29]. As I have discussed in chapter 6, socio-economic position is in general measured by using three descriptive social indicators: income, occupation, and educational level. The studies based on such indicators have provided useful information, and yet they suffer from substantial limitations. The main problem deals with the kind of knowledge that can be derived from this approach: the analysis of the socio-economic indicators highlights *that* socio-economic factors are relevant to health, however it does not establish *why* this is the case. In other words, while these measures can show that there are robust *correlations* between specific socio-economic conditions and health, they

[29] It should be noted that I use the expression "coarse-grained" to define only the way in which the social dimension of health is often studied in social epidemiology. My intention is not to categorise all social studies as coarse-grained, neither I want to argue that this is the aspect that allows researchers to identify the differences between the natural and the social sciences. I am aware that there are several considerations to take into account when the social and the natural sciences are compared, and that they differ in many respects. In what follows, I only suggest one way to overcome one limitation characterising certain social epidemiological studies.

are not sufficient to uncover the *mechanisms* through which such socio-economic conditions affect health.

8.5.2 Tracing processes through sociomarkers: the ACEs example

My suggestion is that the limitation described above can be tackled by using sociomarkers. Sociomarkers can be employed to trace social processes and can help to find the missing links between social factors at the population-level (such as the family's socio-economic position) and health factors at the individual-level (such as inflammatory biomarkers). The combination of sociomarkers and biomarkers, in this way, can finally help researchers to uncover these mixed etiological mechanisms whereby 'the social' influences health.

More specifically, the combination of sociomarkers and biomarkers might enable researchers to identify all the key stages of a mixed mechanism of disease, from the social components triggering the mechanism, to certain key events involving social or biological components, to the biological factors leading to the final outcome. In such a way, researchers might avoid the problematic situations in which detailed reconstructions of the biological processes of health are linked to the social domain only through the correlations between some biomarkers and very coarse-grained social variables.

Recently, researchers have begun to use the label 'markers' for socio-economic measures in a rudimentary way. For instance, measures such as birth weight, household income, education, maternal smoking during pregnancy, mother's body mass index and mother's age at birth, known to affect health conditions, were used to explore how the link between socio-economic position and health is mediated by certain social pathways (Barboza Solís et al., 2016; Davillas et al., 2017; Shin et al., 2018). It is exactly when such measures are used, I argue, that a precise conceptualisation of sociomarkers is needed.

Let us consider an example: it is known that a low socio-economic position in early life is correlated to poor health and increased mortality in adulthood (Anda et al., 2006). In chapter 6 I have observed that, at the biological level, the effects of childhood conditions on health appear to be brought about by changes in the systems responsible for maintaining physiological stability, also known by the name of allostatic systems (Danese & McEwen, 2012).

At the social level, furthermore, an increasing number of projects on life chances and well-being have started focusing on the notion of 'life course' that was originally developed by some sociologists such as Cain (1964) and Elder (1976). The notion of 'life course' has been used in sociology to investigate how disadvantaged conditions in childhood may influence physical health in adulthood. Sociologists have often taken inspiration from Merton's analysis of cumulative advantage (also known by the name of 'the Matthew effect'), according to which early distinction in scientific research might lead to further honour and opportunities (Merton, 1968). Merton's analysis was not focused on health, it simply highlighted the challenges faced by individuals who did not have an auspicious scientific start. Many researchers of the life course, however, used the notion of accumulation to theorise how, in general, negative conditions might produce further negative outcomes. This consideration has been also incorporated in the 'cumulative inequality theory', which combines elements of the cumulative advantage theories (Dannefer, 2003; DiPrete & Eirich, 2006), of life course theories (Elder, 1998) and of stress process theories (Pearlin et al., 2005). Overall, in sociology several accounts prioritise childhood as a pivotal life stage leading to adult health (Ferraro et al., 2016), and claim that early inequalities might lead to discrepancies in how people are exposed to factors that could compromise health.

To detect 'social' signals of the process leading to health outcomes, some researchers have used the so-called Adverse Childhood Experiences (ACEs) (Ferraro & Shippee, 2009). This expression describes certain stressful or traumatic events that can be experienced by children whilst growing up. The original ACEs study was conducted by the Centers for Disease Control and Prevention (CDC) in collaboration with the American integrated managed care consortium Kaiser Permanente (Felitti et al., 1998). In this study, researchers developed a list of 10 categories of childhood adverse experiences, as shown in Table 8:

Adverse Childhood Experiences	
Emotional abuse	Physical neglect
Physical abuse	Emotional neglect
Contact sexual abuse	Not raised by both biological parents

Mother treated violently	Household member was imprisoned
Household member was alcoholic or drug user	Household member was chronically depressed, suicidal, mentally ill, or in a psychiatric hospital

Table 8. Categories of adverse childhood experiences.

Measures of ACEs have all the characteristics required to be used as sociomarkers.

ACEs measures can be employed to identify some key stages connecting social factors at the population-level, such as the family's socio-economic position, and biological factors at the individual-level (biomarkers). In such a way, they can help to establish how socio-economic determinants exert an impact on health by means of biological mechanisms.

More specifically, this is possible for two reasons. One the one hand, at the population-level, ACEs appear to be more prevalent among children in low-income families: in the U.S., for instance, data from the National Survey of Children's Health (NSCH) showed that children in low- and middle-income families are more than 2 times as likely to experience 2 or more ACEs compared to children from high-income families (Halfon et al., 2017). This prevalence, hence, supports the idea that socio-economic condition during childhood could affect health through these adverse experiences. On the other hand, several studies have provided evidence of the associations between ACEs and allostatic load biomarkers, inflammatory biomarkers and cancer biomarkers (Barboza Solís et al., 2015; Kelly-Irving et al., 2013; Tietjen et al.,2012). Such correlations support the idea that ACEs could be the link connecting 'the social' and 'the biological', as described by Danese and McEwen (2012):

> "adverse childhood experiences are likely to signal high levels of environmental threat and to trigger adaptive responses in the nervous, endocrine, and immune systems in children" (Danese & McEwen, 2012, p. 31)

ACEs, hence, offer much more information than traditional socio-economic indicators: rather than providing a representation of certain conditions, they help, through an approach very similar to Vineis and Perera's 'meet-in-the-middle' approach (2007), to investigate the etiological mechanisms whereby social conditions can get under the skin.

There is a caveat that needs to be discussed. Someone may argue that, rather than being markers of the etiological process from socio-economic circumstances to health changes, ACEs identify the real 'social' causes of health: ACEs would not be picking up signals of the process starting from socio-economic conditions, they would be the real social factors with a causal power, as described in Figure 34.

This ambiguity is due to the fact that there is little evidence of the causal process from socio-economic circumstances to ACEs: until now, it has been only demonstrated that children in low- and middle-income families are more likely to experience ACEs (Halfon et al., 2017). To solve this problem, the only solution would be to identify new sociomarkers in the process between socio-economic circumstances and ACEs. For instance, it might be hypothesised that low socio-economic circumstances lead to family economic pressure, that causes parents to increase their working hours and to reduce parental supervision on children, with the consequence that children experience physical and emotional neglect. In this case, measuring family economic pressure, parents' working hours and children supervision might help to obtain new sociomarkers and to trace the process from socio-economic circumstances to ACEs.

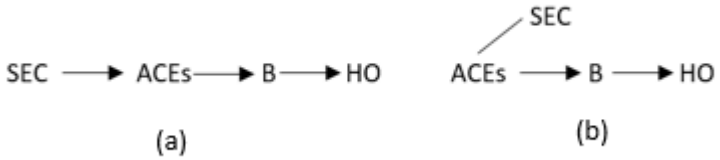

Figure 34. Someone might claim that ACEs are not, like biomarkers (B), signals of the process from socio-economic conditions (SEC) to health outcomes (HO), as shown in (a). ACEs are the causes of the process that, passing through biomarkers (B), could lead to health outcomes (HO). In a similar situation, socio-economic conditions (SEC) might only be correlated to ACEs, without being their cause, like in (b).

It is worth noting that researchers who trace a process from A to B are in general aimed at developing interventions on such a process to change B. Intervening on the process does not entail only intervening on the cause A. Markers help researchers to detect some key stages of the process that, if changed, might lead to a different outcome. It follows

that both the primary causal factor and markers can be the target of intervention. Let us consider the impact of childhood conditions on health: in order to develop an intervention, researchers might take into account population-level and individual-level social factors, such as the socio-economic conditions of the family during childhood and individual experiences of neglect and abuse. Similarly, they might consider biological factors, such as the maintenance of physiological stability. All these dimensions (population, individual, social and biological) could be relevant for the purposes of policy interventions.

8.6 Identifying sociomarkers from sociometric badges' data

The development of new data collection devices might provide new pieces of information that can be used as sociomarkers. A case in point is the development of sociometric badges. Sociometric badges are sensor packed devices with the size of a smartphone, able to generate fine-grained data for analysing patterns concerning face-to-face social interaction, non-verbal communication and turn-taking. Data produced by means of sociometric badges are called by Pentland (2008) 'honest signals': behavioural signals that humans unknowingly exchange and that might express their aim and emotional state.

My proposal is that sociometric badges' data share some important characteristics with biomarkers and can be used to identify *signals* or *markers* within specific research contexts. Sociometric badges appear to collect information mainly about events and characteristics. Such data, indeed, can help researchers to examine characteristics such as speech features, physical proximity to other people, and the length of time certain individuals spend talking face-to-face. Moreover, they can cast light on events such as the use of particular body movements and gestures (Pentland, 2008, p. 102). Like in the case of biomarkers, hence, sociometric badges' data provide information not only about entities, but also about certain characteristics and events.

It is important to note that, as for the ACEs example, measures from sociometric badges can count as sociomarkers if they are used in a specific way. As already highlighted, nothing is a sociomarker per se. The same measures can be used as simple indicators, representing specific characteristics of a concept, or as proxies. For instance, a measure of a specific body movement can be used as an indicator of aggressive behaviours if the measure is thought to represent such aggressive behaviours. Similarly, the same measure could be used as a proxy for autism spectrum (for instance, some body movements such

as hand flapping are known to be associated with autism) if it is thought that the measure is highly correlated to autism. When used as sociomarkers, measures from sociometric badges' data are used to *detect* a causal process at crucial points. As argued above, in this sense sociomarkers are not thought to be just correlated to the causal process, and researchers do not aim to represent a part of this process by using sociomarkers.

To examine how measures from sociometric badges' data can be used as sociomarkers, let us imagine a hypothetical causal study whose aim is to answer the question 'why fewer women than men end up in senior positions?'. The question might be motivated by the observation that gender diversity in the workplace is in general correlated to the number of women ending up in senior positions.

Gender diversity cannot be represented only by the number of men and women in the workplace. This is in general understood by social scientists as a complex concept involving different aspects such as demographic diversity (like the age and the marital status of women and men having the same position) (Riach et al., 2014) and functional diversity (such as differences in education, in hierarchical levels and roles, and in leadership styles) (Humbert & Guenther, 2017; Tost et al., 2012). Due to the multifaceted nature of this concept, researchers use in general aggregate variables to measure it.

To trace the putative process from gender diversity to the number of women ending up in senior positions, researchers can use measures from sociometric badges' data as sociomarkers. They can collect individual-level measures on aspects such as relevant speech features (like how many times a person interrupts another person), the time spent in face-to-face conversations, the voice's tone, the physical proximity to other people. In order to use such measures as sociomarkers, then, researchers need to hypothesise their position in relation to the causal process under study.

To give an example, researchers might hypothesise that a low number of women in the workplace leads men to behave differently with women than with other men, for instance interrupting women more often than other men. This might cause women to avoid expressing their opinions during business meetings, and might also lead them to network only with other women. Since there are fewer senior women, however, women would not network upwards, and men in senior positions might observe that such women do not have the self-confidence needed to be promoted. In this case, sociomarkers such as the measure of the times a person interrupts another person and the measure of the time spent

in face-to-face conversations would be located *in* the process, as in figure 18(a). Other sociomarkers might be outside the process (as in Figure 18(b), 18(c), 18(d)): the amount of time that women spend on individual work, for instance, might be influenced by the fact that networking with a smaller group of person requires less time, and the rest of the time is spent on individual work. In this case, the amount of time spent on individual work would be the effect of the choice to network only with women, as in Figure 18(d).

8.7 Population and individual level

In this chapter, I have mentioned several times the distinction between population and individual levels. In this section, my aim is to investigate better how such levels are linked, and why it is important to distinguish between them. Let us start with the *level at which measurements are performed*.

In scientific research, it is possible to distinguish between individual-level measures, measuring individual characteristics, and aggregate measures, that summarise some of the characteristics of the individuals composing a group and that can be developed, in some cases, aggregating several individual-level measures (Illari & Russo, 2014, p. 43). An example of an aggregate measure is, for instance, a measure of socio-economic position obtained aggregating educational, income and occupational measures. An example of individual-level measure is, instead, the income of each individual in the sample.

If we consider biomarkers and sociomarkers, most of the time[30] they are located at the individual level. Biomarkers, for instance, are often identified measuring biosamples of real, particular patients. Similarly, ACEs are identified by examining real individuals: they can be experienced only by real, individual children (this is particularly relevant if we consider that, for each individual, adverse childhood experiences can vary in duration, timing, and type, and that individual characteristics can lead to very different responses to very similar adversities). Sociometric badges, finally, monitor specific individual behaviours.

Yet, such individual-level measures are studied to *say something at the population level*. In biomarkers studies, for instance, researchers obtain biomarkers from individual-level

[30] There are cases in which biomarkers are aggregate measures. As I have discussed in section 6.4, allostatic load is measured by using more than 25 biomarkers, and quite often such indicators are combined to obtain a single aggregate variable (Juster, McEwen, & Lupien, 2010, pp. 10–11).

data (bio-samples pertain to individuals); but they make an epidemiological study out of it. This, in the early days of molecular epidemiology created tension in the field. As Schulte put it, it might sound an oxymoron to do epidemiology with molecular data (1993, p. 10). Like Schulte, I argue that this is not the case, and I claim that, to disambiguate the problem, it is necessary to rethink the relation between population- and individual-level causal knowledge in markers studies.

In the case of biomarkers the change of the scale of measurement, the size of the sample and the way in which inferences from '-omics' analyses are made, allow researchers to use individual-level markers to make inferences about the population, in order to understand population-level phenomena. In analogy with the study of biomarkers, also sociomarkers can be used to trace social process at the individual level. Such studies, however, can help researchers to gain understanding of the social mechanisms that operate at the individual level, but within a group or population. In both cases, hence, researchers can collect evidence of population-level mechanisms by measuring individual-level markers[31].

The use of individual-level markers does not exclude the use of aggregate variables or of multi-level methods that use both population- and individual-level measures. In some cases, for instance, it could be possible to use population-level sociomarkers that describe the role of structural causes (this would go into the same direction as Galea and Link's, according to which epidemiology is 'all social' (Galea & Link, 2013, p. 844)).

8.8 From big data to markers

The notion of sociomarkers, I suggest, is useful to shed light on how big data can be used in the social sciences. More specifically, this concept helps to clarify how big data can contribute to uncovering etiological mechanisms operating in the social realm, in parallel with how biological data can help to uncover etiological mechanisms operating at the biological level.

[31] Although these considerations might appear in line with the approach known by the name of methodological individualism, it is important to consider that my proposal is that sociomarkers can be measured to trace etiological processes. Methodological individualists, instead, are in general interested in constitutive mechanisms, where the constitutive parts are claimed to be the individuals, whose actions and organizations can constitute macro-level phenomena. An illustrating example is the segregation mechanism studied by Schelling and described in section 3.2.2.

Let us compare biomarkers and sociomarkers: in both cases, we can establish that most of the data used to identify markers can be defined as big data. By considering the example above and in chapter 6, it is possible to say that both datasets containing data from '-omic' technologies, and the collection of data generated through sociometric badges, are characterised by large volume and are in general collected through automated processes. To give an example, transcriptomics allows researchers to obtain information about fifty thousand genes transcribed into RNA molecules. Similarly, sociometric badges can produce datasets containing thousands of measures: in their study, for instance, Lepri et al. (2012) collected 15725.35 hours of bodily activity, 15894.63 of audio data and thousands of Bluetooth detections concerning the proximity between devices. Moreover, thanks to new automated forms of data collection such data can potentially be characterised by 'high velocity' (in terms of the speed with which data can be shared, and the frequency of incoming data that need to be processed). Indeed, when using novel technologies like '-omic' methods and sociometric badges, researchers can collect huge amounts of information in very short periods of time, and the time interval between one collection and another can be really short.

Claiming that big data can be used to identify markers, however, does not entail denying that the same data can be used properly to measure proxies or indicators. The same measures, in different research contexts, might be used to represent a phenomenon or to trace a mechanism. Furthermore, it is perfectly possible to use traditional data to obtain sociomarkers, like in the case of ACEs data. My argument is that the novel idea of sociomarkers can help to answer the question of whether big data can be used, in the social sciences, to obtain evidence of causality that goes beyond correlations. While this possibility does not emerge in a clear way when big data are used to obtain indicators and proxies, when researchers use big data to identify sociomarkers, they obtain evidence of etiological mechanisms.

This consideration, I suggest, can also provide a useful answer to the question of what should be done of all available data in the social sciences. While in some cases new data can be used to obtain more accurate indicators and proxies, in some occasions the amount of data could be so vast that collecting new data might not improve the way in which social factors are measured and represented. New data, however, if used to identify markers, could help to trace new key stages of causal processes.

8.9 Conclusion

This chapter has provided an answer to the question of whether big data can be used, in the social sciences, to obtain evidence of etiological mechanisms. At the beginning of the chapter, I have argued that social phenomena are in general studied by means of indicators and proxies. Such types of measures can help researchers to identify correlations, but in general cannot be used to uncover the causal mechanisms responsible for them.

By considering the differences between biomarkers and social indicators and proxies, I have observed that it might be possible to use social data to obtain sociomarkers. To clarify what sociomarkers are and how they can be used, I have described two examples. First, I have argued that measures of adverse childhood experiences can be used to trace the causal processes that, from low socio-economic position, can lead to the development of diseases. Second, I have suggested that measures obtained from sociometric badges' data might be used as sociomarkers to uncover causal processes operating in the workplaces. In both these examples, I have argued, researchers can use individual-level sociomarkers to make inferences about the population. Finally, I have suggested that the notion of sociomarkers can be particularly helpful to understand how to use big data in causal studies. Not only, indeed, the examples I have examined support the idea that big data can be used to obtain sociomarkers, but this observation might also provide an answer to the question on what should be done of all available data in the social sciences. Rather than using big data to measure new indicators and proxies and, in turn, to identify new correlations or to confirm the presence of correlations already studies, big data might be used to find some key stages of etiological mechanisms.

9 Conclusion

The aim of this thesis was to investigate how big data are changing the ways in which causal studies are conducted in the social sciences. In chapter 2 I have clarified that big data do not provide magical solutions to tackle some of the methodological challenges typical of causal studies (such as confounding, Simpson's paradox and overfitting), and that data quality and interpretation might still cause problems in causal studies. Such observations have then been used to identify three key challenges concerning the use of big data in the study of social phenomena: i) how to overcome the limitations of data-driven causal studies, ii) how to understand the role of thick data in causal studies based on big data, iii) how to use big data, in the social sciences, to obtain evidence of causality that goes beyond correlations.

9.1 Overcoming the limitations of data-driven causal studies

I have provided a solution to the first challenge in chapter 3, where I have argued that most of the problems characterising data-driven studies can be tackled by complying with the Russo-Williamson thesis, according to which causation can be established only if it is possible to identify both a correlation between the cause and the effect, and a mechanism linking the cause and the effect. I have argued that RWT can be used in the social sciences, and I have defended it from criticisms.

More specifically, I have identified four criticisms discussed in the literature: i) RWT is not used in the social sciences, ii) RWT does not work, iii) causal concepts do not overlap and iv) there are better alternatives to RWT. To answer such criticisms, I have observed that, in general, RWT is used successfully in the social sciences, and that it is supposed to reduce inferential mistakes, not to cancel all of them. I have next argued that the cases where causal concepts do not overlap are rare and can be solved by thinking counterfactually, and that RWT is the approach most consistent with how social scientists establish causal relationships. I have finally described some examples where the identification both of a correlation and of a mechanism between the cause and the effect might help to uncover data-driven problems such as Simpson's paradox and low-quality data.

9.2 Thick data

I have explored the use of thick data in chapter 4, starting from some of the observations discussed in the mixed methods research literature. The parallel between MMR and thick data discussions is justified by a couple of observations: both MMR and thick data researchers aim to cast light on the role and importance of qualitative data, in addition there are different cases where the qualitative/quantitative distinction and the thick data/big data distinction overlap. Starting from such a parallel, I have claimed that, as for qualitative data in MMR, thick data in new big data studies can help scientists to collect evidence to establish causal claims. Indeed, while big data are generally studied to identify correlations between variables, thick data can be analysed to collect evidence of mechanisms. The combination of big and thick data, hence, can be used to meet the requirements of RWT.

Furthermore, I have proposed that thick data, like qualitative data in MMR, can also be used to obtain a more comprehensive image of causal phenomena. On the one hand, thick data might allow researchers to mix different ontological categories such as the categories of singular and general causation. While in chapter 4 this observation has been supported by analysing Wood's study (2003) on the reasons why peasants in El Salvador decided to join rebel movements, the detailed case study analysed in chapter 5 has provided clearer evidence on how such ontological categories can be combined within a single study. On the other hand, thick data are collected by observing the world from the point of view of the actors. Consequently, they can be used to examine phenomena from an espistemological perspective that is different from the one used in general in data-driven studies, where the general aim is to maintain the distance from the object under study.

9.3 Big data and evidence of mechanisms

In the third part of the thesis, I have proposed two discussions to overcome the third challenge (how to use big data, in the social sciences, to obtain evidence of causality that goes beyond correlations). The first discussion, offered in chapter 7, deals with the notion of constitutive mechanisms, according to which the constitutive entities and activities of a mechanism are those that actualise the capacity that makes the phenomenon-to-be-explained possible. Current research on constitutive mechanisms is mostly philosophical, and is focused on metaphysical questions that are investigated through formal discussions. I have examined such discussions and I have developed a new account starting from the existing ones with the aim to tackle some methodological challenges

that social scientists have to overcome when they study constitutive mechanisms. I have argued that, while in specific cases big data might help social scientists to use this account to collect evidence of constitutive mechanisms, in numerous cases this account is not feasible. The limitations of this and of any similar account, I have observed, are partly associated with the nature of social mechanisms. In several cases, indeed, social mechanisms are not taken to be unified wholes, but composite systems whose constituents are defined analytically and whose organisational levels are not clearly separated. Applying this account, furthermore, is made more difficult by the fact that it should be possible to intervene (through horizontally surgical interventions) on the mechanism and its constitutive entities. Such interventions, in the case of social mechanisms, are often unfeasible. Finally, I have observed that any attempt to improve the current accounts of constitutive mechanisms to provide a useful approach to scientists faces one main challenge: due to the distance between such discussions and scientific practices, it is difficult to clarify how social scientists collect, or should collect, evidence of constitutive mechanisms. These observations have led me to conclude that a more scientifically-informed discussion might provide more insights into how social scientists can obtain evidence of mechanisms.

The second answer to the third question is more scientifically-informed and has been developed by analysing a real-life research programme, LIFEPATH, which has been discussed in chapter 6. The analysis of how LIFEPATH researchers use biological markers to trace etiological mechanisms has been used as a theoretical basis to develop the concept of social markers, that, according to my proposal, can be used to trace social processes. I have suggested that sociomarkers, like biomarkers, can help researchers to pick up signals of social processes or of social-to-biological processes. I have finally argued that the notion of sociomarkers is useful to clarify how big data can be used in the social sciences. In addition to being used to obtain indicators or proxies, big data can be used to identify markers, like in the case of the identification of sociomarkers from sociometric badges' data.

The challenges associated with causal studies based on big data are at the heart of several discussions both in philosophy and in the social sciences. To tackle them, it is often essential to work both on conceptual problems and on methodological issues associated with scientific practices. With my thesis, I hope to pave the way for further interdisciplinary discussions on the use of big and thick data to study causal phenomena.

Bibliography

Adler, N. E., & Ostrove, J. M. (1999). Socioeconomic Status and Health: What We Know and What We Don't. *Annals of the New York Academy of Sciences, 896*(1), 3–15. https://doi.org/10.1111/j.1749-6632.1999.tb08101.x

Albers, C. J. (2015). Dutch research funding, gender bias, and Simpson's paradox. *Proceedings of the National Academy of Sciences*, *112*(50), E6828–E6829. https://doi.org/10.1073/pnas.1518936112

Anda, R. F., Felitti, V. J., Bremner, J. D., Walker, J. D., Whitfield, Ch., Perry, B. D., … Giles, W. H. (2006). The enduring effects of abuse and related adverse experiences in childhood: A convergence of evidence from neurobiology and epidemiology. *European Archives of Psychiatry and Clinical Neuroscience, 256*(3), 174–186. https://doi.org/10.1007/s00406-005-0624-4

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/

Ankeny, R. A., & Leonelli, S. (2016). Repertoires: A post-Kuhnian perspective on scientific change and collaborative research. *Studies in History and Philosophy of Science Part A, 60*, 18–28. https://doi.org/10.1016/j.shpsa.2016.08.003

Armitage, E. G., & Barbas, C. (2014). Metabolomics in cancer biomarker discovery: Current trends and future perspectives. *Journal of Pharmaceutical and Biomedical Analysis, 87*, 1–11. https://doi.org/10.1016/j.jpba.2013.08.041

Ashworth, S., Berry, C. R., & Bueno de Mesquita, E. (2015). All Else Equal in Theory and Data (Big or Small). *PS: Political Science & Politics*, *48*(01), 89–94. https://doi.org/10.1017/S1049096514001802

Baggerly, K. A., Morris, J. S., & Coombes, K. R. (2004). Reproducibility of SELDI-TOF protein patterns in serum: Comparing datasets from different experiments. *Bioinformatics*, *20*(5), 777–785. https://doi.org/10.1093/bioinformatics/btg484

Barboza Solís, C., Fantin, R., Castagné, R., Lang, T., Delpierre, C., & Kelly-Irving, M. (2016). Mediating pathways between parental socio-economic position and allostatic load in mid-life: Findings from the 1958 British birth cohort. *Social Science & Medicine*, *165*, 19–27. https://doi.org/10.1016/j.socscimed.2016.07.031

Barboza Solís, C., Kelly-Irving, M., Fantin, R., Darnaudéry, M., Torrisani, J., Lang, T., & Delpierre, C. (2015). Adverse childhood experiences and physiological wear-and-tear in midlife: Findings from the 1958 British birth cohort. *Proceedings of the National Academy of Sciences*, *112*(7), E738–E746. https://doi.org/10.1073/pnas.1417325112

Barnes, J., & Weller, N. (2017). Case Studies and Analytic Transparency in Causal-Oriented Mixed-Methods Research. *PS: Political Science & Politics*, *50*(04), 1019–1022. https://doi.org/10.1017/S1049096517001202

Bassanini, A., & Duval, R. (2006). *Employment patterns in OECD countries: Reassessing the role of policies and institutions* (No. 42; pp. 7–86). OECD.

Baumgartner, M. (2013). Rendering Interventionism and Non-Reductive Physicalism Compatible: Rendering Interventionism and Non-Reductive Physicalism Compatible. *Dialectica*, *67*(1), 1–27. https://doi.org/10.1111/1746-8361.12008

Baumgartner, M., & Casini, L. (2017). An Abductive Theory of Constitution. *Philosophy of Science*, *84*(2), 214–233. https://doi.org/10.1086/690716

Baumgartner, M., Casini, L., & Krickel, B. (2018). Horizontal Surgicality and Mechanistic Constitution. *Erkenntnis*. https://doi.org/10.1007/s10670-018-0033-5

Baumgartner, M., & Gebharter, A. (2016). Constitutive Relevance, Mutual Manipulability, and Fat-Handedness. *The British Journal for the Philosophy of Science*, *67*(3), 731–756. https://doi.org/10.1093/bjps/axv003

Beach, D., & Pedersen, R. (2013). *Process-Tracing Methods: Foundations and Guidelines*. https://doi.org/10.3998/mpub.2556282

Becker, M. C., Lazaric, N., Nelson, R. R., & Winter, S. G. (2005). Applying organizational routines in understanding organizational change. *Industrial and Corporate Change*, *14*(5), 775–791. https://doi.org/10.1093/icc/dth071

Bennett, A., & Checkel, J. T. (Eds.). (2015). *Process tracing: From metaphor to analytic tool*. Cambridge ; New York: Cambridge University Press.

Bickel, P. J., Hammel, E. A., & O'Connell, J. W. (1975). Sex Bias in Graduate Admissions: Data from Berkeley. *Science*, *187*(4175), 398–404. https://doi.org/10.1126/science.187.4175.398

Biddle, C., & Schafft, K. A. (2015). Axiology and Anomaly in the Practice of Mixed Methods Work: Pragmatism, Valuation, and the Transformative Paradigm. *Journal of Mixed Methods Research*, *9*(4), 320–334. https://doi.org/10.1177/1558689814533157

Biomarkers Definitions Working Group.. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology & Therapeutics*, *69*(3), 89–95. https://doi.org/10.1067/mcp.2001.113989

Blok, A., & Pedersen, M. A. (2014). Complementary social science? Quali-quantitative experiments in a Big Data world. *Big Data & Society*, *1*(2), 205395171454390. https://doi.org/10.1177/2053951714543908

Booth, A., Edwards, J. N., & Johnson, D. R. (1991). Social Integration and Divorce. *Social Forces*, *70*(1), 207. https://doi.org/10.2307/2580069

Borgman, C. L. (2016). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA; London: The MIT Press.

Bornakke, T., & Due, B. L. (2018). Big–Thick Blending: A method for mixing analytical insights from big and thick data sources. *Big Data & Society*, *5*(1), 205395171876502. https://doi.org/10.1177/2053951718765026

Bothwell, L. E., Greene, J. A., Podolsky, S. H., & Jones, D. S. (2016). Assessing the Gold Standard — Lessons from the History of RCTs. *New England Journal of Medicine*, *374*(22), 2175–2181. https://doi.org/10.1056/NEJMms1604593

Boumans, M. (2015). *Science outside the laboratory: Measurement in field science and economics*. New York, NY: Oxford University Press.

Boumans, M., & Leonelli, S. (2013). Introduction: On the Philosophy of Science in Practice. *Journal for General Philosophy of Science*, *44*(2), 259–261. https://doi.org/10.1007/s10838-013-9232-6

Brady, H., & Collier, D. (Eds.). (2004). *Rethinking social inquiry: Diverse tools, shared standards*. Lanham, Md: Rowman & Littlefield.

Braveman, P., & Gottlieb, L. (2014). The Social Determinants of Health: It's Time to Consider the Causes of the Causes. *Public Health Reports*, *129*(1_suppl2), 19–31. https://doi.org/10.1177/00333549141291S206

Buhaug, H., & Gates, S. (2002). The Geography of Civil War. *Journal of Peace Research*, *39*(4), 417–433. https://doi.org/10.1177/0022343302039004003

Bühlmann, P., & Geer, S. A. (2011). *Statistics for high-dimensional data: Methods, theory and applications*. Heidelberg ; New York: Springer.

Butler, D. (2013). When Google got flu wrong. *Nature*, *494*(7436), 155–156. https://doi.org/10.1038/494155a

Cain, L. D. Jr. (1964). Life Course and Social Structure. In F. Rel (Ed.), *Handbook of Modern Sociology* (pp. 272–309). Chicago: Rand McNally.

Canali, S. (2016). Big Data, epistemology and causality: Knowledge in and knowledge out in EXPOsOMICS. *Big Data & Society*, *3*(2), 205395171666953. https://doi.org/10.1177/2053951716669530

Carrington, W. J., Detragiache, E., & Vishwanath, T. (1996). Migration with Endogenous Moving Costs. *The American Economic Review*, *86*(4), 909–930.

Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge ; New York: Cambridge University Press.

Castagné, R., Delpierre, C., Kelly-Irving, M., Campanella, G., Guida, F., Krogh, V., … Chadeau-Hyam, M. (2016). A life course approach to explore the biological embedding of socioeconomic position and social mobility through circulating inflammatory markers. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep25170

Castles, S. (2000). International Migration at the Beginining of the Twenty-First Century: Global Trends and Issues. *International Social Science Journal*, *52*(165), 269–281. https://doi.org/10.1111/1468-2451.00258

Chadeau-Hyam, M., Athersuch, T. J., Keun, H. C., De Iorio, M., Ebbels, T. M. D., Jenab, M., … Vineis, P. (2011). Meeting-in-the-middle using metabolic profiling – a strategy for the identification of intermediate biomarkers in cohort

studies. *Biomarkers*, *16*(1), 83–88. https://doi.org/10.3109/1354750X.2010.533285

Chang, H. (2014). *Is water H_2O? evidence, realism and pluralism*. Cambridge, England ; New York: Springer.

Check, E. (2004). Proteomics and cancer: Running before we can walk? *Nature*, *429*(6991), 496–497. https://doi.org/10.1038/429496a

Chen, X., & Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, *108*(21), 8589–8594. https://doi.org/10.1073/pnas.1017031108

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., … Cutler, D. (2016). The Association Between Income and Life Expectancy in the United States, 2001-2014. *JAMA*, *315*(16), 1750. https://doi.org/10.1001/jama.2016.4226

Chorpita, B. F., Weisz, J. R., Daleiden, E. L., Schoenwald, S. K., Palinkas, L. A., Miranda, J., … Research Network on Youth Mental Health. (2013). Long-term outcomes for the Child STEPs randomized effectiveness trial: A comparison of modular and standard treatment designs with usual care. *Journal of Consulting and Clinical Psychology*, *81*(6), 999–1009. https://doi.org/10.1037/a0034200

Christakis, N. A., & Fowler, J. H. (2008). The Collective Dynamics of Smoking in a Large Social Network. *New England Journal of Medicine*, *358*(21), 2249–2258. https://doi.org/10.1056/NEJMsa0706154

Church, J., & King, I. (1993). Bilingualism and Network Externalities. *The Canadian Journal of Economics*, *26*(2), 337. https://doi.org/10.2307/135911

Clark, W. R., & Golder, M. (2015). Big Data, Causal Inference, and Formal Theory: Contradictory Trends in Political Science? *PS: Political Science & Politics*, *48*(01), 65–70. https://doi.org/10.1017/S1049096514001759

Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the Evidence Hierarchy. *Topoi*, *33*(2), 339–360. https://doi.org/10.1007/s11245-013-9220-9

Claveau, F. (2012). The Russo–Williamson Theses in the social sciences: Causal inference drawing on two types of evidence. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(4), 806–813. https://doi.org/10.1016/j.shpsc.2012.05.004

Cohen, J. H. (2004). *The culture of migration in Southern Mexico* (1st ed). Austin: University of Texas Press.

Collier, P., & Hoeffler, A. (1998). On Economic Causes of Civil War. *Oxford Economic Papers*, *50*(4), 563–573.

Constantinou, A., & Fenton, N. (2018). Things to know about Bayesian networks: Decisions under uncertainty, part 2. *Significance*, *15*(2), 19–23. https://doi.org/10.1111/j.1740-9713.2018.01126.x

Cook, S., Conrad, C., Fowlkes, A. L., & Mohebbi, M. H. (2011). Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic. *PLoS ONE*, *6*(8), e23610. https://doi.org/10.1371/journal.pone.0023610

Cornelius, W. A. (2001). Death at the Border: Efficacy and Unintended Consequences of US Immigration Control Policy. *Population and Development Review*, *27*(4), 661–685. https://doi.org/10.1111/j.1728-4457.2001.00661.x

Cortez, V., & Rica, S. (2015). Dynamics of the Schelling Social Segregation Model in Networks. *Procedia Computer Science*, *61*, 60–65. https://doi.org/10.1016/j.procs.2015.09.148

Craver, C. F. (2006). When mechanistic models explain. *Synthese*, *153*(3), 355–376. https://doi.org/10.1007/s11229-006-9097-x

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford : New York : Oxford University Press: Clarendon Press.

Creswell, J. W. (1994). *Research design: Qualitative & quantitative approaches*. Thousand Oaks, Calif: Sage Publications.

Crick, F. H. C., Barnett, L., Brenner, S., & Watts-Tobin, R. J. (1961). General Nature of the Genetic Code for Proteins. *Nature*, *192*(4809), 1227–1232. https://doi.org/10.1038/1921227a0

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*(2), 116–127.

Culotta, A. (2014). Reducing sampling bias in social media data for county health inference. *Joint Statistical Meetings Proceedings*.

Cunningham, P. (2008). Dimension Reduction. In M. Cord & P. Cunningham (Eds.), *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval* (pp. 91–112). https://doi.org/10.1007/978-3-540-75171-7_4

Curran, J. (2013). Big Data or 'Big Ethnographic Data'? Positioning Big Data within the ethnographic space. *Ethnographic Praxis in Industry Conference Proceedings*, *2013*(1), 62–73. https://doi.org/10.1111/j.1559-8918.2013.00006.x

Curran, S. R., Garip, F., Chung, C. Y., & Tangchonlatip, K. (2005). Gendered Migrant Social Capital: Evidence from Thailand. *Social Forces*, *84*(1), 225–255. https://doi.org/10.1353/sof.2005.0094

Cyr, J., & Mahoney, J. (2016). The enduring influence of historical-structural approaches. In P. R. Kingstone & D. J. Yashar (Eds.), *Routledge handbook of Latin American politics* (pp. 433–446). New York London: Routledge.

d'Errico, A., Ricceri, F., Stringhini, S., Carmeli, C., Kivimaki, M., Bartley, M., … LIFEPATH Consortium. (2017). Socioeconomic indicators in epidemiologic research: A practical example from the LIFEPATH study. *PLOS ONE*, *12*(5), e0178071. https://doi.org/10.1371/journal.pone.0178071

Dalkin, S. M., Greenhalgh, J., Jones, D., Cunningham, B., & Lhussier, M. (2015). What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science*, *10*(1). https://doi.org/10.1186/s13012-015-0237-x

Danese, A., & McEwen, B. S. (2012). Adverse childhood experiences, allostasis, allostatic load, and age-related disease. *Physiology & Behavior*, *106*(1), 29–39. https://doi.org/10.1016/j.physbeh.2011.08.019

Dannefer, D. (2003). Cumulative Advantage/Disadvantage and the Life Course: Cross-Fertilizing Age and Social Science Theory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *58*(6), S327–S337. https://doi.org/10.1093/geronb/58.6.S327

Danzig, M. E. (1976). GLEN H. ELDER, JR. Children of the Great Depression: Social Change of Life Experience. Pp. xxiii, 400. Chicago, Ill.: The University of Chicago Press, 1974. $15.00. *The ANNALS of the American Academy of Political and Social Science*, *425*(1), 171–172. https://doi.org/10.1177/000271627642500136

Davillas, A., Benzeval, M., & Kumari, M. (2017). Socio-economic inequalities in C-reactive protein and fibrinogen across the adult age span: Findings from

Understanding Society. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-02888-6

De Veaux, R. D., & Hand, D. J. (2005). How to Lie with Bad Data. *Statistical Science*, *20*(3), 231–238. https://doi.org/10.1214/088342305000000269

Debnath, M., Bisen, P., & Prasad, G. B. K. S. (2010). *Molecular Diagnostics Promises and Possibilities*. Retrieved from http://dx.doi.org/10.1007/978-90-481-3261-4

Denscombe, M. (2008). Communities of Practice: A Research Paradigm for the Mixed Methods Approach. *Journal of Mixed Methods Research*, *2*(3), 270–283. https://doi.org/10.1177/1558689808316807

di Bella, E., Leporatti, L., & Maggino, F. (2018). Big Data and Social Indicators: Actual Trends and New Perspectives. *Social Indicators Research*, *135*(3), 869–878. https://doi.org/10.1007/s11205-016-1495-y

DiMaggio, P., & Garip, F. (2012). Network Effects and Social Inequality. *Annual Review of Sociology*, *38*(1), 93–118. https://doi.org/10.1146/annurev.soc.012809.102545

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments. *Annual Review of Sociology*, *32*(1), 271–297. https://doi.org/10.1146/annurev.soc.32.061604.123127

Dowe, D. (2007). *Physical causation*. Cambridge: Cambridge University Press.

Dowe, P. (1992). Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory. *Philosophy of Science*, *59*(2), 195–216. https://doi.org/10.1086/289662

Eells, E. (1991). *Probabilistic Causality*. Cambridge: Cambridge University Press.

Egner, P. A., Wang, J.-B., Zhu, Y.-R., Zhang, B.-C., Wu, Y., Zhang, Q.-N., … Kensler, T. W. (2001). Chlorophyllin intervention reduces aflatoxin-DNA adducts in individuals at high risk for liver cancer. *Proceedings of the National Academy of Sciences*, *98*(25), 14601–14606. https://doi.org/10.1073/pnas.251536898

Elder, G. H. (1998). The Life Course as Developmental Theory. *Child Development*, *69*(1), 1. https://doi.org/10.2307/1132065

Elsenbroich, C. (2012). Explanation in Agent-Based Modelling: Functions, Causality or Mechanisms? *Journal of Artificial Societies and Social Simulation*, *15*(3). https://doi.org/10.18564/jasss.1958

Elster, J. (1989). *Nuts and bolts for the social sciences*. Cambridge ; New York: Cambridge University Press.

Elster, J. (1993). *Political psychology*. Cambridge [England] ; New York, NY, USA: Cambridge University Press.

Falleti, T. G., & Lynch, J. F. (2009). Context and Causal Mechanisms in Political Analysis. *Comparative Political Studies*, *42*(9), 1143–1166. https://doi.org/10.1177/0010414009331724

Fearon, J. D., & Laitin, D. D. (2008). Integrating Qualitative and Quantitative Methods. In H. E. Brady & D. Collier (Eds.), *The Oxford Handbook of Political Methodology* (pp. 756–776). Oxford, United Kingdom: Oxford University Press.

Feldman, M. S. (2000). Organizational Routines as a Source of Continuous Change. *Organization Science*, *11*(6), 611–629. https://doi.org/10.1287/orsc.11.6.611.12529

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., … Marks, J. S. (1998). Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults. *American*

Journal of Preventive Medicine, *14*(4), 245–258. https://doi.org/10.1016/S0749-3797(98)00017-8

Ferraro, K. F:, Schafer, M. H., & Wilkinson, L. R. (2016). Childhood Disadvantage and Health Problems in Middle and Later Life: Early Imprints on Physical Health? *American Sociological Review*, *81*(1), 107–133. https://doi.org/10.1177/0003122415619617

Ferraro, K. F., & Shippee, T. P. (2009). Aging and Cumulative Inequality: How Does Inequality Get Under the Skin? *The Gerontologist*, *49*(3), 333–343. https://doi.org/10.1093/geront/gnp034

Fleming, T. R., & DeMets, D. L. (1996). Surrogate end points in clinical trials: are we being misled? *Annals of Internal Medicine*, *125*(7), 605–613.

Floridi, L. (2012). Big Data and Their Epistemological Challenge. *Philosophy & Technology*, *25*(4), 435–437. https://doi.org/10.1007/s13347-012-0093-4

Ford, H. (2016). The Person in the (Big) Data: A Selection of Innovative Methods, Strategies and Perspectives for Social Research in the Age of (Big) Data. *Working Papers of the Communities & Culture Network+*.

Ford, J. K., Schmitt, N., Schechtman, S. L., Hults, B. M., & Doherty, M. L. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes*, *43*(1), 75–117. https://doi.org/10.1016/0749-5978(89)90059-9

Franklin-Hall, L. R. (2015). Explaining Causal Selection with Explanatory Causal Economy: Biology and Beyond. In P. Braillard & C. Malaterre, *Explanation in Biology* (Vol. 11, pp. 413–438). https://doi.org/10.1007/978-94-017-9822-8_18

Fredriksson, P., & Holmlund, B. (2006). Improving Incentives in Unemployment Insurance: A Review of Recent Research. *Journal of Economic Surveys*, *20*(3), 357–386. https://doi.org/10.1111/j.0950-0804.2006.00283.x

Friedrichs, J. (2016). Causal Mechanisms and Process Patterns in International Relations: Thinking Within and Without the box. *St Antony's International Review*, *12*(1), 76–89.

Fushiki, T., Fujisawa, H., & Eguchi, S. (2006). Identification of biomarkers from mass spectrometry data using a "common" peak approach. *BMC Bioinformatics*, *7*(1). https://doi.org/10.1186/1471-2105-7-358

Fussell, E., & Massey, D. S. (2004). The Limits to Cumulative Causation: International Migration From Mexican Urban Areas. *Demography*, *41*(1), 151–171. https://doi.org/10.1353/dem.2004.0003

Galea, S., & Link, B. G. (2013). Six Paths for the Future of Social Epidemiology. *American Journal of Epidemiology*, *178*(6), 843–849. https://doi.org/10.1093/aje/kwt148

Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., & Davey Smith, G. (2006). Indicators of socioeconomic position (part 1). *Journal of Epidemiology & Community Health*, *60*(1), 7–12. https://doi.org/10.1136/jech.2004.023531

Garcia, P. R., & Jutila, S. T. (1988). Socio-economic Stratification Generated by International Migration Loops. *International Migration*, *26*(1), 57–69. https://doi.org/10.1111/j.1468-2435.1988.tb00611.x

Garip, F., & Asad, L. (2016). Network Effects in Mexico–U.S. Migration: Disentangling the Underlying Social Mechanisms. *American Behavioral Scientist*, *60*(10), 1168–1193. https://doi.org/10.1177/0002764216643131

Garni, A. (2010). Mechanisms of Migration: Poverty and Social Instability in the Postwar Expansion of Central American Migration to the United States. *Journal of Immigrant & Refugee Studies*, *8*(3), 316–338. https://doi.org/10.1080/15562948.2010.501295

Gayo-Avello, D. (2011). Don't turn social media into another "Literary Digest" poll. *Communications of the ACM*, *54*(10), 121. https://doi.org/10.1145/2001269.2001297

Geertz, C. (1973). *The interpretation of cultures: Selected essays*. New York: Basic Books.

George, A. L., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. Cambridge, Mass: MIT Press.

Gerrets, R. (2015). Charting the road to eradication: Health facility data and malaria indicator generation in rural Tanzania. In R. Rottenburg, S. E. Merry, S. Park, & J. Mugler (Eds.), *The World of Indicators* (pp. 151–187). https://doi.org/10.1017/CBO9781316091265.007

Gerring, J. (2006). *Case Study Research: Principles and Practices*. https://doi.org/10.1017/CBO9780511803123

Geyer, S. (2006). Education, income, and occupational class cannot be used interchangeably in social epidemiology. Empirical evidence against a common practice. *Journal of Epidemiology & Community Health*, *60*(9), 804–810. https://doi.org/10.1136/jech.2005.041319

Ghiara, V. (2016). *Mind the gap: from data to big data in the economic domain* (Erasmus University Rotterdam). Retrieved from file:///C:/Users/Virginia/Downloads/Master-Virginia-Ghiara-43784820%20(4).pdf

Ghiara, V. (2019). Disambiguating the Role of Paradigms in Mixed Methods Research. *Journal of Mixed Methods Research*, 155868981881992. https://doi.org/10.1177/1558689818819928

Ghiara, V., & Russo, F. (2019). Reconstructing the mixed mechanisms of health: the role of bio- and sociomarkers. *Longitudinal and Life Course Studies*, *10*(1), 7–25. https://doi.org/10.1332/175795919X15468755933353

Giczi, J., & Szőke, K. (2018). Official Statistics and Big Data. *Intersections*, *4*(1). https://doi.org/10.17356/ieejsp.v4i1.408

Giere, R. N. (2006). *Scientific perspectivism*. Chicago: University of Chicago Press.

Gillies, D. (2010). The Russo–Williamson thesis and the question of whether smoking causes heart disease. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 110–125). Oxford: Oxford University Press.

Glennan, S. (2002). Rethinking Mechanistic Explanation. *Philosophy of Science*, *69*(S3), S342–S353. https://doi.org/10.1086/341857

Gogoshin, G., Boerwinkle, E., & Rodin, A. S. (2017). New Algorithm and Software (BNOmics) for Inferring and Visualizing Bayesian Networks from Heterogeneous Big Biological and Genetic Data. *Journal of Computational Biology*, *24*(4), 340–356. https://doi.org/10.1089/cmb.2016.0100

Goldring, L. (2004). Family and Collective Remittances to Mexico: A Multi-dimensional Typology. *Development and Change*, *35*(4), 799–840. https://doi.org/10.1111/j.0012-155X.2004.00380.x

Goldthorpe, J. H. (2016). *Sociology as a population science*.

Greene, J. C. (2006). Toward a methodology of mixed methods social inquiry. *Research in the Schools*, *13*(1), 93–99.

Guba, E. G., & Lincoln, Y. S. (1994). *Competing paradigms in qualitative research* (N. K. Denzin & Y. S. Lincoln, Eds.). Thousand Oaks, CA: SAGE.

Guo, H. (2017). Big Earth data: A new frontier in Earth and information sciences. *Big Earth Data*, *1*(1–2), 4–20. https://doi.org/10.1080/20964471.2017.1403062

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, *3*, 1157–1182.

Halfon, N., Larson, K., Son, J., Lu, M., & Bethell, C. (2017). Income Inequality and the Differential Effect of Adverse Childhood Experiences in US Children. *Academic Pediatrics*, *17*(7), S70–S78. https://doi.org/10.1016/j.acap.2016.11.007

Harinen, T. (2014). Mutual manipulability and causal inbetweenness. *Synthese*, *195*(1), 35–54. https://doi.org/10.1007/s11229-014-0564-5

Heckerman, D., Geiger, D., & Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, *20*(3), 197–243. https://doi.org/10.1007/BF00994016

Henderson, R. F., Bechtold, W. E., Bond, J. A., & Sun, J. D. (1989). The Use of Biological Markers in Toxicology. *Critical Reviews in Toxicology*, *20*(2), 65–82. https://doi.org/10.3109/10408448909017904

Herrmann-Pillath, C. (2015). Constitutive Explanations as a Methodological Framework for Integrating Thermodynamics and Economics. *Entropy*, *18*(1), 18. https://doi.org/10.3390/e18010018

Hesslow, G. (1976). Two Notes on the Probabilistic Approach to Causality. *Philosophy of Science*, *43*(2), 290–292. https://doi.org/10.1086/288684

Hitchcock, C. (2009). Causal Modelling. In H. Beebee & C. Hitchcock (Eds.), *The Oxford handbook of causation* (Oxford University Press, pp. 299–314). Oxford, United Kingdom.

Hitchcock, C., & Sober, E. (2004). Prediction Versus Accommodation and the Risk of Overfitting. *The British Journal for the Philosophy of Science*, *55*(1), 1–34. https://doi.org/10.1093/bjps/55.1.1

Hobbs, G., Vignoles, A., & Centre for the Economics of Education (Great Britain). (2007). *Is free school meal status a valid proxy for socio-economic status (in schools research)?* Retrieved from http://cee.lse.ac.uk/cee%20dps/ceedp84.pdf

Horgan, R. P., & Kenny, L. C. (2011). 'Omic' technologies: Genomics, transcriptomics, proteomics and metabolomics: The Obstetrician & Gynaecologist. *The Obstetrician & Gynaecologist*, *13*(3), 189–195. https://doi.org/10.1576/toag.13.3.189.27672

House, J. S., Lantz, P. M., & Herd, P. (2005). Continuity and Change in the Social Stratification of Aging and Health Over the Life Course: Evidence From a Nationally Representative Longitudinal Study From 1986 to 2001/2002 (Americans' Changing Lives Study). *The Journals of Gerontology: Series B*, *60*(Special_Issue_2), S15–S26. https://doi.org/10.1093/geronb/60.Special_Issue_2.S15

Howe, L. D., Galobardes, B., Matijasevich, A., Gordon, D., Johnston, D., Onwujekwe, O., … Hargreaves, J. R. (2012). Measuring socio-economic position for epidemiological studies in low- and middle-income countries: A methods of measurement in epidemiology paper. *International Journal of Epidemiology*, *41*(3), 871–886. https://doi.org/10.1093/ije/dys037

Hox, J. (2017). Computational Social Science Methodology, Anyone? *Methodology*, *13*(Supplement 1), 3–12. https://doi.org/10.1027/1614-2241/a000127

Humbert, A. L., & Guenther, E. A. (2017). *The Gender Diversity Index, preliminary considerations and results* (p. 31) [Project Deliverable]. Retrieved from https://www.gedii.eu/wp-content/uploads/D3.1GenderDiversityIndex_final.pdf

Hunter, A., & Brewer, J. (2003). Multimethod research in sociology. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods in Social and Behavioral Research* (pp. 577–594). Thousand Oaks, CA: SAGE.

Illari, P. M. (2011). Mechanistic Evidence: Disambiguating the Russo–Williamson Thesis. *International Studies in the Philosophy of Science*, *25*(2), 139–157. https://doi.org/10.1080/02698595.2011.574856

Illari, P., & Russo, F. (2014). *Causality: Philosophical theory meets scientific practice* (First edition). Oxford, United Kingdom: Oxford University Press.

Illari, P., & Russo, F. (2016). Information Channels and Biomarkers of Disease. *Topoi*, *35*(1), 175–190. https://doi.org/10.1007/s11245-013-9228-1

International Programme on Chemical Safety (Ed.). (2001). *Biomarkers in risk assessment: Validity and validation*. Geneva: World Health Organization.

Iwashyna, T. J., & Liu, V. (2014). What's So Different about Big Data?. A Primer for Clinicians Trained to Think Epidemiologically. *Annals of the American Thoracic Society*, *11*(7), 1130–1135. https://doi.org/10.1513/AnnalsATS.201405-185AS

J. Hunter, D., Losina, E., Guermazi, A., Burstein, D., N. Lassere, M., & Kraus, V. (2010). A Pathway and Approach to Biomarker Validation and Qualification for Osteoarthritis Clinical Trials. *Current Drug Targets*, *11*(5), 536–545. https://doi.org/10.2174/138945010791011947

Jain, K. K. (2015). *The handbook of biomarkers*.

Jensen, E. A. (2017). Putting the methodological brakes on claims to measure national happiness through Twitter: Methodological limitations in social media analytics. *PLOS ONE*, *12*(9), e0180080. https://doi.org/10.1371/journal.pone.0180080

Jensen, F. V., & Nielsen, T. D. (2007). *Bayesian Networks and Decision Graphs*. https://doi.org/10.1007/978-0-387-68282-2

Johnson, R. B. (2012). Dialectical Pluralism and Mixed Research. *American Behavioral Scientist*, *56*(6), 751–754. https://doi.org/10.1177/0002764212442494

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, *33*(7), 14–26. https://doi.org/10.3102/0013189X033007014

Johnson, R. B., Russo, F., & Schoonenboom, J. (2017). Causation in Mixed Methods Research: The Meeting of Philosophy, Science, and Practice. *Journal of Mixed Methods Research*, 155868981771961. https://doi.org/10.1177/1558689817719610

Juster, R. P., McEwen, B. S., & Lupien, S. J. (2010). Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience & Biobehavioral Reviews*, *35*(1), 2–16. https://doi.org/10.1016/j.neubiorev.2009.10.002

Kaiser, M. I., & Krickel, B. (2016). The Metaphysics of Constitutive Mechanistic Phenomena. *The British Journal for the Philosophy of Science*, axv058. https://doi.org/10.1093/bjps/axv058

Kandel, E. R. (2006). *In search of memory: The emergence of a new science of mind* (1st ed). New York: W. W. Norton & Company.

Kandel, W., & Kao, G. (2006). The Impact of Temporary Labor Migration on Mexican Children's Educational Aspirations and Performance1. *International Migration Review*, *35*(4), 1205–1231. https://doi.org/10.1111/j.1747-7379.2001.tb00058.x

Karasz, A., & Singelis, T. M. (2009). Qualitative and Mixed Methods Research in Cross-Cultural Psychology. *Journal of Cross-Cultural Psychology*, *40*(6), 909–916. https://doi.org/10.1177/0022022109349172

Katz, R. (2004). Biomarkers and surrogate markers: An FDA perspective. *NeuroRX*, *1*(2), 189–195. https://doi.org/10.1602/neurorx.1.2.189

Kauffman, S. (1971). Gene Regulation Networks: A Theory For Their Global Structure and Behaviors. In *Current Topics in Developmental Biology* (Vol. 6, pp. 145–182). https://doi.org/10.1016/S0070-2153(08)60640-7

Kelly, M. P., Kelly, R. S., & Russo, F. (2014). The Integration of Social, Behavioral, and Biological Mechanisms in Models of Pathogenesis. *Perspectives in Biology and Medicine*, *57*(3), 308–328. https://doi.org/10.1353/pbm.2014.0026

Kelly, M. P., & Russo, F. (2017). Causal narratives in public health: The difference between mechanisms of aetiology and mechanisms of prevention in non-communicable diseases. *Sociology of Health & Illness*, *40*(1), 82–99. https://doi.org/10.1111/1467-9566.12621

Kelly-Irving, M., Lepage, B., Dedieu, D., Lacey, R., Cable, N., Bartley, M., … Delpierre, C. (2013). Childhood adversity as a risk for cancer: Findings from the 1958 British birth cohort study. *BMC Public Health*, *13*(1). https://doi.org/10.1186/1471-2458-13-767

Khan, M. A., Uddin, M. F., & Gupta, N. (2014). Seven V's of Big Data understanding Big Data to extract value. *Proceedings of the 2014 Zone 1 Conference of the*

American Society for Engineering Education, 1–5. https://doi.org/10.1109/ASEEZone1.2014.6820689

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00513

Kitchin, R. (2014a). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 205395171452848. https://doi.org/10.1177/2053951714528481

Kitchin, R. (2014b). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1–14. https://doi.org/10.1007/s10708-013-9516-8

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), 205395171663113. https://doi.org/10.1177/2053951716631130

Kohler, H., Behrman, J. R., & Watkins, S. C. (2001). The Density of Social Networks and Fertility Decisions: Evidence From South Nyanza District, Kenya. *Demography*, *38*(1), 43–58. https://doi.org/10.1353/dem.2001.0005

Korb, K. B., & Nicholson, A. E. (2004). *Bayesian artificial intelligence*. Boca Raton: Chapman & Hall/CRC.

Krickel, B. (2018). Saving the mutual manipulability account of constitutive relevance. *Studies in History and Philosophy of Science Part A*, *68*, 58–67. https://doi.org/10.1016/j.shpsa.2018.01.003

Krieger, N. (2001). A glossary for social epidemiology. *Journal of Epidemiology & Community Health*, *55*(10), 693–700. https://doi.org/10.1136/jech.55.10.693

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions (second edition enlarged)*. Chicago: University of Chicago Press.

Laney, D. (2001). *3D data management: Controlling data volume, velocity and variety. Technical report*. META Group.

Latzko-Toth, G., Bonneau, C., & Millette, M. (2016). Small Data, Thick Data: Thickening Strategies for Trace-based Social Media Research. In L. Sloan & A. Quan-Haase (Eds.), *The SAGE Handbook of Social Media Research Methods* (pp. 199–213). London: SAGE.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., … Van Alstyne, M. (2009). SOCIAL SCIENCE: Computational Social Science. *Science*, *323*(5915), 721–723. https://doi.org/10.1126/science.1167742

Leonelli, S. (2012a). Introduction: Making sense of data-driven research in the biological and biomedical sciences. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 1–3. https://doi.org/10.1016/j.shpsc.2011.10.001

Leonelli, S. (2012b). When humans are the exception: Cross-species databases at the interface of biological and clinical research. *Social Studies of Science*, *42*(2), 214–236. https://doi.org/10.1177/0306312711436265

Leonelli, S. (2014a). Data Interpretation in the Digital Age. *Perspectives on Science*, *22*(3), 397–417. https://doi.org/10.1162/POSC_a_00140

Leonelli, S. (2014b). What difference does quantity make? On the epistemology of Big Data in biology. *Big Data & Society*, *1*(1), 205395171453439. https://doi.org/10.1177/2053951714534395

Leonelli, S. (2015). What Counts as Scientific Data? A Relational Framework. *Philosophy of Science*, *82*(5), 810–821. https://doi.org/10.1086/684083

Lepri, B., Staiano, J., Rigato, G., Kalimeri, K., Finnerty, A., Pianesi, F., … Pentland, A. (2012, September). *The SocioMetric Badges Corpus: A Multilevel Behavioral Dataset for Social Behavior in Complex Organizations*. 623–628. https://doi.org/10.1109/SocialCom-PASSAT.2012.71

Lerman, K. (2018). Computational social scientist beware: Simpson's paradox in behavioral data. *Journal of Computational Social Science*, *1*(1), 49–58. https://doi.org/10.1007/s42001-017-0007-4

Leuridan, B. (2012). Three Problems for the Mutual Manipulability Account of Constitutive Relevance in Mechanisms. *The British Journal for the Philosophy of Science*, *63*(2), 399–427. https://doi.org/10.1093/bjps/axr036

Levins, R. (1970). Complex Systems. In C. H. Waddington (Ed.), *Towards a Theoretical Biology* (Vol. 3, pp. 73–88). Chicago: Aldine Publishing.

Lewis, D. (2004). Causation as influence. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 75–106). Cambridge, Massachusetts London, England: The MIT Press.

Libman, A. (2012). Learning to be different: Quantitative research in economics and political science. *Rationality, Markets and Morals*, *3*.

Lincoln, Y. S., & Guba, E. G. (2000). The Only Generalization Is: There is no Generalization. In R. Gomm, M. Hammersley, & P. Foster (Eds.), *Case Study Method* (pp. 27–44). London: SAGE.

Little, D. (2015). Mechanisms and Method. *Philosophy of the Social Sciences*, *45*(4–5), 462–480. https://doi.org/10.1177/0048393115580420

Little, D. (2018). Disaggregating Historical Explanation: The Move to Social Mechanisms in the Philosophy of History. In S. Glennan & P. Illari (Eds.), *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (pp. 413–422). New York: Routledge.

Liu, C., Jiang, J., Gu, J., Yu, Z., Wang, T., & Lu, H. (2016). High-dimensional omics data analysis using a variable screening protocol with prior knowledge integration (SKI). *BMC Systems Biology*, *10*(4).

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton, N.J: Princeton University Press.

Lynd, R. S., & Lynd, H. M. (1929). *Middletown: A Study in [contemporary] American Culture*. New York: Harcourt, Brace & Company.

Lynd, R. S., & Lynd, H. M. (1937). *Middletown in transition*. New York: Harcourt, Brace & Company.

Ma, W. (2017). *China's mobile economy: Opportunities in the largest and fastest information consumption boom*. Southern Gate, Chichester, West Sussex, UK: Wiley.

Ma, Y. Z., & Ma, M. M. (2011). Simpson's Paradox and other reversals in basketball: examples from the 2011 NBA playoffs. *International Journal of Sports Science and Engineering*, *5*(3), 145–154.

Macaulay, I. C., Carr, P., Gusnanto, A., Ouwehand, W. H., Fitzgerald, D., & Watkins, N. A. (2005). Platelet genomics and proteomics in human health and disease. *Journal of Clinical Investigation*, *115*(12), 3370–3377. https://doi.org/10.1172/JCI26885

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, *67*(1), 1–25. https://doi.org/10.1086/392759

Mackenbach, J. P., Stirbu, I., Roskam, A., Schaap, M. M., Menvielle, G., Leinsalu, M., & Kunst, A. E. (2008). Socioeconomic Inequalities in Health in 22 European Countries. *New England Journal of Medicine*, *358*(23), 2468–2481. https://doi.org/10.1056/NEJMsa0707519

Malik, M. M., Lamba, H., Nakos, C., & Pfeffer, J. (2015). Population Bias in Geotagged Tweets. *Standards and Practices in Large-Scale Social Media Research*. Presented at the 2015 ICWSM Workshop.

Malinowski, B. (1935). *Coral gardens and their magic*. New York: American Book Co.

Martin, J. F. (2017). *Exploring Service Learning Outcomes in Students: A Mixed Methods Study for Nursing*. ProQuest LLC. 789 East Eisenhower Parkway, P.O. Box 1346, Ann Arbor, MI 48106. Tel: 800-521-0600; Web site: http://www.proquest.com/en-US/products/dissertations/individuals.shtml.

Marwick, A. (2013). Ethnographic and Qualitative Research on Twitter. In K. Weller, A. Bruns, C. Puschmann, J. Burgess, & M. Mahrt (Eds.), *Twitter and Society* (pp. 109–122). New York: Peter Lang.

Marx, S. (2008). Rich Data. In L. Given, *The SAGE Encyclopedia of Qualitative Research Methods* (p. 795). https://doi.org/10.4135/9781412963909.n408

Massey, D. S. (2005). *Backfire at the Border: Why Enforcement without Legalization Cannot Stop Illegal Immigration*. Washington, D.C.: CATO Institute.

Massey, D. S., Goldring, L., & Durand, J. (1994). Continuities in Transnational Migration: An Analysis of Nineteen Mexican Communities. *American Journal of Sociology*, *99*(6), 1492–1533. https://doi.org/10.1086/230452

Massey, D. S., & Zenteno, R. (2000). A Validation of the Ethnosurvey: The Case of Mexico-U.S. Migration. *International Migration Review*, *34*(3), 766. https://doi.org/10.2307/2675944

Maxwell, J. A. (2004a). Causal Explanation, Qualitative Research, and Scientific Inquiry in Education. *Educational Researcher*, *33*(2), 3–11. https://doi.org/10.3102/0013189X033002003

Maxwell, J. A. (2004b). Using Qualitative Methods for Causal Explanation. *Field Methods*, *16*(3), 243–264. https://doi.org/10.1177/1525822X04266831

Maxwell, J. A. (2016). Expanding the History and Range of Mixed Methods Research. *Journal of Mixed Methods Research*, *10*(1), 12–27. https://doi.org/10.1177/1558689815571132

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work and think*. London: Murray.

Mayo, E. (1993). *The Human Problems of an Industrial Civilization*. Macmillan.

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. https://doi.org/10.15252/embr.201541001

McEwen, B. S., & Stellar, E. (1993). Stress and the individual. Mechanisms leading to disease. *Archives of Internal Medicine*, *153*(18), 2093–2101. https://doi.org/10.1001/archinte.153.18.2093

McGinnis, J. M. (2004). The Immediate vs the Important. *JAMA*, *291*(10), 1263. https://doi.org/10.1001/jama.291.10.1263

McGinnis, J. Michael, & Foege, W. H. (1993). Actual Causes of Death in the United States. *JAMA: The Journal of the American Medical Association*, *270*(18), 2207. https://doi.org/10.1001/jama.1993.03510180077038

Meek, C., & Glymour, C. (1994). Conditioning and Intervening. *The British Journal for the Philosophy of Science*, *45*(4), 1001–1021. https://doi.org/10.1093/bjps/45.4.1001

Merton, R. K. (1968). The Matthew Effect in Science: The reward and communication systems of science are considered. *Science*, *159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56

Milgram, S. (1974). *Obedience to authority: An experimental view* (Nachdr.). New York, N.Y.: HarperPerennial.

Mislove, A., Lehmann, S., Ahn, Y., Onnela, J., & Rosenquist, J. N. (2011). Understanding the demographics of twitter users Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, , and J. Niels Rosenquist. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM'11)*. Presented at the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM'11), Spain.

Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place. *PLoS ONE*, *8*(5), e64417. https://doi.org/10.1371/journal.pone.0064417

Moneta, A., & Russo, F. (2014). Causal models and evidential pluralism in econometrics. *Journal of Economic Methodology*, *21*(1), 54–76. https://doi.org/10.1080/1350178X.2014.886473

Monroe, B. L., Pan, J., Roberts, M. E., Sen, M., & Sinclair, B. (2015). No! Formal Theory, Causal Inference, and Big Data Are Not Contradictory Trends in Political Science. *PS: Political Science & Politics*, *48*(01), 71–74. https://doi.org/10.1017/S1049096514001760

Morgan, D. L. (2007). Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods. *Journal of*

Mixed Methods Research, *1*(1), 48–76. https://doi.org/10.1177/2345678906292462

Morgan, M. S. (2012). Case Studies: One Observation or Many? Justification or Discovery? *Philosophy of Science*, *79*(5), 667–677. https://doi.org/10.1086/667848

Neapolitan, R. E. (2004). *Learning Bayesian networks*. Upper Saddle River, NJ: Pearson Prentice Hall.

Nelson, R. R., & Winter, S. G. (1982). *An evolutionary theory of economic change* (digitally reprinted). Cambridge, Mass.: The Belknap Press of Harvard Univ. Press.

Noor, E., Eden, E., Milo, R., & Alon, U. (2010). Central Carbon Metabolism as a Minimal Biochemical Walk between Precursors for Biomass and Energy. *Molecular Cell*, *39*(5), 809–820. https://doi.org/10.1016/j.molcel.2010.08.031

Nussbaum, M. C. (2013). *Creating capabilities: The human development approach* (First Harvard University Press paperback edition). Cambridge, Massachusetts London, England: The Belknap Press of Harvard University Press.

Ogburn, W. F., & Nimkoff, M. F. (1964). *A Handbook of Sociology* (5th ed.). London: Routledge and K. Paul.

Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and Qualitative Research: An Oxymoron? *Quality & Quantity*, *41*(2), 233–249. https://doi.org/10.1007/s11135-006-9000-3

Palinkas, L. A. (2014). Causality and Causal Inference in Social Work: Quantitative and Qualitative Perspectives. *Research on Social Work Practice*, *24*(5), 540–547. https://doi.org/10.1177/1049731514536056

Palinkas, L. A., Weisz, J. R., Chorpita, B. F., Levine, B., Garland, A. F., Hoagwood, K. E., & Landsverk, J. (2013). Continued Use of Evidence-Based Treatments After a Randomized Controlled Effectiveness Trial: A Qualitative Study. *Psychiatric Services*, *64*(11), 1110–1118. https://doi.org/10.1176/appi.ps.004682012

Passel, J. S., & Woodrow, K. A. (1987). Change in the undocumented alien population in the United States, 1979-1983. *International Migration Review*, *21*, 1304–1334.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, Calif: Morgan Kaufmann Publishers.

Pearl, J. (1998). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco, Calif: Kaufmann.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, U.K. ; New York: Cambridge University Press.

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. New York: Basic Books.

Pearlin, L. I., Schieman, S., Fazio, E. M., & Meersman, S. C. (2005). Stress, Health, and the Life Course: Some Conceptual Perspectives. *Journal of Health and Social Behavior*, *46*(2), 205–219. https://doi.org/10.1177/002214650504600206

Pentland, A. (2008). *Honest signals: How they shape our world*. Cambridge, Mass: MIT Press.

Petricoin, E. F., Ardekani, A. M., Hitt, B. A., Levine, P. J., Fusaro, V. A., Steinberg, S. M., … Liotta, L. A. (2002). Use of proteomic patterns in serum to identify ovarian cancer. *The Lancet*, *359*(9306), 572–577. https://doi.org/10.1016/S0140-6736(02)07746-2

Pietsch, W. (2016). The Causal Nature of Modeling with Big Data. *Philosophy & Technology*, *29*(2), 137–171. https://doi.org/10.1007/s13347-015-0202-2

Portes, A. (1997). Immigration Theory for a New Century: Some Problems and Opportunities. *International Migration Review*, *31*(4), 799. https://doi.org/10.2307/2547415

Portes, A., & Sensenbrenner, J. (1993). Embeddedness and Immigration: Notes on the Social Determinants of Economic Action. *American Journal of Sociology*, *98*(6), 1320–1350. https://doi.org/10.1086/230191

Radermacher, W. (2005). The Reduction of Complexity by Means of Indicators: Case Studies in the Environmental Domain. In The Organisation for Economic Co-operation and Development (Ed.), *Statistics, Knowledge and Policy Key Indicators to Inform Decision Making : Key Indicators to Inform Decision Making* (pp. 163–174). OECD Publishing.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems*, *2*(1). https://doi.org/10.1186/2047-2501-2-3

Ramström, G. (2018). Coleman's Boat Revisited: Causal Sequences and the Micro-macro Link. *Sociological Theory*, *36*(4), 368–391. https://doi.org/10.1177/0735275118813676

Reichenbach, H. (1949). *The theory of probability*. Berkeley: University of California Press.

Reichenbach, H. (1958). The Direction of Time. *The Philosophical Quarterly*, *8*(30), 72. https://doi.org/10.2307/2216858

Reichert, J. (1981). The Migrant Syndrome: Seasonal U.S. Wage Labor and Rural Development in Central Mexico. *Human Organization*, *40*(1), 56–66. https://doi.org/10.17730/humo.40.1.c6148p5743512768

Reiss, J. (2009). Causation in the Social Sciences: Evidence, Inference, and Purpose. *Philosophy of the Social Sciences*, *39*(1), 20–40. https://doi.org/10.1177/0048393108328150

Reyes, B. (2004). U.S. Immigration Policy and the Duration of Undocumented Trips. In J. Durand & D. S. Massey (Eds.), *Crossing the Border: Research from the Mexican Migration Project* (pp. 299–320). New York: Russell Sage Foundation.

Riach, K., Rumens, N., & Tyler, M. (2014). Un/doing Chrononormativity: Negotiating Ageing, Gender and Sexuality in Organizational Life. *Organization Studies*, *35*(11), 1677–1698. https://doi.org/10.1177/0170840614550731

Robinson, S., & Mendelson, A. L. (2012). A Qualitative Experiment: Research on Mediated Meaning Construction Using a Hybrid Approach. *Journal of Mixed Methods Research*, *6*(4), 332–347. https://doi.org/10.1177/1558689812444789

Roethlisberger, F. J., & Dickson, W. J. (1964). *Management and the worker: an account of a research program conducted by the Western Electric Company, Hawthorne Works, Chicago, by FJ Roethlisberger and William J. Dickson, with the assistance and collaboration of Harold A. Wright*. Cambridge, Mass: Harvard University Press.

Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese*, *192*(11), 3731–3755.

Rose, D., & Harrison, E. (2007). The European Socio-economic Classification: A New Social Class Schema for Comparative European Research. *European Societies*, *9*(3), 459–490. https://doi.org/10.1080/14616690701336518

Ross, M. L. (2004). How Do Natural Resources Influence Civil War? Evidence from Thirteen Cases. *International Organization*, *58*(01). https://doi.org/10.1017/S002081830458102X

Rossman, G. B., & Wilson, B. L. (1985). Numbers and Words: Combining Quantitative and Qualitative Methods in a Single Large-Scale Evaluation Study. *Evaluation Review*, *9*(5), 627–643. https://doi.org/10.1177/0193841X8500900505

Russo, F. (2009). *Causality and causal modelling in the social sciences: Measuring variations*. Dordrecht: Springer.

Russo, F., & Williamson, J. (2007). Interpreting Causality in the Health Sciences. *International Studies in the Philosophy of Science*, *21*(2), 157–170. https://doi.org/10.1080/02698590701498084

Ruzzene, A. (2014). Process Tracing as an Effective Epistemic Complement. *Topoi*, *33*(2), 361–372. https://doi.org/10.1007/s11245-013-9195-6

Ryle, G. (1968). *The Thinking of Thoughts: What Is 'Le Penseur' Doing?'*. University of Saskatchewan.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, N.J: Univ. Pr.

Salmon, W. C. (1997). Causality and Explanation: A Reply to Two Critiques. *Philosophy of Science*, *64*(3), 461–477. https://doi.org/10.1086/392561

Sapsford, R. (2006). Indicator. In V. Jupp (Ed.), *The Sage Dictionary of Social Research Methods*. London: SAGE Publications Ltd.

Sauer, P. J. J., Nicholson, A., & Neubauer, D. (2016). Age determination in asylum seekers: Physicians should not be implicated. *European Journal of Pediatrics*, *175*(3), 299–303. https://doi.org/10.1007/s00431-015-2628-z

Schelling, T. (1978). *Micromotives and. Macrobehavior*. New York: W. W. Norton & Company.

Schulte, P. A. (1993). A conceptual and historical framework for molecular epidemiology. In P. A. Schulte & F. P. Perera (Eds.), *Molecular epidemiology: principles and practices*. San Diego: Academic Press.

Schulte, P. A., & Perera, F. P. (Eds.). (1993). *Molecular epidemiology: Principles and practices*. San Diego: Academic Press.

Schultze, U., & Avital, M. (2011). Designing interviews to generate rich data for information systems research. *Information and Organization*, *21*(1), 1–16. https://doi.org/10.1016/j.infoandorg.2010.11.001

Schutt, R., & O'Neil, C. (2013). *Doing data science* (First edition). Beijing ; Sebastopol: O'Reilly Media.

Selvitella, A. (2017). The ubiquity of the Simpson's Paradox. *Journal of Statistical Distributions and Applications*, *4*(1). https://doi.org/10.1186/s40488-017-0056-5

Shaffer, P. (2018). *Causal pluralism and mixed methods in the analysis of poverty dynamics* (No. 2018/115; pp. 1–18). Helsinki: The United Nations University World Institute for Development Economics Research.

Shin, E. K., Mahajan, R., Akbilgic, O., & Shaban-Nejad, A. (2018). Sociomarkers and biomarkers: Predictive modeling in identifying pediatric asthma patients at risk of hospital revisits. *Npj Digital Medicine*, *1*(1). https://doi.org/10.1038/s41746-018-0056-y

Shmueli, G. (2017). Research Dilemmas with Behavioral Big Data. *Big Data*, *5*(2), 98–119. https://doi.org/10.1089/big.2016.0043

Silipo, R. (2007). Neural Networks. In M. R. Berthold & D. Hand (Eds.), *Intelligent Data Analysis* (2nd ed., pp. 269–320). New York: Springer-Verlag Berlin Heidelberg.

Simpson, E. H. (1951). The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *13*(2), 238–241.

Simpson, N. (2017). Demographic and economic determinants of migration. *IZA World of Labor*. https://doi.org/10.15185/izawol.373

Skocpol, T. (1979). *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. New York and Cambridge: Cambridge University Press.

Smith, R. C. (2006). *Mexican New York: Transnational lives of new immigrants*. Berkeley: University of California Press.

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search* (2nd ed). Cambridge, Mass: MIT Press.

Steel, D. (2004). Social Mechanisms and Causal Inference. *Philosophy of the Social Sciences*, *34*(1), 55–78. https://doi.org/10.1177/0048393103260775

Steel, D. (2008). *Across the boundaries: Extrapolation in biology and social science*. Oxford ; New York: Oxford University Press.

Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North Holland Publishing.

Tang, Y., Xu, Z., & Zhuang, Y. (2016). Bayesian Network Structure Learning from Big Data: A Reservoir Sampling Based Ensemble Method. In H. Gao, J. Kim, & Y. Sakurai (Eds.), *Database Systems for Advanced Applications* (Vol. 9645, pp. 209–222). https://doi.org/10.1007/978-3-319-32055-7_18

Tashakkori, A., & Teddlie, C. (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: SAGE.

Taylor, C. F., Field, D., Sansone, S., Aerts, J., Apweiler, R., Ashburner, M., … Wiemann, S. (2008). Promoting coherent minimum reporting guidelines for biological and biomedical investigations: The MIBBI project. *Nature Biotechnology*, *26*(8), 889–896. https://doi.org/10.1038/nbt.1411

Tietjen, G. E., Khubchandani, J., Herial, N. A., & Shah, K. (2012). Adverse Childhood Experiences Are Associated With Migraine and Vascular Biomarkers. *Headache: The Journal of Head and Face Pain*, *52*(6), 920–929. https://doi.org/10.1111/j.1526-4610.2012.02165.x

Tilly, C. (2007). Trust Networks in Transnational Migration. *Sociological Forum*, *22*(1), 3–24. https://doi.org/10.1111/j.1573-7861.2006.00002.x

Titiunik, R. (2015). Can Big Data Solve the Fundamental Problem of Causal Inference? *PS: Political Science & Politics*, *48*(01), 75–79. https://doi.org/10.1017/S1049096514001772

Tost, L. P., Gino, F., & Larrick, R. P. (2012). Power, competitiveness, and advice taking: Why the powerful don't listen. *Organizational Behavior and Human Decision Processes*, *117*(1), 53–65. https://doi.org/10.1016/j.obhdp.2011.10.001

Triantafillou, S., Lagani, V., Heinze-Deml, C., Schmidt, A., Tegner, J., & Tsamardinos, I. (2017). Predicting Causal Relationships from Biological Data: Applying Automated Causal Discovery on Mass Cytometry Data of Human Immune Cells. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-08582-x

Trinchero, R. (2014). Five research principles to overcome the dualism quantitative-qualitative. *Education Sciences & Society*, *5*(1).

United Nations, Department of Economic and Social Affairs, Population Division. (2017). *International Migration Report 2017: Highlights* (No. ST/ESA/SER.A/404). United Nations.

Uprichard, E., & Dawney, L. (2016). Data Diffraction: Challenging Data Integration in Mixed Methods Research. *Journal of Mixed Methods Research*, 155868981667465. https://doi.org/10.1177/1558689816674650

Varian, H. R., Farrell, J., & Shapiro, C. (2004). *The economics of information technology: An introduction*. Cambridge ; New York: Cambridge University Press.

Vineis, P., Avendano-Pabon, M., Barros, H., Chadeau-Hyam, M., Costa, G., Dijmarescu, M., … Stringhini, S. (2017). The biology of inequalities in health: The LIFEPATH project. *Longitudinal and Life Course Studies*, *8*(4). https://doi.org/10.14301/llcs.v8i4.448

Vineis, P., Chadeau-Hyam, M., Gmuender, H., Gulliver, J., Herceg, Z., Kleinjans, J., … Wild, C. P. (2017). The exposome in practice: Design of the EXPOsOMICS project. *International Journal of Hygiene and Environmental Health*, *220*(2), 142–151. https://doi.org/10.1016/j.ijheh.2016.08.001

Vineis, P., Illari, P., & Russo, F. (2017). Causality in cancer research: A journey through models in molecular epidemiology and their philosophical interpretation. *Emerging Themes in Epidemiology*, *14*(1). https://doi.org/10.1186/s12982-017-0061-7

Vineis, P., & Kriebel, D. (2006). Causal models in epidemiology: Past inheritance and genetic future. *Environmental Health*, *5*(1). https://doi.org/10.1186/1476-069X-5-21

Vineis, P., & Perera, F. (2007). Molecular Epidemiology and Biomarkers in Etiologic Cancer Research: The New in Light of the Old. *Cancer Epidemiology Biomarkers & Prevention*, *16*(10), 1954–1965. https://doi.org/10.1158/1055-9965.EPI-07-0457

Vineis, P., van Veldhoven, K., Chadeau-Hyam, M., & Athersuch, T. J. (2013). Advancing the application of omics-based biomarkers in environmental epidemiology: Application of Omics in Environmental Epidemiology. *Environmental and Molecular Mutagenesis*, *54*(7), 461–467. https://doi.org/10.1002/em.21764

Vromen, J. I. (2011). Routines as multilevel mechanisms. *Journal of Institutional Economics*, *7*(02), 175–196. https://doi.org/10.1017/S1744137410000160

Wang, J., Tang, Y., Nguyen, M., & Altintas, I. (2014, December). *A Scalable Data Science Workflow Approach for Big Data Bayesian Network Learning*. 16–25. https://doi.org/10.1109/BDC.2014.10

Wang, T. (2013). Big Data needs thick data. Retrieved July 18, 2018, from Ethnography Matters website: http://ethnographymatters.net/2013/05/13/big-data-needs-thick-data/

Wang, T. J., Gona, P., Larson, M. G., Levy, D., Benjamin, E. J., Tofler, G. H., … Vasan, R. S. (2007). Multiple Biomarkers and the Risk of Incident Hypertension. *Hypertension*, *49*(3), 432–438. https://doi.org/10.1161/01.HYP.0000256956.61872.aa

Warburg, O. (1956). On the origin of cancer cells. *Science*, *123*, 309–314.

Weber, E. (2009). How Probabilistic Causation Can Account for the Use of Mechanistic Evidence. *International Studies in the Philosophy of Science*, *23*(3), 277–295. https://doi.org/10.1080/02698590903197757

Weiss, R. S. (1995). *Learning from strangers: The art and method of qualitative interview studies* (First Free Press paperback ed). New York: Free Press.

Wiest, R. E. (1973). Wage-Labor Migration and the Household in a Mexican Town. *Journal of Anthropological Research*, *29*(3), 180–209. https://doi.org/10.1086/jar.29.3.3629935

Wild, C. P. (2012). The exposome: From concept to utility. *International Journal of Epidemiology*, *41*(1), 24–32. https://doi.org/10.1093/ije/dyr236

Wilde, M., & Parkkinen, V. (2017). Extrapolation and the Russo–Williamson thesis. *Synthese*. https://doi.org/10.1007/s11229-017-1573-y

Wilkinson, R. G., Marmot, M. G., World Health Organization, Regional Office for Europe, WHO Healthy Cities Project, & WHO International Centre for Health and Society. (2003). *The solid facts*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=102456

Williams, D. R. (1999). The Monitoring of Racial/Ethnic Status in the USA: Data Quality Issues. *Ethnicity & Health*, *4*(3), 121–137. https://doi.org/10.1080/13557859998092

Williams, J. (2013). *Constructing New Professional Identities: Career Changers in Teacher Education*. Rotterdam: SensePublishers.

Williamson, J. (2005). *Bayesian nets and causality: Philosophical and computational foundations*. New York: Oxford University Press.

Williamson, J. (2010). *In defence of objective Bayesianism*. Oxford ; New York: Oxford University Press.

Williamson, J. (2018). Establishing causal claims in medicine. *International Studies in the Philosophy of Science*.

Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, Mass: Harvard University Press.

Winters, P., de Janvry, A., & Sadoulet, E. (2001). Family and community networks in Mexico- U.S. migration. *Journal of Human Resources*, *36*, 159–184.

Wolcott, H. F. (1995). *The art of fieldwork*. Walnut Creek: AltaMira Press.

Wood, E. J. (2003). *Insurgent collective action and civil war in El Salvador*. New York: Cambridge University Press.

Woodson, D. G. (2007). What Do Indicators Indicate? Reflections on the Trials and Tribulations of Using Food Aid to Promote Development in Haiti. In L. W. Field & R. G. Fox (Eds.), *Anthropology put to work* (pp. 129–148). Oxford ; New York: Berg.

Woodward, J. (2005). *Making things happen: A theory of causal explanation* (Oxford Univ. Press paperback). Oxford: Oxford Univ. Press.

Woodward, J. (2014). A Functional Account of Causation; or, A Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard That Matters—Usefulness (as Opposed to Metaphysics or Agreement with Intuitive Judgment). *Philosophy of Science*, *81*(5), 691–713. https://doi.org/10.1086/678313

Woodward, J., & Hitchcock, C. (2003). Explanatory Generalizations, Part I: A Counterfactual Account. *Nous*, *37*(1), 1–24. https://doi.org/10.1111/1468-0068.00426

Yoshikawa, H., Weisner, T. S., Kalil, A., & Way, N. (2013). Mixing qualitative and quantitative research in developmental science: Uses and methodological choices. *Qualitative Psychology*, *1*(S), 3–18. https://doi.org/10.1037/2326-3598.1.S.3

Zeng, T., Zhang, W., Yu, X., Liu, X., Li, M., & Chen, L. (2016). Big-data-based edge biomarkers: Study on dynamical drug sensitivity and resistance in individuals. *Briefings in Bioinformatics, 17*(4), 576–592. https://doi.org/10.1093/bib/bbv078