Describing and Evaluating the Social Sciences : Lessons from a Conceptual Scheme and its Example Implementation

Bordo Garcia, Luis Alejandro

Describing and Evaluating the Social Sciences: Lessons from a Conceptual Scheme and its Example Implementation

Luis Alejandro Bordo García

Thesis presented for the degree of Master of Arts in Philosophy with Specialization in the Philosophy of Science

Directed by Prof. Marcel Weber
Department of Philosophy, University of Geneva
Switzerland

August, 2020

Acknowledgments

I would like to thank the following people, in no particular order. Professors Christian Wüthrich and Claudio Calosi, who indirectly influenced this work by exchanging ideas with me numerous times. Professor Lorenzo Casini, Dr. María José Ferreira Ruiz, Dr. Edgar Phillips, Dr. Silvia de Cesare, and Michaëla Egli, for their invaluable feedback on my first partial draft, presented at an lgBIG meeting on November 2019. Professor Catherine Herfeld, for her feedback on an early partial draft. Chiara Bigarella, for her help in inspecting many of my drafts and ideas, and in guiding me in using better digital tools for writing and organizing my thesis project. And Professor Marcel Weber, for his patience and feedback throughout the development of my thesis project.

Abstract

I first elaborate a descriptive and evaluative conceptual scheme intended to organize the analysis of any social-scientific framework. The descriptive aspect regards the historical development of ideas of the framework under question. The evaluative aspect regards this development both (a) in relative terms to what practitioners find important, and (b) in more general terms, proposing some explicit evaluative criteria. Then, I show how this scheme could be implemented using a restricted part of the historical development of the account of individual decision-making of Expected Utility Theory (EUT) as an example. To do this, I use Network Analysis to organize an analysis of citations between texts to guide the historical description of this development, which is then used as a base to propose possible lines of evaluation. Lastly, I present a series of reflections and lessons that the elaboration of this scheme and its example implementation leave for the Philosophy and History of the Social Sciences.

Contents

Chapter 1

Introduction

I will begin this work by presenting my background motivations and perspective, which are important to understand the spirit of the following pages. Then, I will present the general considerations to have in mind to understand surrounding the features of this work. Finally, I will present my research questions and the general structure of this thesis.

1.1 Motivation and Perspective

In my view, in the history of Social Sciences, there have been some commitments about the nature of reality, and about how to create knowledge about reality, that have not always received sufficient attention and proper analysis from the part of practitioners. Because of this, some ideas present in social scientific theories, related to said commitments, may potentially imply problems. For example, some models of individual decision-making in Economics are criticized because they assume that agents make flawless ("rational") calculations, with perfect information, about the consequences of their choices in advance and then act accordingly (Epstein, 2015, p. 5). Critics to these models note that by committing to a flawed representation of individual decision-making, economic models end up not correctly predicting people's real behavior. In this line, my main motivation for the present work is that I believe that the Social Sciences can benefit greatly from the identification and discussion of such type of issues.

To understand the spirit of this work, it is important to make explicit that I sympathize with *naturalism* as a perspective for studying both social reality and the Social Sciences, in the sense noted by Harold Kincaid (2012):

> [Naturalism] denies that there is something special about the social world that makes it unnameable to scientific investigation, and also denies that there is something special about philosophy that makes it independent or prior to the sciences in general and the social sciences in particular. (p. 3)

From this it follows that, if we accept that social reality is amenable to scientific investigation (as is the case, in my opinion), then we should do our best effort to develop correct and accurate accounts of it.

And to complement the second part of Kincaid's quote, I would add that Philosophy of Science contains a series of analytical tools that may help us with this task, for example by allowing us to identify possible theoretical or methodological inconsistencies in an existing scientific theory. Or by allowing us to look back into the general picture, after specialization makes us lose track of all the assumptions into play and their potential problems. In particular, I find appealing the contextualization of scientific practices that historical approaches in the Philosophy of Science allow.

With this in mind, an important part of this work is dedicated to the development of a schematic way of historically describing and evaluating social-scientific accounts of reality and showing a limited implementation of it as an example. This descriptive-evaluative scheme and its further implementation is the result of my personal attempt to unify different concerns I had during my time in the Master's Program. In my view, these concerns are all connected by the idea that, while a part of the Philosophy of Science de facto attempts to make descriptions and evaluations of scientific disciplines, it may sometimes lack a clear structure for doing this. So, my general ambition in this work is to present a series of reflections on the Philosophy of Science based on the exercise of developing and implementing my descriptive-evaluative scheme.

1.2 General Considerations

Now, a first consideration regarding my approach is that this work is *not* dedicated to compare and criticize different approaches in the Philosophy of Science to then develop a proposal. This work instead starts with a proposal, discussing and building on different ideas that I found to be consistent with each other. The main ones include Ludwig Fleck's (1935/1979) approach to analyze the *historical development of scientific ideas*, the debate around *folk psychology* and its role in the Social Sciences, and Paul Hoyningen-Huene's (2013) concept of *systematicity*.

As I intend to show, when put together and implemented these ideas allow for a nuanced description of the historical development of social-scientific accounts of reality, which then further serves as a base for a contextualized evaluation. The descriptive-evaluative scheme that I will propose has therefore a dual nature: it proposes a way of *describing* the historical development of ideas; and it further proposes a way of *evaluating* this development, based on its features and context.

After presenting the descriptive-evaluative scheme, this work then continues by presenting an example implementation of it, on which there are two considerations to note. First, I chose to study a particular type of Decision Theory, called Expected Utility Theory (EUT), as has been understood and developed mainly in Economics. There are many interesting and important debates surrounding EUT's representation of *individual decision-making*, so I decided that would be a good fit for this work.

Second, the particular way I implemented my scheme is through the means of the research method known as *Network Analysis*, which has merits and appeal on its own[1], and that I found to be compatible with Fleck's approach to studying the development of scientific ideas. However, as I will discuss later, several limitations impeded this to be a *full* implementation of my scheme as I would have wanted, and so my results do not allow for concluding remarks regarding EUT's historical development. The value of this work surrounds, instead, the lessons and reflections that my approach leaves for the Philosophy and History of the

[1] See for example Herfeld and Malte (2018).

Social Sciences.

1.3 Research Questions and Structure of this Work

With these considerations in mind, the following research question will guide the present work:

(**Question 1**) What methodological lessons are there to be learned from the exercise of developing a explicit scheme to describe and evaluate the historical development of social-scientific accounts of reality, and showing an example of its implementation using Network Analysis?

As this is a very broad question implying many steps, before attempting to answer it I will address two complementary questions:

(**Question 1.1**) How may we develop a explicit scheme to guide the historical description and an evaluation of the development of any given social-scientific account of reality?
(**Question 1.2**) How may we implement such a scheme on a real social-scientific account of reality?

To answer these questions, my objectives and the structure of this work are the following. In Chapter 2 I will address Question 1.1 by first discussing the notions of *social-scientific framework*, *folk psychology*, and *systematicity*. Then, I will put together a synthesis of the ideas discussed in this chapter in the form of a conceptual scheme aimed at guiding both a historical analysis and a contextualized evaluation of the development of social-scientific ideas.

In Chapter 3 I will address Question 1.2 by first operationalizing the conceptual scheme developed in Chapter 2 using network analysis, explicitly stating the limitations of my work. Then, I will present the example implementation itself which is composed of a historical description of the development of Expected Utility Theory's (EUT) account of individual decision-making, as developed mainly (but not exclusively) in Economics, and a characterization and evaluation of this development based on what was described.

Finally, in Chapter 4 I will address Question 1 by reflecting on the main lessons for the Philosophy and History of the Social Sciences left by the exercise of developing and operationalizing this descriptive-evaluative scheme, and exemplifying its implementation on a part of the historical development of EUT.

Chapter 2

Descriptive and Evaluative Conceptual Scheme

In this chapter, I will discuss a series of concepts that, in virtue of being unified by the broad aims of describing and evaluating scientific disciplines, will help me propose a scheme to describe and evaluate the development of social-scientific accounts of reality. The main objective of this chapter is, thus, to present a conceptual descriptive and evaluative scheme.

Now, the first two sections in this chapter correspond to two threads of discussion. They are (1) the notion of *social-scientific framework* and (2) the notions of *folk psychology*, *systematicity*, and their roles in scientific disciplines. Hence, the objective of these sections is to clarify these concepts, particularly in such a way that it is clear that they are consistent when put together.

This discussion is important because, on one hand, there is no clear consensus on what a "scientific framework" (or "tradition", "school", etc.) exactly is, so if I were to just assume this term, the delimitation criteria to choose the particular entities (e.g. particular ideas, authors, texts, etc.) with which to represent an object of study (e.g. Economics) would not be clear. By building my scheme on an explicit definition of social-scientific framework, I expect to be able to frame my object of study in such a way that there is an easy transition to applying some method of empirical research[2].

[2] That is, a method that allows in an organized way to gather and analyze information about

On the other hand, there is no consensus in the debate on what exactly folk psychology is and the role it plays in social-scientific disciplines. I will defend, however, particular positions in these debates which will then allow me to use of this concept for the evaluative aspect of my scheme. In contrast, the concept of systematicity has been thoroughly discussed by Hoyningen-Huene (2013), and I will largely base my understanding of this idea on his book. By building part of my scheme on the concepts of folk psychology and systematicity, I expect to have a rich and nuanced base from which to characterize and evaluate the historical development of ideas of social-scientific frameworks.

2.1 Social-Scientific Framework

First, I think it is relevant to pose the following general question: what exactly is a scientific framework? To begin with, the first word of the term "scientific framework" certainly touches the problem of demarcation between science and non-science. This is a non-trivial issue and I will make a brief clarification in order to not leave gaps in the development of my ideas. Now, certainly, the discussion on this may very well be a separate work on its own, as it is a major topic in the Philosophy of Science[3]. And as we will see in a moment, when we consider the Social Sciences in the light of the demarcation problem the discussion becomes complicated very quickly.

So, to stay in the tracks of what the present work is about, I am going to take a simple and pragmatic posture on the matter. I will assume that any framework used by *institutionalized science* is scientific. And by this I will refer to the knowledge-oriented disciplines practiced in the context of both, (a) an officially recognized (i.e. by its home State) and (b) a peer-recognized (i.e. by other similar organizations) research organization, such as a university or a research institute. That is, I will simply assume that *Science is what institutionally recognized scientists do*. And in a similar vein, *Social Science is what institutionally recognized social*

the historical development of social-scientific ideas.

[3] For more detailed discussion on this, see Hansson (2008/2017).

scientists do[4].

Sure enough, there are potential criticisms to be made to this perspective, but then again the discussion in the literature is far from settled and there is no agreement on what the demarcation criteria for 'Science' should be (Hoyningen-Huene, 2013, pp. 5-6). Since this work is *not* about the demarcation problem, the idea here is just to settle in a functional definition for my purposes: I am putting into hold a potentially time-consuming discussion on what exactly characterizes Science in general, while allowing me to be inclusive enough of any potential social-scientific framework of interest, as understood intuitively.

One of my main concerns in this chapter is that failing to be inclusive, in my view, may put my conceptual scheme at risk. The reason is that some parts of the institutionalized Social Sciences have particular epistemic goals that may conflict with positive definitions of "Science", and I do not want to exclude them before even looking at them in detail. To be more clear, the following subsection is dedicated to further explain the point on the epistemic goals of the Social Sciences.

2.1.1 Epistemic Goals of the Social Sciences

It is important to note that, in this work, I am taking *Social Sciences* to mean a group of disciplines that pretend to account for phenomena related to human behavior and relations, including (mainly, but not exclusively) Anthropology, Economics, History, Political Science, Psychology, and Sociology. However, my interest is not on the institutional disciplines themselves but on *how to analyze social-scientific frameworks*, most of which transcend the institutional disciplinary boundaries of what we call 'the Social Sciences'[5]. So, for my purposes, a remark to consider is that it would not be relevant to map how frameworks relate to each

[4] Of course, this is not a novel perspective since it is close to a (social or cultural) relativistic perspective. As I will argue later, however, I will not commit to relativism when advancing judgments on the social-scientific framework to be analyzed. That is, even if something is institutionally (or socially) recognized as a science, it does not mean that it is infallible or immune to errors. Instead, I chose this definition because it is functional for my descriptive purposes.

[5] For example, the aforementioned Expected Utility Theory is present mainly in Economics, but also in some other disciplines, such as Political Science and Sociology. For more on what has been called the *hybridization of social knowledge* (in the sense of how social-scientific traditions exceed institutional boundaries), see Dogan (1996).

institutional discipline, a task that may be worthy of a different work.

With this in mind, consider now that different social-scientific frameworks may not differ only in their scopes (i.e. the part of reality that they study) or on their basic assumptions, as may be the case in the Natural Sciences. Social-scientific frameworks may have different, and even antagonistic, epistemic goals. For example, while some try to explain and predict what are considered objective facts of social reality, others deny the possibility of objective knowledge and will focus on an empathetic approach to clarify what it means to 'understand' social reality (Della Porta & Keating, 2008, p. 23). This second kind of approach is what Della Porta and Keating (2008, p. 25) call "humanistic", and is concerned with ontological and epistemological issues. Its methods of research focus mainly on the meanings of the ideas and worldviews of the subjects of research, where empathetic interactions between researchers and the subjects are considered a crucial issue (2008, pp. 31-32).

So, the Social Sciences are not a unified project, but a patchwork of different activities with different epistemic goals, values, and methods. Given this situation, it is understandably hard to construct a positive definition of "Social Science" that is inclusive enough for all of them. Now, moving onto the next topic, a similar remark applies to the "framework" part of "social-scientific framework": give a definition that is too restrictive and have as a consequence that some institutional social-scientific traditions are excluded.

2.1.2 Thought Style, Thought Collective, and Social-Scientific Framework

Conveniently, there is an author whose account of the notion of scientific framework is inclusive enough and is intimately tied with a historical approach, being hence functional for this work. In *Genesis and Development of a Scientific Fact*, Ludwig Fleck (1935/1979) uses the case example of the development of the concept of syphilis, from its origins as a term until its twentieth-century scientific formulation, to discuss more generally the process of how scientific concepts develop historically. An important part of his book relies on the concepts of *thought collective* and *thought style* (1935/1979):

> If we define "thought collective" as a community of persons mutually exchanging ideas or maintaining intellectual interaction, we will find by implication that it also provides the special "carrier "for the historical development of any field of thought, as well as for the given stock of knowledge and level of culture. This we have designated thought style. The thought collective thus supplies the missing component. (p. 39)

A crucial part of these definitions is that a thought collective *carries* a thought style. In the light of these terms, Fleck mentions that "someone recognizes something (. . .) in a particular thought style, in a particular thought collective" (1935/ 1979, p. 39). The "someone" may be, for example, medical researchers, while the "something" may be a concept such as syphilis. The idea is that a concept only makes sense in the context of a thought style and thought collective, never by its own[6]. Furthermore, this is compatible with people belonging to more than one thought collective (1935/1979, p. 42), leaving open the possibility that a community can work in lines of research that are part of different thought styles[7].

Using these definitions, we may instantiate them understanding the "someone" as social scientists, while the "something" as their accounts of social phenomena. The complete phrase would be, then, *social scientists recognize social phenomena in a particular thought style, in a particular thought collective*. Consider again the remarks made about how the social-scientific traditions may sometimes be isolated from each other, depending on the differences in their epistemic goals (Della Porta & Keating, 2008, pp. 19-21) (sometimes even very openly and voluntarily in virtue of their differences). Given this situation, it makes sense to state that the Social Sciences are composed of different, sometimes even isolated, thought collectives of people that internally maintain intellectual interaction, and

[6] An 'isolated concept' is meaningless under Fleck's account.

[7] In any case, given what has been called the hybridization of social knowledge (Dogan, 1996), if we want to be inclusive of most of the social-scientific traditions there seems to be no other alternative but to accept this.

which may not necessarily correspond to the institutional disciplinary boundaries.

With the discussion made so far, under the light of Fleck's definitions we can be inclusive enough of what we find in the contemporary practice of the Social Sciences. So, I am now in a position to clarify what I intend by using the term "social-scientific framework". I will understand it as *a particular thought style about human behavior and relations, developed in the context of at least one of the institutionally recognized Social Sciences, carried by a thought collective whose people mainly are institutionally recognized social scientists*. As desired, this approach to social-scientific frameworks is inclusive of any kind of institutionally recognized thought collective, regardless of the epistemic goals that are part of their thought style.

While this approach may not be flawless, there does not seem to be any approach exempt from critiques. And since this work is not about the definition of "Social Science" itself (or of "framework"), I will keep with this approach in virtue of it being both convincing enough and functional for my purposes.

In particular, this definition has the important advantage of being compatible with a method of empirical research, through an operationalization of the idea of "intellectual interaction" (which is key for the definition of thought collective, and thus of social-scientific framework). This method is known as *Network Analysis*, and I will discuss its key features pertinent for my work at the beginning of Chapter 3. It is important to note here that Network Analysis is a *method*, not a theoretical approach. It gives us a systematic way of organizing information about interactions between entities, as well as some techniques to study the properties of the structure of said interactions and entities represented. Thus, on its own, this method does not tell us what we should analyze and hence this *what* may end up being very different things (e.g. hierarchical relations in an organization; economic interactions; chains of e-mails; etc.).

With this in mind, I will now discuss some features of scientific frameworks that will be useful for my descriptive purposes.

2.1.3 Features of a Scientific Framework

This section discusses some particular features of scientific frameworks that will serve as a base for their analysis. It is important to note that what follows is not meant to be an exhaustive list of all the possible features of scientific frameworks, and I will only present the ones I think are relevant for the present work. As a base for different levels of inquiry, in the example implementation of the next chapter I hope to show the usefulness of defining these features.

Existential Implications and Methodological Standards

The first two features related to scientific frameworks that I will preset are its *existential implications* and its *methodological standards*.

Surrounding the first of these features, I found Rudolf Carnap's (1950) ideas to be largely compatible with Fleck's account as described so far. Regardless of Carnap not explicitly using this term, I will understand *existential implications*[8] in the sense that "if someone accepts a framework of entities, then he is bound to admit its entities as possible designata" (Carnap, 1950, p. 35). It should be noted that Carnap's term "framework" is arguably best understood in the sense of "linguistic framework", as expressed by Carnap himself: "If someone wishes to speak in his language about a new kind of entities, he has to introduce a system of new ways of speaking, subject to new rules; we shall call this procedure the construction of a framework for the new entities" (1950, p. 21)[9].

But even with this difference in the use of the same labels, Carnap's notion seems to suggest that *every (scientific) thought style is tied with a linguistic framework* (i.e. ways of speaking subject to rules), bringing thus together Fleck's and Carnap's accounts rather easily. If we accept this, then we ought to accept that both the features of a social-scientific framework include both a (Carnap's) linguistic

[8] The term is in fact borrowed from Bricker (2014/2016, §1.1), and I use it to contrast it from the more common "ontological implications". The critique of this latter notion presented by Bricker arguably renders it incompatible with Fleck's account, since it seems to imply that we ought to define a single meta-scientific framework from which to advance evaluations on scientific practice.

[9] So, the reader must be careful in noting that Carnap's term "framework" is different from the term "framework" in "social-scientific framework" as defined earlier

framework and existential implications.

Extending further this discussion, arguably *the entities that the people in a scientific framework will try to observe are tied to the scientific framework's existential implications*. And since different methods of observation are needed depending on what exactly we need to observe, arguably how the research will be done in practice will be intimately linked to the linguistic framework and overall thought style of the practitioners. To reflect this, I will define an additional feature of social-scientific frameworks, which I will name "methodological standards", using Gerald Doppelt's (1990) ideas[10] which are largely compatible with Fleck's and Carnap's:

> [Methodological standards are] cognitive standards which assert in a general way, for all empirical inquiry, what counts as empirical knowledge, proof, evidence, or explanation. Such standards ought to be consistent, mutually plausible and coherent. They embody a generic view of what science is, or ought to be. (p. 16)

Two General Values

To close this subsection, In what I consider to be the spirit of naturalism, I will present here two basic values that any framework with scientific pretensions must have and that, in my view, do not seem to be renounceable.

First, I will assume that logical consistency is desirable, in particular respecting the *principle of no contradiction*[11]. And second, I will assume that any theory with scientific pretensions should try to capture the dynamics found by the observation of reality.

A few words on this last point. While Fleck's (1935/1979, pp. 84-98) remark that a change in thought style brings also change in the way that observations in science are made, here I will take "empirical observation" to mean not a single act of observation but *a triangulation of different observations using different methods,*

[10] While Doppelt uses the terms 'methodological rules' and 'methodologies', in the literature 'methodological standards' seem to be the most common term to refer to what is defined here.

[11] To be explicit, I am referring to the principle of classical logic that claims that a proposition and its negation cannot both be true at the same time.

replication, and statistical rigor, that together help justify the claim that a phenomenon is in fact real. In my view, my understanding of the term *empirical observation* differs from Fleck's historical examples, where there is usually only one or few observers and one or few methods of observation (usually only plain sight), and where there is no mention of triangulation, replication, or statistical rigor.

I consider that these elements make observations *robust*, strengthening the claim of they being *real* observations (i.e. by lowering the probability that what is being observed is the product of error of any kind). Errors may be, for example, incorrect inferences, such as drawing incorrect conclusions from statistical information; or incorrect observations, such as an artifact in the measurements[12]. Additionally, this is a more nuanced representation which is arguably also closer to the actual practice of researchers that gather empirical information (be it experimenters gathering experimental information or other kinds of researchers gathering observational information).

2.1.4 The Genesis of Scientific Ideas

An interesting consideration bridges my definition of the concept of social-scientific framework with the notion of folk psychology. At the genesis of a social-scientific framework, everyday simple (and often biased) ideas about social reality would have been the only account available from which to begin social-scientific inquiries. In Fleck's account of the historical development of scientific thought, we find more generally ideas on the development of scientific concepts from common sense[13]. In the author's terms (1935/1979, p. 24), some very old "hazy proto-ideas" underwent a development that included a stage where they "received a modern expression"[14]. But of course, not all common sense accounts

[12] A simple characterization of artifacts includes two types (Guala, 2005, p. 96): (a) either artifacts in the sense of distortion in what is being observed coming from a defect in the measuring instrument; or (b) artifacts in the sense of influence of the experimenter or observer, who may inadvertently cause a phenomenon in the test subject that otherwise would have not been there.

[13] In this work, I may use "folk", "common sense" and "every day" as synonyms to refer to non-systematic ideas about reality.

[14] Fleck claims that this developmental pattern can be observed for the "theories of the elements and of chemical composition, the law of conservation of matter, the principle of a spherical earth as well as the heliocentric system" (Fleck, 1935/1979, p. 24).

of reality are proto(scientific)-ideas in this sense, and some never find a place in a scientific framework.

The ones that do find a place, however, may later judged to be not valid scientifically, when looking backward in time from the perspective of its modern framework. But making such a judgment is invalid in Fleck's perspective: a concept should only be judged as valid or invalid in a particular moment of its historical development in relation with its surrounding framework, just like an animal can be judged adapted or unadapted only relative to its environmental context, and not in an abstract sense (1935/1979, pp. 25-26). For me this point is crucial, since accepting this approach prevents us from making a priori judgments on the work of disciplines that at present may not be recognized as 'scientific', but that could eventually be developed into a powerful account of reality. The point is, for disciplines that may appear to be in a promising path, to first analyze its features and only then advance judgments on the quality of its scientific practice.

Then, the importance of proto(scientific)-ideas lies not in their absolute truth or validity (in any justifiable sense of these terms) but in their capacity of being a step in the development of a proper scientific concept. And in this process of development, at first, a proto-idea is always "too broad and insufficiently specialized", but then becomes differentiated or specialized from a broad idea into nuanced, technical concept (Fleck, 1935/1979, p. 27). Interestingly, this account may shed some light on how to identify if a particular idea in a social-scientific framework is closer to a proto-idea (i.e. an every day, common sense idea) or to a scientific one: by evaluating its "haziness", vagueness, broadness, and lack of differentiation. Unfortunately, Fleck does not further discuss the properties of common sense ideas and their role in the development of scientific ideas. And so, in order to develop explicit criteria to identify and evaluate them, we need more than just his account.

2.2 Folk Psychology and Systematicity

A very interesting debate in the Philosophy of Mind and the Philosophy of Science regards the notion of *folk psychology* and its relation with scientific inquiry. There are different ways in which we may understand what folk psychology is (Ravenscroft, 1997/2019), but for the purposes of this work, by using this term I will refer only to *the understanding people have about human behavior and mind-related phenomena, that has been developed in a non-systematic way*. So it is "folk" in the sense that folk psychology is *the way the folk understands human behavior and mind*.

Now, part of the debate includes the issue of whether folk psychology can be understood as a theory with a structure comparable to scientific theories. However, I will not enter into the details of this discussion[15]. For my work, it is enough to note that I will assume that there are at least two features shared by folk psychology and scientific accounts of reality: proposing the existence of some phenomena and proposing a way of explaining or understanding these phenomena.

What is of interest here is, rather, how to identify when we are in front of some account of reality that *has been developed in a non-systematic way*.

2.2.1 Systematicity and Identifying Folk Psychology

In his book *Systematicity: The Nature of Science*, Paul Hoyningen-Huene (2013) advances a proposal to address the demarcation problem (i.e. how to differentiate scientific from non-scientific knowledge) based on the notion of *systematicity*.

While his proposal is well constructed and convincing, I already noted above that I will not try to argue for or against some view on the demarcation problem. So, instead of taking Hoyningen-Huene's systematicity (from now on, HHS) to differentiate Science from non-Science, I will take a different approach and use the features of HHS to both *characterize social-scientific thought styles* and *identify*

[15] The account that considers that folk psychology *is* a theory is called in the literature the "theory-theory". For a more detailed discussion on this as well as a critique of this account, see Ravenscroft (1997/2019, §2.1).

folk ideas (what the author calls "non-scientific, everyday knowledge")[16]. By extension, this will allow me to identify folk ideas surviving from the genesis into later in the history of a social-scientific framework. And as I will develop next, HHS can also be used as a base to characterize and evaluate the development of ideas in social-scientific frameworks in terms of their systematicity.

The Dimensions of Systematicity

To begin clarifying the concept of systematicity, the author states that there are many non-comparable intuitions on what a *non-systematic idea* is (Hoyningen-Huene, 2013, pp. 26-30). For example, it may lack some expected order of its components, or it may be incomplete in terms of accounting for an expected amount of classes. Since there is more than one sense in which we may understand it, then, nine *dimensions* are presented such that scientific knowledge may be said to be more systematic than other kinds of knowledge, individually (and depending on the context) for each of these nine dimensions (which is the particular thesis of the author). So these nine dimensions must be taken as a *family of concepts*, that are descendants of the abstract intuition of systematicity mentioned above (in the sense of order, completeness), and that may not be comparable to each other when analyzing concrete examples.

Briefly put, the nine dimensions in which scientific knowledge may be said to be more systematic than everyday knowledge, depending on the standards of each specific scientific framework, are the following (Hoyningen-Huene, 2013, Chapter 3)[17]:

- *Descriptions*: present in most scientific fields, regardless of their aims in terms of other dimensions, concerns the *abstract description* of the phenom-

[16] The author himself concludes that his approach diverges from the classical Philosophy of Science, where it is attempted to differentiate (and judge) Science from Pseudo-Science, claiming that he is trying to instead compare and describe scientific knowledge and non-scientific knowledge (e.g. "everyday knowledge") (Hoyningen-Huene, 2013, p. 209). This goes in line with my choice of using his account to characterize folk knowledge as non-systematic.

[17] I will characterize each dimension briefly, but I will devote some additional lines to dimensions that may be harder to grasp.

ena in the domain of the scientific framework[18]. This dimension is characterized by *what* questions (e.g. "what has happened?"; "what is the case?").

- *Explanations*: to account for the phenomena describe, which are in general characterized by *why* questions (e.g. "why is x the case").
- *Predictions*: the author characterizes this dimension only in the literal sense of prediction, as *prediction of future events* (and not as *implications of a theory or model*).
- *The defense of knowledge claims*: key for most scientific disciplines, related to the idea that they are more systematic than everyday knowledge in guaranteeing the quality of its knowledge claims by minimizing the risk of errors (e.g. biases, false assumptions, dogmatic ideas, fraud, etc.).
- *Institutionalized critical discourse*: concerns the social organization of scientific disciplines that evaluates and decides which ideas will be kept in the "pool of knowledge" of the discipline (e.g. standards for publications; quality control mechanisms in Big Science projects; disciplinary meetings).
- *Epistemic connectedness*: concerns the relationships between the knowledge claims of specific scientific disciplines with the knowledge claims of a wider, more general, scientific community (based on the notion that "no research field is completely independent from all the rest of science" (2013, p. 120)). For example, some particular piece of knowledge may be said to be more systematic when it has applications in many scientific fields, or when it is systematically connected to theories in other fields previously developed.
- *An ideal of completeness*: in a very broad sense, this refers to the idea that scientific disciplines aim at "knowing everything" (2013, p. 126) about their domain of reality[19]. And so scientific knowledge may be said to be more

[18] Abstract in the sense that descriptions "do not cover each and every possible aspect of the phenomenon described" (Hoyningen-Huene, 2013, p. 38).

[19] This dimension is presented as a *description* of an ideal in contemporary scientific communities, regardless of whether this ideal is in fact achievable.

systematic, in the sense of "thorough and complete", than everyday knowledge. Additionally, the author mentions that this ideal is often reflected in the decisions made by the scientific communities.

- *Knowledge generation*: concerns the systematic order in the way in which scientific communities research new topics to generate new knowledge. Since the author's thesis is comparative (and not absolute), particular events of apparently chaotic nature (such as exceptionally creative scientists at the forefront of revolutions in their fields) do not invalidate the idea.

- *The representation of knowledge*: concerns how the scientific disciplines represent their corpus of knowledge with systems of terms that exceed everyday language to refer to the world (e.g. when naming abstract mathematical objects, rare chemical elements or reactions, and others that do not have a correlate in everyday language).

It is worth mentioning that there are many potential relationships between these dimensions, as well as possible positive reinforcements. For example, a theory that produces better explanations and that is epistemically connected to many particular scientific fields, may be used to improve the defense of knowledge claims, which in turn may change institutionalized critical discourse practices. Or research showing the incompleteness of scientific knowledge about a particular topic may start a systematic generation of new knowledge with further research.

Identifying Folk Psychology in Social-Scientific Frameworks

Now, consider that Hoyningen-Huene argues that the relevance given to each dimension will depend on each particular scientific discipline (2013, p. 30). So, together with Fleck's claim that the importance of the context of ideas is important, it arguably follows that a particular configuration of an idea of systematicity (i.e. the role that all or some dimensions are meant to play) must be always understood in the context of a particular scientific framework, and so it will most likely vary for different social-scientific frameworks.

Similarly, as scientific frameworks change historically, their concepts of systematicity may also change (e.g. in the history of some scientific fields, as quantification became more important the way systematicity was understood for descriptions, explanations, and predictions changed). Note that from these considerations, it follows that a framework's concept of systematicity *is a part of their methodological standards* (as defined earlier), as it can be understood as a set of cognitive standards for all or some dimensions, that put together embodies in a general sense how the practice of the scientific discipline ought to be done.

Hoyningen-Huene's notes that because of these last considerations (which he presents in his terms, without mention of Fleck's ideas), his thesis risks being too relativistic when considering scientific fields (in an enlarged sense) that lack a clear consensus on many of their fundamentals issues. Namely, the Social Sciences and the Humanities (2013, p. 32). The only way, then, to say that these fields are *more systematic than* folk knowledge (Hoyningen-Huene's main thesis), even when each social-scientific framework may strongly disagree with others on matters such as epistemic goals and methodological standards, is by being very specific.

That is, we should only compare what a social-scientific framework has to say about its particular domain with what folk knowledge may tell us about that domain, instead of comparing the Social Sciences as a whole with folk knowledge. In the words of the author (2013):

> So I should constrain my thesis about the higher degree of systematicity in controversial research areas [e.g. the Social Sciences] to each individual school [i.e. individual social-scientific frameworks] and not apply it to the area as a whole. Of course, that makes our argumentative task even more difficult because in such areas, each individual school needs to be discussed separately. (p. 32)[20]

These considerations are directly relevant to how to recognize if a particular idea is folk. A possible way of doing this could be attempting to map a significant

[20] Considering the patchwork, non-unified nature of the Social Sciences as discussed above, in my opinion, this is the only correct approach (discussing each social-scientific framework separately).

amount of folk ideas that share the same domain as the social-scientific frameworks' ideas and then comparing both groups of ideas[21]. However, since my work is about how to analyze social-scientific frameworks and *not* about folk psychology itself, I will not do this kind of mapping. My approach will be, instead, to evaluate folk psychological ideas present in the social-scientific frameworks under inquiry, *through a contrast with the framework's own synchronic methodological standards* (in particular, its concept of systematicity). It is important to note that by this I mean considering the status of the social-scientific framework *and* its methodological standards as they are the same period in time.

As we see, Hoyningen-Huene's systematicity goes many steps further than Fleck's brief account of common sense ideas. This is furthermore compatible with the theoretical considerations presented in the previous section: analyzing the social-scientific framework's concept of systematicity will allow me to identify a relevant part of its methodological standards[22]. So, in Chapter 3, after presenting a historical analysis, I will present a characterization of the concept of systematicity of the social-scientific framework under analysis, in terms of the roles that the relevant dimensions of HHS play.

Before moving on, an important remark on the historical continuity between folk and contemporary scientific ideas is due here. Coming back to Fleck's account on the historical development of scientific ideas, here I am assuming that, at their genesis, all historical social-scientific thought collectives have based their ideas on folk psychology. But this only in the sense that, even if we find that at the present day there is nothing that could be identified as folk psychology in a social-scientific framework, we can still trace a history of intellectual interactions that leads back to a folk account of social reality. This includes, for example, the hypothetical interactions between (a) the social scientists that had folk ideas, and (b) the ones that at a certain historical point rejected or modified those ideas based on some concept of systematicity.

[21] For example, gather by observation (questionnaires, interviews, historical records, etc.) the folk ideas on a particular social phenomenon (e.g. individual decision-making) present in different societies (past or contemporary).

[22] Relevant because it is this part of the methodological standards (i.e. the particular concept of systematicity) that will allow me to identify folk-psychological ideas in its historical development.

I do not think that it is controversial to claim that all scientific disciplines at their origin were based on many folk concepts, even the more mature natural sciences. This is one of the ideas that Fleck very clearly develops throughout his book (1935/1979), as well as an idea defended by Hoyningen-Huene (2013, §5.1.2; §5.2.1). Hence, it is only in the following sense that I would say there is a continuity between folk ideas and contemporary (social-)scientific ideas. It is *not a continuity of the content* of folk and scientific accounts of reality (which may not be the case if all folk ideas have been discarded or heavily refined), but a *continuity of historical interactions* (which is certainly always the case).

Moving on to the next section, I will now discuss the relationship between folk psychology and social-scientific frameworks.

2.2.2 The Role of Folk Psychology

Hoyningen-Huene presents the notion of systematicity as being *descriptive* of scientific practice, and *not prescriptive* of how "good science" must be (2013, p. 21), so he does not present us an explicit reflection on the potential problems caused by the presence of everyday knowledge in scientific disciplines or on the role it may play. He only hints that, in the historical context of scientific disciplines, new systematic ideas outperform perform previous, less systematic (and more common sense) ideas in terms of scientific values. In particular, sometimes producing more systematic knowledge is a value itself (2013, p. 196), but he also argues that the theory of systematicity, by being descriptive, is not helpful to decide normative criteria ("how science should be?"). Hence, to bridge the evaluative aspect of my scheme, in what follows I will present some personal reflections based on the literature on the role of folk ideas in scientific frameworks.

Consider first that, under HHS, it is clear that there are good reasons to believe that some characteristics of scientific inquiry differentiate it from folk reasoning. In particular, some of these characteristics certainly lead scientific accounts of reality to be comparatively *more accurate than* folk ones as a representation of reality. While it is hard to show precisely which special characteristics account for the more-accurate-than nature of scientific knowledge, Hoyningen-Huene's account is convincing enough to believe that some features of the different dimensions of systematicity contain (at least some of) these special characteristics. For example, a careful and systematic defense of knowledge claims makes scientific knowledge inherently better at trying to prevent biases and errors in reasoning than everyday thinking. However, the author notes that in some cases more systematicity may not be related to some kind of "scientific progress" and, in those particular cases, less systematicity may be preferable (2013, p. 187).

These considerations point again to the direction of first analyzing the practice of a particular scientific framework before trying to sketch an evaluation. To defend this kind of approach, in the rest of this subsection I will take position on the debate surrounding the role that folk ideas play in scientific disciplines.

Eliminativism versus Suggested Defenses of Folk Psychology

In the discussion and evaluation of folk psychology in terms of its accuracy as an account of human behavior, we find some strong positions. On one side of the debate, *eliminativism* (also called *eliminative materialism*) claims that the folk understanding of human behavior and the mind is *always* false, and so should be eliminated (hence the name) from every framework with scientific pretensions (W. Ramsey, 2003/2019). A notable proponent of this perspective is Paul M. Churchland, who claims that "(...) folk psychology is a hopelessly primitive and deeply confused conception of our internal activities" (1984/1988, p. 45).

The two main justifications for this perspective are the following (Churchland, 1984/1988, p. 46). First, the explanatory poverty of folk psychology to explain phenomena outside its usual domain (for example, explaining behavioral and cognitive problems in people with brain damage, being that the usual domain of folk psychology is people that do not have brain damage). And second, an induction made by observing the history of science, whose conclusion is that, since early folk theories about physical, chemical, and biological phenomena were "wildly off the mark", an accurate account of the surely more complex phenomena studied by psychology (and the Social Sciences) must be very different from folk psychology. It is claimed that the reason why folk psychology has survived is not that it is accurate, but rather because "the phenomena addressed are so surpassingly difficult that any useful handle on them, no matter how feeble, is unlikely to be displaced in a hurry" (1984/1988, p. 46).

The point raised by eliminativists is particularly relevant for this work, since *if* folk psychology is intrinsically false or inaccurate, useless, and even detrimental, *then* any folk-psychological idea present in a social-scientific framework should be automatically discarded as quickly as possible after being identified. On the other side of the debate, however, there is an important amount of critiques in the literature against eliminativism. While these critiques are made in several fronts[23], to not extend this discussion more than necessary I will only present two arguments in defense of folk psychology[24].

Now, it is important to remember that my interest surrounds the social-scientific accounts of social reality in general. This includes a dimension of *dynamics* and *relationships*, and so we are not talking only about an ontology (i.e. an account of what entities exist). Hence, in this work, I use mostly the term folk psychology and not folk ontology. But in virtue of being related, one way of defending folk *ontology* is relevant for this section. It has been argued that folk ontology (i.e. folk accounts of what entities exist in reality) is *warranted* by the evolution of our sensory organs, that have adapted to perceive real-world phenomena (Osborne, 2016). And furthermore, going into a radical opposite perspective to eliminativism, it has also been proposed that social-scientific theories must have an "adequacy requirement": they *must* be consistent with our folk ontology intuitions (Hindriks & Guala, 2015, p. 3). After all the considerations made in this chapter, however, positions defending the infallibility of folk psychology or an adequacy requirement seem to be rather reckless.

[23] For example, one line of discussion surrounds the viability of the theory-theory perspective. For more on this see Churchland (1989) and W. Ramsey (2003/2019, §4.2).

[24] So, in what follows I will not map all possible positions but just illustrate the debate to introduce my perspective. The point of presenting Churchland's eliminativism, in particular, was to show an extreme position from which to open relevant lines of discussion.

Against Unconditional Defenses of Folk Psychology

First, even with our very powerful sensory organs, we can be way off in our ontology of the world as, for example, the history of astronomical models of the cosmos since antiquity shows. In particular, it is relevant to point out that empirical research in psychology has shown that people may make systematic reasoning errors when making judgments (Kahneman & Tversky, 1996)[25], and even that under certain conditions people may generate false memories about the past or false perceptions about the present (Hirstein, 2009).

There is no reason, then, to assume that our folk accounts of individual action are necessarily warranted and correct. Second, even if in disagreement with them, we should give some credit to eliminativists for noting that scientific theories when sufficiently developed tend to be way more powerful in terms of explanation and accuracy than folk accounts of reality, and that the concepts that are part of these scientific theories often carry different or more elaborated expressions than the folk idea at its historical genesis. And third, considering again Hoyningen-Huene's argument around systematicity, folk accounts of reality are *less systematic than* scientific accounts in all the relevant dimensions of systematicity (e.g. by being less attentive to possible errors or biases in reasoning, or by relying on a non-systematic observation of reality).

These are certainly strong reasons against the infallibility of folk psychology that put into question its usefulness for scientific practice. However, in my view this debate is not so simple and discarding an unconditional defense of folk psychology does not imply accepting eliminativism. Even if one determines that certain social-scientific concept comes from folk psychology, as in *it has secured its place in a contemporary social-scientific framework even if it is non-systematic under the social-scientific framework's own methodological standards*, the a priori stance of eliminativism towards discarding all folk psychology does not seem entirely convincing, as I will discuss next. Most certainly it does not go in the line of carefully analyzing a scientific framework before sketching conclusions on its practice

[25] Kahneman and Tversky (1996) consider them to be "systematic" in the sense that these errors are consistent in their deviation from the correct answers in different empirical studies, and appear frequently in different subjects.

(which is my general approach in this work).

To continue the present discussion, I will now focus on the rejection of the idea that folk psychology is always false and useless, and call it *non-eliminativism*. In particular, I will argue in favor of a perspective claiming that folk psychology should be *revised* (not automatically discarded).

Revisionist Non-Eliminativism

In the Philosophy of the Social Sciences, the views on the value of folk psychology are less radical than eliminativism and are arguably non-eliminativist in general. Some authors such as Hindriks and Guala (2015, p. 3) have argued that probing a convergence between folk accounts of reality and a social-scientific theory might be fruitful and serve as a starting point to propose hypotheses (but not conclusions) about social phenomena. In particular, they argue that their theory of institutions[26] is "by and large consistent" with common sense and by this claim they defend that the folk concepts in social-scientific theories are not, in principle, a problem. Hindriks and Guala's approach is, then, non-eliminativist. This goes in line with Fleck's and Hoyningen-Huene's ideas on the origin of scientific concepts and disciplines: a strict eliminativism perspective is problematic at the genesis of a scientific discipline since, as discussed earlier, there may not be any alternative than to start working with everyday ideas about the world.

Furthermore, in his book *The Ant Trap*, Brian Epstein (2015, pp. 6-10) discusses what he calls *anthropocentrism*: an (in his view) unjustified and common sense assumption in many theories in the Social Sciences that, he claims, should be critically revised and possibly discarded[27]. Consider that in virtue of being unjustified and common sense, it can be safely argued that anthropocentrism is non-systematic (for example, because its claims are not systematically defended) and hence a folk idea as defined previously.

Now, while Epstein does not discuss the issue of common sense or folk con-

[26] As developed by Guala (2016b).

[27] Anthropocentrism is described as a metaphysical assumption stating that "the objects of the social sciences are built out of individual people much as an ant colony is built out of ants, or a chimpanzee community is built out of chimpanzees, or a cell is built out of organelles." (Epstein, 2015, p. 7).

cepts in relation to social-scientific frameworks in general, we can at least infer from his work a non-eliminativist, revisionist position: the common sense assumptions should at least be considered suspects of causing problems in a scientific framework. I consider this position as non-eliminativist because, in contrast with Churchland's approach, it seems that Epstein is not arguing in favor of *automatically discarding* folk ideas when they are identified as such. Instead, he is arguing that these ideas should be *critically revised*. In my opinion, this is a much more sensible position that goes in line with my general approach of analyzing a scientific framework before making claims about its practice.

2.2.3 Evaluating Folk Psychology

A revisionist non-eliminativism of folk psychology implies that instead of by principle rejecting or accepting folk psychology in social-scientific frameworks, all the non-systematic social-scientific ideas (e.g. that still resembles the folk idea at its genesis) should be *considered suspects* of causing potential problems. And of course, being such suspects, they should be carefully analyzed to see if the social-scientific framework using them could be modified to improve its practice. In the rest of this work, this is going to be my approach regarding the evaluation of any part of a social-scientific framework that could be identified as folk psychology.

To pursue the evaluative nature of this work, however, I still need to discuss more precisely what kinds of problems may be caused by the presence of folk psychology in social-scientific frameworks. It seems to me that there are at least two kinds of these. First, from the way I will identify folk psychology (as discussed in 2.2.1), it immediately follows that by contrasting non-systematic ideas with each social-scientific framework's own methodological standards (including their concepts of systematicity and more general values), we could identify problems in the sense that folk psychology may (a) hinder the progress of the social-scientific frameworks *in terms of their own methodological standards*. Coming back to Fleck's considerations, this evaluation does not mean that identified folk ideas are simply useless, but rather *unadapted* (and in need of further refinement or modification) given a social-scientific framework's current standards. Then, in this work, I will also argue that potentially there are problems that may (b) hinder the progress of the social-scientific frameworks *in terms of some more general scientific values* (e.g. the ones mentioned in subsection 2.1.3). So, to conclude this section, I will now discuss this second kind of problem.

Folk Psychology and Systematicity

Now, consider first that from the previous discussion it does not follow that if an account is not folk, then it necessarily becomes 'scientific'. I am leaving this point open because, again, I am not entering the debate on the demarcation problem. Again, what worries me is being too restrictive before even starting an analysis of social-scientific practice. Given the lack of consensus on what science is in the first place, I do not want to commit to classifying something as science if and only if it complies with some a priori criteria.

So, to evaluate social-scientific frameworks, my idea here is just to argue that certain features of Hoyningen-Huene's systematicity are *desirable* (not necessary nor sufficient) properties for any account that pretends to be a correct or accurate representation of reality, because they go (broadly) in line with the two general values of logical consistency and empirical consistency (as discussed above). My claim here is that, for scientific frameworks, *in the long run, a systematic (contextualized in the practice of each discipline) development of ideas lowers the probability of producing an inaccurate or inconsistent picture of the world*. Notice however that a lower probability does not imply impossibility: I am also leaving open the possibility that folk ideas are compatible with a correct or accurate account of reality.

Now, remember that it does not seem to be the case that folk psychological accounts are infallible by nature, even when they are about the own mental states of the person in question. In this line, I will not assume that folk psychology is (a) a true theory of the mind or (b) a scientific theory. Additionally, because of how I defined it earlier, folk psychology is not systematic (in the sense of HHS) and hence I will not assume that it is (c) a logically consistent set of ideas that agrees with empirical observations[28]. And I will neither assume that folk psychology is

[28] For example, when pushed too much in a certain way (e.g. in a psychological experiment),

(d) a single account, so I am leaving open the possibility that different communities or individuals may have different common sense ways of understanding the mind[29].

But now, if HHS is only a descriptive approach and each social-scientific framework has its own values, why is the infallibility of folk psychology in terms of consistency and accuracy a problem? Given the important differences in epistemic goals between social-scientific frameworks, could we have social-scientific frameworks whose values do not include, in any possible sense, the search for truth and accuracy? While Hoyningen-Huene may be reluctant to give a definitive answer to this question, this is where I extend his descriptive aims by committing to the two general values that should be part of the methodological standards of any scientific framework. I am committing to this last point since, otherwise (e.g. defending a relativistic approach), the evaluative aspect of my scheme would be limited to the problems of type (a) mentioned earlier. And, ultimately, I consider that this limitation would not allow me to fulfill one of the main motivation for this work[30].

So, I am attempting to present a more general way of evaluating social-scientific practice, in terms of the problems of type (b). But remember that, to do this, I will first discuss a social-scientific framework in detail and only after considering the lessons of this analysis I expect to be in a position to emit a general evaluation. In my view, this is not incompatible with Hoyningen-Huene's suggestion of carefully contrasting each particular social-scientific framework with everyday knowledge (instead of the Social Sciences in general). What I am doing is to

---

some folk ideas or ways of reasoning may show to be flagrantly inaccurate or simply produce false claims. The research on cognitive biases, such as that of Kahneman and Tversky (1996), is supportive of this claim.

[29] In my view, folk knowledge is not just given. It depends, at least partially, on the social context. What we may consider everyday knowledge about the world (and also about human action) may be very different from what is considered as such in other social contexts. This makes me a *pluralist* about folk psychology, a view that I think is more congruent with what social scientists observe.

For more on how common sense ideas and ways of reasoning may vary across different societies, see Hallpike (2011, Introduction and §3. Primitive thought).

[30] As outlined in the introductory chapter, remember that my main motivation for this work surrounds the identification of problems and, from here, try to think in ways of improving social-scientific practice.

take Hoyningen-Huene's suggestion as a first step, then (exceeding Hoyningen-Huene's descriptive aims) evaluating the framework in terms of its own values, and only then attempt to advance an evaluation of the framework's practice in a more general sense.

So, to recapitulate, my approach towards folk psychology in social-scientific frameworks will be that we should at the very least be careful of them. And that it would be better if we critically revised ideas as soon as they are identified as folk. With this in mind, in Chapter 3 I will revise the folk ideas and its potential problems in the context of a particular social-scientific framework. And then, building upon this, I will present a more general evaluation on the appropriateness of the social-scientific framework's standards to maximize the quality of its knowledge.

2.3 Descriptive and Evaluative Scheme

With all of these considerations in mind, I close this chapter with the following synthesis of the ideas discussed in the form of a scheme to (i) describe the historical development of social-scientific accounts of reality and (ii) characterize and evaluate this development given the historical description:

- Definition of the object of study: Describe what is meant by the *historical development of the account of a part of social reality of a social-scientific framework*, in terms of specific objects of analysis (e.g ideas, authors, texts, etc.) and how they will be selected

- Descriptive dimension: Follow the historical development of the social-scientific framework's account on the phenomenon of interest, focusing on finding clues about the features of thought style(s)

- Evaluative dimension (relative):

  - Characterize the thought style(s) in the object of study in terms of their identifiable features, such as their existential implications and methodological standards

  - Explore the historical development of the social-scientific framework for any potential issue under the light of the (synchronic) methodological standards identified

- Evaluative dimension (general):

  - Evaluate the development of the social-scientific framework in terms of the two general values of logical consistency and empirical consistency with robust observations

  - Revise and evaluate the role of folk ideas in the historical development of the social-scientific framework's account of the phenomenon of interest

As I mentioned in the introductory chapter, this scheme is a positive and particular suggestion on how to historically analyze and evaluate social-scientific accounts of reality. While there may be different ways of doing this, I have not found explicit alternatives. In the final chapter, some of my conclusions will surround the comparison of (a) what I propose in this work with (b) some examples with which I disagree at a methodological level. So, if so far the need of an *explicit conceptual scheme* is not clear enough, I hope that my final reflections will show its relevance.

Chapter 3

Example Implementation of the Conceptual Scheme

In this chapter I will present an example implementation of the descriptive and evaluative scheme developed.

As I will discuss in the next section, this is a *restricted* example implementation on the historical development of a type of Decision Theory[31], called Expected Utility Theory (EUT). In particular, surrounding its fundamental representation of individual decision-making based on the concepts of *utility* and *preference*. EUT has had an enormous influence in the Social Sciences, in particular in Economics, where it occupies a central role in ares such as macro-economical models and Game Theory (which is further applied in Economics but also in Political Science), as well as in the debates surrounding the notion of rationality[32]. Additionally, EUT is one of the few examples of a social-scientific theory that has been (i) clearly formalized in an axiomatic system, and (ii) put to experimental test, in a history with many nuances that looks inviting as an object of study.

So, in this chapter, I will first discuss the methodological features and constraints of the example implementation below, surrounding the use of Network Analysis to obtain a network of citations between authors. With the use of this

[31] What I have been calling so far *Decision Theory* (DT) can be understood as a broad term to describe different accounts of decision-making in Philosophy and the Social Sciences.

[32] This is a very popular topic, surrounding what rationality is and what it implies. In Economics, theories of rationality tend to build over some decision theory (such as EUT). I will not, however, discuss this broader subject here.

method, I produced a network of *detailed citations* between a restricted set of texts related to Expected Utility Theory, which used as a guide in for a description of the historical development of EUT's account of individual decision-making. Finally, considering both the methodological constraints of this example implementation and the (limited) historical description done, I will suggest some ways of evaluating the part of the development of EUT's account of individual decision-making I had access to.

3.1 Method and Design

In this chapter, I will use *Network Analysis* as a way to operationalize my object of study based on the definition of social-scientific framework, as I find this method compatible with the notion of *intellectual interaction* behind the definition of thought collective. While in principle my descriptive-evaluative scheme should be compatible with other methods of empirical research, based on the lessons learned during the implementation of my scheme, in the next chapter I will argue in favor of Network Analysis in particular.

Briefly put, Network Analysis is a research method to gather, process, and analyze empirical information about relations between entities. Because it is a general method accepting different instances of relations and entities, it has been used in different scientific disciplines for decades and there are many research communities basing their work on this method. Under its most basic form, empirical data is organized as 'nodes' (which represent entities) linked by 'edges' (which represent relations), together forming a *network* (hence the name). While this method may be used to study many kinds of phenomena, I will use a network of *texts as nodes* and *citations between texts* as edges. Then, when used to analyze social phenomena, this method is usually called *Social Network Analysis* (SNA), and hence in the rest of this work I will use this name[33].

Now, as discussed in subsection 2.1.2, the idea of thought collective behind social-scientific frameworks has a strong focus on the notion of *intellectual inter-*

[33] For an introduction to the basic concepts of SNA, see Wellman (1988).

action. While this last term may sound vague, a method such as SNA allows us to give more substance to it, since it has been developed to systematically analyze structures of social interactions. But regardless of the usefulness of SNA, its use implies practical challenges. This together with a commitment of being transparent with the delimitation of my object of study, during the research for this work I had to impose the constraints described in the rest of this section. The limitations of derived from these constraints taken into account for the description and evaluation presented below in this chapter, and reflections surrounding these limitations will be presented in the next chapter.

3.1.1 Constraint 1: On the Idea of Intellectual Interaction

Now, the notion of intellectual interaction is very general and it may be argued that it includes events such as informal conversations between social scientists, mail and e-mail exchanges, discussions in talks, and so on. However, in face of the many challenges implied by being too inclusive with the notion of intellectual interaction, the first methodological constraint introduced is that I will interpret this notion as *citation* in published texts. Behind this constraint there are two basic assumptions that I will not put into question in this work, since both seem reasonable enough and allow me to move forward. The first assumption is that, if a text T_B of author B has a citation to text T_A of author A, then author B must have read T_A and reflect on the ideas encoded there. The second assumption is that, at the macro-level, by this reflection leading to finally producing and publishing texts with new ideas, authors are, in a sense, intellectually interacting.

3.1.2 Constraint 2: On the Number of Texts

Now, an important advantage of using SNA as a method is that there are well documented techniques to gather the information necessary to construct a network from scratch. One well known and simple technique is *snowballing* (Heckathorn, 2011, p. 356): we start with a set of known nodes for our network in what we call a first "wave"; then we register the nodes linked to our first wave (e.g. people mentioned by the people in the first wave) as the second wave; then we repeat this process in successive waves until no new nodes appear or until some threshold of representativeness is achieved. For this implementation, I started by consulting different texts of Philosophy of Economics and Decision Theory that (at least partially) focus on the history of Expected Utility Theory for individual decision-making, and then registering all the texts they mentioned as nodes for a potential first wave.

However, because of the vast amount of texts that would be included in a potential second wave, I took the decision of limiting the size and elements of the network presented below. First, I limited the network to the first wave of nodes, which already covered an important amount of material. Unfortunately, this meant abandoning much of the most recent literature, a limitation to which I will come back in the final chapter. The reason is that analyzing a complete network (i.e. exhausting all necessary waves until no new nodes appear) in the way I have done here (categorizing the registered citations between authors by how they treat each other's ideas, as I will describe in a moment) proved to greatly exceed the time constraints for this work.

Additionally, I had a rather small number of exponents for some debates[34], which meant that I was having a largely unfair representation of them, and would have led to sacrifice depth in other debates. So, second an additional less systematic limitation was introduced by filtering out texts *not* belonging to one of the following topics in EUT's history:

- The early precursors of Expected Utility Theory
- The development of Political Economy
- The early twentieth century development of Economics
- The classical axiomatization of EUT in the mid-twentieth century
- Savage's and Jeffrey's extensions of classical EUT

[34] These debates surround: (a) non-descriptive uses of Decision Theory, as expressed by authors who explore a normative use (E.F. MacClennan; R. Bradley) or an interpretive use (D. Davidson; D. Lewis; J.L. Bermúdez); and (b) the problem of the individuation of the objects of the preference relation (J. Broome; P. Pettit; J. Dreiser).

- The debate on the Revealed Preferences Theory
- The Allais Paradox and the Preference Reversal Phenomenon
- The reactions to these phenomena in the late twentieth century

Regardless of this limitation, these topics still gave me enough material for an example implementation of the descriptive-evaluative scheme.

3.1.3 Construction of the Citation Network

With these constraints in mind, then, the main sections or chapters of the commentators' texts from where the nodes for the network were extracted were the following: Guala's chapters 5 ("Prediction") and 6 ("Elimination") (2005); Hausman's section 1.1 ("The emergence of economics and of economies") (2018); Thoma's sections 1 ("Decision Problems and the uses of Decision Theory"), 2 ("Representation Theorems"), and subsection 3.1 ("Interpretations of Utility") (2019); and Thoma's sections 1 ("Introduction) and 2 ("Behaviourism vs Mentalism about Preference") (Forthcoming).

The idea here is that, in virtue of being present in Philosophy of Social Sciences' accounts on aspects of the history of EUT, the authors mentioned by the commentators form a thought collective and presumably many of them share similar thought styles. In virtue then of most of them being institutionally recognized social scientists, then I expected that this selection of texts represented a part of a (certainly) much larger social-scientific framework.

To construct the network, the procedure was the following. First, I consulted the four commentators' texts and registered a first wave of nodes (i.e. EUT-related texts surrounding individual decision-making). Second, I filtered out the nodes not belonging to one of the topical categories mentioned above. Third, I consulted each node (i.e. text), registering the mentions to other nodes in the filtered network, and also the characteristics of this mention (e.g. if they agreed, disagreed, under which conditions, etc.). Fourth, I defined a property for the edges in the network, called "type of mention", with three possible values: agreement; disagreement; and neutral mention.

Fifth, I assigned one of three possible values to each edge in the network, based on the characteristics of the mention. "Agreement": if it could be described as a total or partial agreement and the author built on at least on a part of the ideas mention to propose new original ideas. "Disagreement": if it could be described as a critique in order to present an alternative (e.g. an alternative concept or theory). "Neutral": if it could be described as a simple mention, without building on or criticizing the ideas (even if the mention is positive).

The network thus constructed, which can be found in Appendix A, contains:

- 53 texts as nodes, registered in the Citation Network Texts section of the References
- 236 total citations between texts as edges, with the following distribution of *mention types*: 85 agreements (36.02%); 54 disagreements (22.89%); and 97 neutral mentions (41.01%)

Additionally, I have made publicly available the list of nodes and the list of edges I built manually[35], for anyone interested in controlling my implementation or extending the work I did. A final note to be made here is that I could not locate a copy (either physical or digital) of one of the texts (Chew & MacCrimmon, 1979), so this node sends no edges and I could not directly read the authors' ideas.

The constraints and assumptions mentioned here limit the power of my example implementation, such that I am not able to have conclusive remarks on the debate around EUT as a whole (and evidently much less about Economics as a whole). I believe, however, that what will be presented next is enough to be considered as a pilot or proof of concept of my descriptive-evaluative scheme. And that the lessons and discussions that stem from this exercise are enough to satisfy the objectives of this work.

With all of this in mind, in the rest of this chapter I will use the network to guide the historical description and evaluation of EUT's account of individual decision-making.

[35] Together with the digital image of the network, they can be found with the following DOI: https://doi.org/10.5281/zenodo.4005963
The edge list contains a "reason" column, where I briefly explain why I categorized the way I did some potentially unclear mention types.

3.2 Historical Description

In this section I will present the historical development of some of the most basic ideas behind the social-scientific use of Expected Utility Theory to understand individual decision-making. This will be done by reviewing both the ideas themselves and how they were developed historically as reactions to previous ideas, as suggested by the network of curated citations which will serve as a guide (see Appendix A).

In what follows, I will not necessarily follow a chronological order, but rather present the texts by topics as mentioned in page 35.

3.2.1 Early Precursors

While general reflections on the nature of human decision-making can certainly be found in ancient Philosophy, the oldest ideas mentioned by the four commentators as precursors of contemporary EUT (Thoma, 2019) are the inquiries on probability and risk in gambling, found in a correspondence between Fermat and Pascal (de Fermat and Pascal, 1654/1929). Pascal and Fermat's mathematical ideas are at the base of much of the mathematics of later EUT development, and both made further contributions in texts that do not appear in the network. In particular, their inquiries lead to the idea that (expressed in contemporary terms) when comparing gambles, we ought to consider not only the magnitudes of each outcome of the gamble (e.g. monetary prizes), but also how likely each outcome is, by multiplying the probability by the magnitude. This lead to the idea of calculating the *expected value* of a gamble as *the mean of the probability-weighted magnitudes of each possible outcome of the gamble*.

The next fundamental contribution came in Daniel Bernoulli's text (Bernoulli, 1738/1954), where he explores the relation between his concept of *moral expectation* and the notion of utility[36]. For Bernoulli, the moral expectation is a *quantity* representing the utility that gamble outcomes have for an agent *given its wealth*,

[36] It has also been argued, however, that these ideas were defended earlier by G. Cramer (Joyce, 1999). However, since Cramer did not appear in any of the four commentators' texts, consequently he does not appear in the network.

in contrast with considering only the probability-weighted magnitudes. For example, for two agents, one of which is very wealthy and the other very poor, a given amount of money may not make much difference for the former agent but may be very meaningful for the latter.

In particular, Bernoulli proposed a function to represent the decreasing utility (moral expectation) of each unit of money as wealth increases. But while this particular function has not been very influential, two ideas of this time will later become fundamental in Expected Utility Theory: - The notion of *utility*, which throughout history has meant to reflect the property of an entity (such as the outcome of a gamble, money, some consumer product, etc.) of satisfying a want of an agent. - What I will call *Bernoulli's thesis*, stating that we can, in principle, numerically represent utility Many of the neutral mentions of Bernoulli's text in later texts, represented in the network, reflect this point.. This thesis was considered, however, a relatively unimportant idea and sometime later explicitly rejected, until its mid-twentieth century recovery.

3.2.2 Political Economy

In contrast with the mathematical ideas of the early precursors, a different thread of discussion in eighteenth century philosophy gave rise to the discipline known as *Political Economy*, mainly through the works of the french Physiocrats[37] (Hausman, 2018, §1.1). Now, apparently much more influential than the Physiocrats themselves was the discussion of their ideas by David Hume (1752/1994), which where further developed by Adam Smith (1776/1937). The ideas of the latter on the relationship between the wealth of nations and the choices of individuals are considered to be one of the milestones in the development of contemporary Economics in general. The contribution of these authors regards mainly in shaping it as an independent discipline, and, as shown in the network, they do not appear as a particularly crucial step in the development of EUT.

Towards the nineteenth century, Political Economy was defined as the study

[37] No specific text from the Physiocrats is mentioned in any of the four commentators' texts, so they do not appear in the network.

of the individual pursuit of *pleasure or happiness*, and the consequences of this pursuit, by authors such as Bentham (Bentham, 1789/2007), J.S. Mill (Mill, 1861/1998; Mill, 1843/2011) and Jevons (Jevons, 1871) (Hausman, 2018, §1). As is the case for present day Economics, Political Economy covered many topics, from individual decision-making to macro-economics. Jevons made the notable contribution of clarifying the idea of *marginal utility*, leading to what some commentators call the *marginalist* or neoclassical revolution in Economics (Hausman, 2018, §1.1; Guala, 2005, p. 16). In short, the marginal utility is meant to represent that, for an agent, the utility of an additional unit of a good will be inversely proportional to amount of units of said good[38]. Jevons' text is also notable because it heralded what I consider to be a major shift in the historical development discussed here. This author repeatedly criticizes some political economists (including Smith and Bentham) of having ideas implying quantities, but having a very poor mathematical treatment of these quantities.

3.2.3 Increasing Formalization and the Preference Relation

With the shift towards increasing formalization of ideas in Economics, towards the beginning of the twentieth century there is an explicit rejection of the previous assumption that agents maximized their pleasure or happiness by their choice behavior. While Hicks and Allen's text (Hicks and Allen, 1934), does not seem to be as influential as V. Pareto's text (1909/1971), on which it heavily relies[39], it is representative of this shift. Both Hicks and Allen and Pareto focus on the idea of representing the preferences of agents in a *preference scale* of alternatives that the agent faces. This is because, as Pareto argues (and Hicks and Allen accept), from empirical observation we may infer an agent's preference scale, but neither from experience nor from the preference scale we may infer the numerical amount of utility that agents assign. We see, then, a rejection of Bernoulli's thesis

[38] So, for example, if I want some salt but do not have any, buying some will be more useful than keeping the as money the price of the salt I want. But after I satisfied all of my want for salt, buying an additional unit of it will not be as useful for me as keeping as money the price of this additional salt.

[39] Both texts note that they build on important definitions advanced by F.Y. Edgeworth, a political economist not appearing in the network.

during period. Pareto's approach contrasts with the previous focus of political economists on (externally) building a scale of how much different outcomes or goods promote an agent's wealth or happiness (Hausman, 2018, §1.1), and then study its implications.

These discussions were also accompanied by a mathematical rigor that heavily contrasted with what Jevons had criticized some decades earlier. Starting from Pareto, in almost all the texts in the network proposing a new theory or theoretical concept propose we are presented with a formal axiomatization of the basic ideas and a mathematical derivation of its consequences (including sometimes an annex of mathematical proofs).

Briefly, what Pareto did was abandoning the notion of utility ("ophemility", as he called it) as a primitive, and instead tried to derive models of equilibrium from preference scales. There were many reasons for this abandonment, including that it was considered as vague, empirically inaccessible, and impossible to formalize. Regarding this last point, Pareto showed that if some sort of utility function related to the preference scale exists, it is not unique, while Hicks and Allen derived from these ideas that, additionally, this function is not determined. It seems to me that this problem stemmed from the fact that the relata of the preference relation were usually assumed to be goods, such as bread or wine. For any agent, different goods may be competitive between each other (i.e. they satisfy similar wants, so they may be replaced by another competitive good) or complementary (i.e. they are not replaceable with one another). So, when trying to derive a utility function, not being able to systematically determine which goods are competitive or complementary between each other yielded non-determined utility functions (i.e. because the output changes depending on if the goods are either competitive or complementary).

In any case, the shifts on the fundamentals of Economics of this period are further marked by Robbins's paper, where he defines the discipline as the "science which studies human behaviour as a relationship between ends and scarce means which have alternative uses" (1932/1995, p. 75). Aside from abandoning the previous focus on happiness and pleasure, we see also a shift to an increasing focus on more abstract relata for the preference relation, in contrast with the focus on goods which we still find in Pareto's text[40]. As we will see, the texts that follow take the relata for the preference relation to be things like exclusive outcomes, states of the world, or possible actions, depending on the author's approach.

3.2.4 Classical Axiomatization of Expected Utility Theory

The particular version of Decision Theory proposed by von Neumann and Morgenstern (1944) is now considered by many, as suggested by the amount of mentions it receives and the commentator's opinions, as the base of the classical axiomatization of Expected Utility Theory.

Notably, in what some call a neo-Bernoullian approach (Allais, 1953/1979), von Neumann and Morgenstern (VNM) broke with the earlier paretian approach to preference and brought back the idea of a quantifiable utility (which VNM call "numerical utility"). In a clever move, VNM compared the problem of quantifying utility to that of quantifying heat: the fact that the pre-quantified notion of heat was vague and too subjective to compare[41] did not make impossible constructing temperature scales. Interestingly, VNM reached this conclusion by in fact building on Pareto's ideas, not discarding them. In my reading, the crucial move for their proposal was restricting the relata of the preference relation to be always competitive, rather than either competitive *or* complementary with each other. So, in VNM's EUT, the relata of the preference relation is more intuitively interpreted as exclusive possibilities (such as outcomes, actions, or events) such that the realization of a particular possibility implies that the other possibilities do not occur.

VNM's text became very influential almost immediately, generating numerous debates on the exact meaning of some of their ideas and implications. From the result of these debates, other authors (Marschak, 1950; Herstein and Milnor, 1953) proposed what many consider to be refinements of VNM's ideas (Guala,

[40] As a side note, Robbin's text is also notable because he advances a much cited (and now classical) defense of Economics as a scientific (descriptive, positive) discipline, in parallel to other uses for it such as insights for policy-making.

[41] As in, what is "cold" for me may not be "cold" for another person.

2005; Camerer, 1995), into what I will call *the classical axiomatization of EUT*. In a nutshell, classical EUT shows that any agent whose preferences comply with certain axioms (that will be discussed below) can be treated as an *expected utility maximizer*: an agent whose decisions and choices are aimed towards obtaining the maximum expected utility, given their preferences and the potential outcomes.

I present here a brief description of two of EUT's axioms that will be discussed later, as they appear in Guala's book (2005, pp. 98-100). Consider first the following notation. Let the letters x, y, z represent *options* or *outcomes*; the letters p, q represent *probabilities*; let the symbol $\succ$ denote the strict preference relation; and let U represent a utility function. With this, it follows that we can represent: a binary lottery x as $py + (1-p)z$ [42]; phrases of the form "option x is strictly preferred to option y" as $x \succ y$; and the (numerical, quantified) utility of an option or outcome x as $U(x)$.

The axioms then are:

- (A1) Strict preference, $\succ$, imposes a *weak-ordered relation* on the elements of its domain (i.e. relata such as options, outcomes, lotteries, etc.). This implies two notable properties:

$$\text{Asymmetry: } (x \succ y) \rightarrow \neg(y \succ x)$$
$$\text{Transitivity: } (x \succ y \wedge y \succ z) \rightarrow (x \succ z)$$

- (A2) Independence: For all p such that $0 < p \leq 1$

$$(x \succ y) \leftrightarrow (px + (1-p)z \succ py + (1-p)z)$$

Then, that an agent's behavior respects these axioms implies that they are *maximizing their expected utility* (EU), defined as the sum of the products of an outcome's utility times its respective probability of occurring:

$$EU = \sum p_i U(x_i) \tag{3.1}$$

[42] Assuming in this case that: (i) y and z's magnitudes are expressed as a numerical value; (ii) $y \neq z$; and (iii) y and z are mutually exclusive and the only possible outcomes of x.

From what is presented up to this point, VNM's approach allow the derivation of the *representation theorem* of classical EUT: *if an ordering relation* $\succ$ *satisfies all of the necessary axioms*[43]*, then there exists a utility function (defined on outcomes)* U *such that for any lotteries* x *and* y

$$(x \succ y) \leftrightarrow EU(x) > EU(y) \tag{3.2}$$

A point to note here is that an alternative but equivalent formalization may be advanced in terms of the *weak preference relation* used to represent the statements of the form "x is *at least as preferred as* y", usually denoted by $\succeq$. As we will see, later authors have proposed other representation theorems usually with the same starting point of a strict preference imposing weak order on the elements of its domain, with the usual aim of deducing an equivalent of the last formula (Thoma, 2019, pp. 66-67). While the classical EUT possesses many more features, the ones described here are arguably the most fundamental and will be enough for the discussion to follow.

Finally, it is worth summarizing here previous ideas and approaches on which much the classical axiomatization rests:

- Bernoulli's thesis
- The formalization of the ideas of preference scales and the preference relation
- The value of mathematical and logical rigor
- The abstraction of decision theoretical ideas by defining non-material relata as the domain of the preference relation

3.2.5 Two Extensions of Classical EUT

There are two important branches of further refinement of classical EUT that are worth mentioning, since they have been mentioned in some important debates.

[43] That is, (A1), (A2), and others that I have left out.

On the one hand, Savage (1954) developed what is sometimes called *Subjective Expected Utility Theory* (SEUT). His text builds on an early representation theorem advanced by F. P. Ramsey's (1926/1988) which allows Savage to deduce *probabilities* directly from the preference order. In Savage's scheme, these probabilities represent "subjective degrees of belief" (Guala, 2005, p. 98) of agents towards *possible states of the world*. Additionally, in SEUT a utility function is also derived from the representation theorem of preferences, which is a function *over possible outcomes*. These possibilities are in turn a consequence of the agent's actions, but dependent on the state of the world. Both SEUT's utility function and preference relation comply with the classic EUT axioms, so the formula 3.2 is deducible from SEUT's axioms.

While SEUT largely builds on classic EUT, there are important differences. In the classic approach, the probabilities of occurrence of outcomes or options are given by the characteristics of the lotteries under consideration, while in SEUT's approach probabilities are instead derived from the preference scale of agents. SEUT is usually interpreted as treating probabilities as *uncertain*, to represent the real-life scenario of decision-makers who may not be able to calculate how likely it is that they are in a particular state of the world. For these reasons, classical EUT is sometimes considered to be a kind of decision theory *under risk* (since we only know the outcomes' probabilities of occurring, not which one will occur), while SEUT is considered as a kind of decision theory *under uncertainty*.

On the other hand, Jeffrey (1965/1983) builds on the criticism done to SEUT (a part of which I will review below) and his own criticism to propose an alternative EUT-based theory. Now, Jeffrey revised his original 1965 text and, by building on the mathematical ideas of Bolker (1966), he published an improved second edition of the text in 1983[44]. Jeffrey's approach draws heavily from logic and bayesianism, with the aim of exploring concepts as *subjective probability* and *desirability* (the latter of which replaces the concept of utility). Years later Joyce (1999) further developed this theory. He praised the Jeffrey-Bolker approach for

[44] Regarding the network, all mentions of this text are on the second edition, which apparently has been more discussed than the original.

being immune to most of the serious criticism made to Savage's SEUT, while keeping the approach of treating probabilities as not given.

Jeffrey and Joyce's texts are an interesting addition to the network, since they are instances of philosophers formally building on the logic and mathematics of EUT for their philosophical inquiries. In this process, Jeffrey in particular discussed enough of the fundamentals of EUT to call the attention of some other authors in the network. And while Joyce presents his text basically as an improvement of Jeffrey-Bolker's ideas, unfortunately the limitations of my network impede an assessment of any potential dialogue that it generated later, since none of authors in the network mention this text. In any case, Jeffrey-Bolker's and Joyce's ideas are not discussed in detail by any of the authors in the network, so I will not further explain their approach here.

3.2.6 The Debate on the Revealed Preferences Theory

In parallel to the development and extensions of the classical EUT, further reflection on fundamentals led to other kind of discussion. The formalization of the preference relation first, and then the classical axiomatization of EUT, led to debates on how to properly obtain empirically relevant claims from them.

In his 1938 text, Samuelson outlines the main features of what later will be known as the *Revealed Preferences Theory* (RPT). This was later refined in his 1947 book, which also appears in the network. Samuelson builds heavily on the ideas of Pareto (1909/1971), Hicks and Allen (1934)[45], to develop a theory from which to obtain "operationally meaningful propositions" from the idea of utility-maximizing behavior. The key idea here is that the abstract preferences of agents (inferred indirectly by economists and decision theorists) are "revealed" by the agents' (real, observable) choice behavior.

In two later texts, Little presents a defense and refinement of Samuelson's main ideas (1949; 1957), which is arguably more extreme. This is because Little (1949) shows that if the observable choice behavior of an agent complies with

[45] And an additional another 1939 book by Hicks (*Value and Capital*) that does not appear in the network.

certain consistency conditions, then this *behavior* (not just the preferences) can be represented as expected-utility-maximizing (as in classical EUT). Further refinement of these consistency conditions is due to Houthakker (1950) and Afriat (1967). A few years later, an axiomatization and systematization of the fundamental ideas around Revealed Preferences so far is presented by Sen (1971). The ideas here, in particular Little's, may be understood as implying *reductionism* over preferences, which are understood as completely determined by the observable choice behavior of agents.

Now, reductionism has been put into question many times and, from the texts in the network, we can identify at least two sources of criticism. First, conceptual criticism to reductionism's assumptions and implications from what is a different interpretation of RPT. An early (in the network) example of this is Sen (1973), on the grounds that the reductionist assumptions on people's behavior are unrealistic. Notably, Sen also argues that (intuitively) it is unlikely that behavior is the only valid source of information about preferences, an idea maintained and refined in a later text (1993). This leaves open the possibility that there are other mental entities not described by EUT[46] that play a role (in conjunction with preferences) in determining the observable choice behavior of agents. Under this reading, these ideas are still compatible with a less strong interpretation of RPT that claims that preferences are reflected by observable choice behavior *to some extent*, but that they are not entirely determined by it (Guala, 2005, p. 91). I will refer to this as the mentalism-compatible interpretation of RPT (or *mentalism*, for short).

Later, and building on many of Sen's ideas, Hausman's (2000; 2012) presents harsh criticism against reductionism, and further defends the use of including other mental entities (than the ones included in EUT) for Decision Theory and Economics[47].

[46] Beliefs, for example, are often mentioned entities in this debate. Motivations and objectives are also sometimes mentioned.

[47] I should mention here that in these texts Hausman uses the label "(original) Revealed Preferences Theory" for what I have been calling "reductionism". He then mentions that what I call "mentalism" is in fact a different project, in a convention that seems consistent with Thoma's texts (2019; Forthcoming). This contrasts, however, with Guala's (2005) understanding of RPT as a more neutral theory with different interpretations (reductionism and mentalism), on which I

An important note here is Gul and Pesendorfer (2008) as the last explicit defense of reductionism in the network, in a rather unclear text about their basic understanding of Economics as a discipline. Dietrich and List's (2016) response to Gul and Pesendorfer reconstructs their position as implying that the "body of evidence for any theory in economics is restricted to agents' choice behaviour." (Dietrich and List, 2016, p. 266)[48].

On the basis of this interpretation, we can then understand Gul and Pesendorfer's rejection of mentalism on the grounds that mental entities not described by classical EUT are not empirically relevant for Economics. As noted by Dietrich and List, this is certainly an arbitrary restriction, and the authors defend mentalism by concluding that economists do not really have good arguments to reject by principle the use of other mental entities. Even the chronologically most recent text in the network (Vredenburgh, 2019) intends to defend an approach closer to reductionism, but is not able to express an explicit counterargument against Dietrich and List. These last texts show that the debate between mentalism and reductionism is still ongoing, even if one of the four commentators argues that most economists accept a mentalism-compatible interpretation of RPT today (Guala, 2005, p. 91).

The second kind of criticism to reductionism is related to the debates generated around experimental results that seem challenge EUT's representation of individual decision-making[49]. This kind of criticism is involved in a more general criticism to classical EUT and SEUT, as we will see next.

---

base my labels.

[48] Dietrich and List call this the "epistemological revealed preference thesis", in contrast with the "ontological revealed preference thesis" which is a different candidate to interpret Gul and Pesendorfer's text. However, on the face of what economists do and how they express themselves, Dietrich and List conclude that the ontological thesis is not a charitable reading of Gul and Pesendorfer, 2008.

[49] Additionally, some criticism has been done on the basis that reductionism does not really work with an interpretive use of EUT (Joyce, 1999; Thoma, Forthcoming), but I will not discuss this issue here.

3.2.7 Two Notable Empirical Results

Since very early after its classical formulation, the fundamentals of EUT have been put into question from various fronts on empirical grounds.

A first notable critique is presented in Allais' 1953 text (1953/1979), where the now called *Allais Paradox* (AP) (Camerer, 1995, pp. 622-632) is formulated. For an AP to occur, agents are presented with two sets of choices X and Y. In set X, one may choose between a lottery X1 with a probability of 1 of winning 1 million francs, and a lottery X2 with probabilities (p) 0.89 of winning 1 million francs, 0.1 for 5 million francs, and 0.01 of winning nothing. Consider now for a moment that, in any case, there is a probability of 0.89 of winning 1 million francs, and thus for EUT the difference between X1 and X2 is defined by what happens with the remaining 0.11 probability. Let m mean "millions", and then we have that in X1 what remains is a p of 0.1 for 1m F and 0.01 for 1m F, while in X2 we have a p of 0.1 for 5m F and 0.01 for 0 F. Let also $x1 = (0.11)(1mF) + (0.01)(1mF)$ and $x2 = (0.1)(5mF) + (0.01)(0F)$.

Since Allais' text, it has been repeatedly shown that some (risk averse) people will choose X1 over X2, which under classical EUT is represented as

$$x1 + (0.89)z \succ x2 + (0.89)z$$

From this, the independence axiom (A2) of classical EUT leads us to conclude that $x1 \succ x2$, by telling us that the value of z is practically irrelevant for this preference (as long as its equal in both sides of the preference relation).

Then, in the second set of choices Y, agents may choose between Y1, where there is a p of 0.9 for 0 F and 0.1 for 5m F, and Y2 where there is a p of 0.89 for 0 F and 0.11 for 1m F. Notice now that in any case, there is a p of 0.89 for 0 F. So, we may represent what happens with the remaining probability of Y1 as $(0.1)(5mF) + (0.01)(0F) = x2$ and what happens with the remaining probability of Y2 as $(0.11)(1mF) + (0.01)(1mF) = x1$. The Allais Paradox occurs when an agent who chose X1 over X2 later chooses Y1 over Y2, which is represented as

$$x2 + (0.89)z \succ x1 + (0.89)z$$

This seems to happen *only because there is a different value for z, even if it is always equal in both sides of the relation*, leading us to conclude now from independence (A2) that $x2 \succ x1$. This last implication contradicts the asymmetry property implied by the preference relation imposing a weak order (A1), which is usually considered to be a more fundamental axiom. In my reading, it is for this reason that in the literature the AP is often presented as a challenge to (A2) and not (A1).

From these results, Allais criticizes the classical axiomatization of EUT on the grounds that the independence axiom is being empirically violated under certain conditions. He also criticizes Savage's SEUT *sure-thing principle*, which is SEUT's equivalent to the independence axiom. Overall, later mentions of Allais' text unanimously accept the paradox as a legitimate phenomenon and concern for EUT, in particular after later experimental results showed that the paradox was robust even with less extreme values[50].

A different but even more problematic (for classical EUT) line of experimental results surround the works of the psychologists Lichtenstein and Slovic, who tested hypotheses of their own 1968 paper (Slovic and Lichtenstein, 1968) on a series of experiments carried out some years later (1971/2006b; 1973/2006a). The results of these experiments, if interpreted assuming a very basic weak ordering preference relation (as most economists did even before VNM), suggested the existence of a phenomenon called "preference reversals" (PR).

The basic form of this kind of experiments goes as follows. Let *P-bets* be bets with a high probability of winning a small quantity of money, and a small probability of losing a smaller amount; and let *$-bets* be bets with a small probability of winning a large amount of money, but a moderate probability of losing a small amount. Then, let there be a *choosing task*, where subjects choose one alternative in a pair (or series of pairs) made of one P-bet and one $-bet, and a *pricing task*, where subjects place prices on $-bets and P-bets separately. To form the P-bet and $-bet pairs, the researchers combine the prizes and probabilities in such a way that their expected (monetary) values of both bets in the pair are similar. An

[50] Unfortunately, none of these experiments is present in the network, but they are further discussed by Camerer (1995, pp. 622-624).

example of such a pair would be the following, the expected monetary value of both bets being around $ 3.85:

- P-bet: win $ 4 with $p = 35/36$; lose $ 1 with $p = 1/36$
- $-bet: win $ 16 with $p = 11/36$; lose $ 1.5 with $p = 25/36$

The PR phenomenon occurs when an agent chooses one of the pairs of bets during the choosing task, but then sets a higher price to the other option of the pair of bets during the pricing task, in what is called a "price-choice reversal".

There have been many suggested implications this observed reversals. Lichtenstein and Slovic originally concluded that these results showed either: a violation of the transitivity of preferences, hence a violation of the preference relation as a weak order (A1); or a violation of a more basic assumption, which in the network is not made explicit until this point, known as *procedure invariance*. This assumption states that all economically relevant behavior can be used as evidence for inferring the preference scale of agents (which is unique) (Guala, 2005, p. 94). Thus the preferences inferred should not be different even when asking agents to perform different kind of tasks that arguably reflect their preferences (such as choosing or pricing). A formalization of the PR phenomenon, using an extended notation over the one used for classical axiomatization, may go as follows[51]:

$$\text{(Procedure invariance) } \succ = \succ_{choosing} = \succ_{pricing}\text{, and}$$
$$\text{(Price-choice reversal) P-bet } \succ_{choosing} \text{ \$-bet } \succ_{pricing} \text{ P-bet}$$

In my view, the assumption of procedure invariance is more basic than even the less extreme mentalism interpretation of revealed preferences, and is a principle strongly defended by the majority of economists in the network. The reason is that *if* procedure invariance is not correct, the claim that agents posses a *stable* structure of preferences[52] can be put into serious question (since the preferences would be context-dependent), and with it basically Economics' whole contemporary representation of individual decision-making (Guala, 2005, pp. 93-94).

[51] A similar but simpler formalization is presented by Guala (2005, p. 94)

[52] Which is what allows to understand the preference scales as a weak order (A1), or some other logical order.

For all these reasons, the Allais Paradox and the Preference Reversal phenomenon have caused a great deal of controversy and debate, as we will see next.

3.2.8 Revision of EUT

As Camerer resumes, price-choice reversals have generated three kinds of reactions, all of them reflected in the network in different degrees. It is argued that they are either: 1) an artifact, either as a distortion caused by the instruments of observation or as an effect unwillingly introduced by the experimenters[53]; (2) a violation of transitivity, and hence of preference scales as a weak order(A1), or other of the classic EUT axioms; or (3) a violation of procedure invariance, and hence of the assumption that individual preferences form a stable structure.

Building on the practical knowledge of experimental economics of the time, Grether and Plott (1979) replicated the original PR experiments controlling for thirteen alternative explanations, which included controlling for artifacts introduced by the instruments of observation (e.g. how the questions were asked). At the level of the network, many mention this paper as the proper introduction of the PR phenomenon into economic literature, since it was done considering many of the common concerns in experimental economics. Finding that PRs were robust to the controls introduced, and against their own expectations, the authors concluded that their experimental results not only suggest the intransitivity of preferences, but that "the uniformities in human choice behavior which lie behind market behavior may result from principles which are of a completely different sort from those generally accepted." (1979, p. 623).

Around the time of Grether and Plott's paper, various authors showed serious concern for the by then well-known Allais' Paradox and the violation of the independence axiom (A2). These authors (Chew and MacCrimmon, 1979[54]; Machina, 1982; Quiggin, 1982; Yaari, 1987) proposed different non-expected utility theories

[53] This distinction of different kinds of artifacts is due to Guala (2005, p. 96).

[54] As mentioned earlier, I was not able to find a copy of this text, so in the network they do not send any edges to other nodes and I only had indirect access to its contents (i.e. through the mentions made in other texts).

that some later authors call Ω-theories. While there are differences in their approaches, the common property of these theories is not including independence as an axiom. For example, Quiggin's theory contains classical EUT as a special case, in the sense that under certain conditions Quiggin's theory will make the exact same predictions and represent agents' decision-making in an equivalent way. And notably, under other conditions, Quiggin's theory will predict that agents violate independence.

Furthermore, adding to what is already the most complicated section of the network, as a critical reaction to the Ω-theories, Segal (1988) introduced the PR phenomenon to the discussion. He argued that the challenges to EUT posed by PRs do not show a violation of independence, but a violation of an axiom called *compound lottery reduction* (Camerer, 1995, pp. 656-657) (or, simply, the "reduction axiom"). Camerer explains that this is an either explicit or implicit axiom in most utility theories, which states that "whether a lottery is described as a compound gamble with several probabilistic stages or as a single-stage gamble should not affect preference." (1995, p. 656).

Now, building on the Ω-theories, both Holt (1986) and Karni and Safra (1987) suggested that a conjunction of (i) the violation of independence and (ii) artifacts introduced by experimenters, may result in agents *behaving as if* PR was a real phenomenon, when in fact it is not. In particular, Karni and Safra argued that *if* agents obeyed reduction but violated independence, *then* PRs could have been an introduced artifact by experimenters through the use of some experimental mechanism such as the Becker-DeGroot-Marschak (BDM) mechanism or the Random Lottery Selector (RLS) mechanism[55]. which, as a whole, caused the agents subject to these mechanisms to *not* reveal their true preferences. This convoluted alternative explanation to PRs implied that if agents isolated lotteries by *violating* reduction, then both BDM and RLS *allowed* for true revealed preferences (Camerer, 1995, p. 657). This implication is curious because revealing true preferences is what both mechanisms were designed for in the first place, assuming

[55] Since these alternative explanations of the PR phenomenon do not appear to be correct, I will not describe these mechanisms here. For a more detailed discussion on this topic, see Guala (2005, pp. 97-101)

classical EUT and the reduction axiom.

So far, these alternative explanations of PRs had no empirical support, since the original PR experiments were simply not designed to control for all the factors involved in these explanations (Guala, 2005, p. 108). In two 1990 papers, Safra, Segal, and Spivak probed further Karni and Safra's ideas, but mention that, since all Holt's, Segal, and Karni and Safra's competed as alternative explanations, further experiments were necessary. More notably, Starmer and Sugden's (1991) experimental results were strongly against the hypothesis that the experimental mechanisms prevented agents from showing their true preferences. Later, Keller, Segal, and Wang (1993) suggested that their experimental results contradicted the argumentative thread of Holt, Karni and Safra, and Safra et al., giving instead evidence in favor of Segal's and Starmer and Sugden's ideas.

Note that the discussion on these alternative explanations largely ignored the original Lichtenstein and Slovic's suggestions of PRs being caused by either a violation of transitivity or procedure invariance. From what observers of this dynamic suggest (Camerer, 1995, p. 674; Guala, 2005, p. 126), it seems that this over-focus on the experimental mechanisms comes from the fact that the text that introduced PRs to Economics (Grether and Plott, 1979) relies on them, and their malfunction was the least threatening alternative to explain PRs. By doing this, however, economists ignored that the original PR experiments by Lichtenstein and Slovic had already controlled for the BDM mechanism.

In the network, the discussion on the original conclusions from the PR experiments is continued by Tversky, Slovic, and Kahneman, who claim that their experimental results show that PRs "cannot be adequately explained by violations of independence, the reduction axiom, or transitivity" (1990, p. 204). Tversky et al.'s experimental design is a refinement of what some consider not proper PR experiments (Cox and Epstein, 1989), which tried to control for the effect of the experimental mechanisms used. They further argue that the violation of procedure invariance is a more consistent explanation for PRs, and bring back Grether and Plott's (1979) original suggestion that the PR phenomenon puts into question the whole of Economics' account of individual decision-making. The last two texts in this subsection (Camerer, 1995; Slovic, 1995) both agree that, given the accumulated evidence on the robustness of the PR phenomenon throughout the last two decades and a half, the fairest conclusion is that it is most likely a real phenomenon and that they keep being a central challenge to EUT and its basic assumptions on the preference relation.

At this point, the commentator who gave the most detailed discussion of this part of Economics and EUT history (Guala, 2005) stops mentioning more texts in this line. Thus, here ends my example implementation of the descriptive aspect of my scheme on the development of EUT's account of individual-decision making, with the constraints mentioned earlier.

3.3 Evaluation

In this section, I will exemplify the evaluative aspect of my scheme by implementing it on the historical development described above. Since the historical description above has important constraints, this evaluation will be limited to the part of EUT's history that I described.

With this in mind, I will first characterize the thought styles I found in the network in terms of their evident features given the previous historical description. Then, I will suggest an evaluation of the main developments of EUT's account of individual decision-making relative to these thought styles. Finally, I will outline some ideas regarding a more general evaluation of this historical development considering (a) the two general values mentioned in subsection 2.1.3, and (b) the role of folk psychology in this development.

3.3.1 Characterization

Regarding the network's contents, from the start we can already differentiate some differences in the thought style of the texts analyzed. On the one hand, Fermat, Pascal, and Bernoulli's, as mathematicians, their methodological standards were tied with deductive reasoning. Bernoulli's thesis however existentially commits to utility as a possible real entity, bridging his formalism with an empirical idea. On the other hand, the first political economists (as criticized by Jevons, 1871) are less concerned with mathematical rigor and instead focus on a conceptual discussion about more immediate economic phenomena. At this point, it is evident that these two groups do not carry a unified thought style, and do not even compose a clear thought collective since intellectual interaction between both parts is rather weak.

This changes after Pareto's text, which marks a stark contrast with the previous political economy and a shift towards what we read in later texts. In a sense, this seems to be a unification of what we had before: a mathematical treatment of economic ideas. From here, all the economists in the network after Pareto show clear concern for the *representation of knowledge* dimension of systematicity in particular, by carefully framing their texts around explicit axioms and discussions of their implications. I will label this broad though style as *formalism*, which is furthermore a unifying force which taken together with the (at least indirect) intellectual interaction between the economists on the network as a whole, leads me to conclude that economists after Pareto compose a single, broad social-scientific framework. Even when in disagreement, for example surrounding the interpretation of the Revealed Preferences Theory (RPT), economists will present a formalization of their ideas and discuss them in those terms (Samuelson, 1947; Sen, 1993). Similarly, a part of economists' responses to the experimental challenges to the classical EUT have been formal alternative explanations or theories (e.g. Machina, 1982; Karni and Safra, 1987).

After Allais, a subset of formalists' texts starts to show a growing concern regarding the empirical challenges to classical EUT. This suggests the emergence of a sub-thought style of formalism containing further methodological standards (i.e. not an almost exclusive focus on mathematical rigor), which I will label "empiricism". This thought style is arguably carried by all of the authors who seriously considered Allais's or Lichtenstein and Slovic's arguments as a challenge for EUT. This includes authors in the line of testing the problematic phenomena with further experiments (e.g. Grether and Plott, 1979; Safra et al., 1990a), as well as also those who advanced alternative theories to account for the evidence accumulated (e.g. Machina, 1982; Quiggin, 1982). By implicitly defending the importance of empirical evidence for EUT and Economics as a whole, in my view empiricists are concerned the *descriptions* and *defense of knowledge claims* dimensions of systematicity. As a note, the most extreme empiricists in the network, judged by the potential impact of the conclusions they draw from empirical evidence, were originally formed as psychologists (E.g. Lichtenstein, Slovic, and Tversky).

Now, my description suggests the existence of a second sub-thought style of formalism, opposed to empiricism, which I will call "traditionalism". Authors carrying this thought style argue that the way economic theory represents individual decision-making cannot be judged by how well this describes individual reality. Different reasons may be advanced to defend this claim, one being that this description is outside the scope of Economics (Robbins, 1932/1995), and that Economics has no way of making and testing predictions of non-economic phenomena (Gul & Pesendorfer, 2008). Traditionalists are also formalists, but neglect the challenges posed to economic accounts of individual decision-making by evidence produced and discussed even by other economists, in what is a clear tension between both thought styles. While traditionalism has few exponents in the network (e.g. Robbins, 1932/1995; Little, 1949; Little, 1957), and only one recent defender (Gul and Pesendorfer, 2008), some philosophers of the Social Sciences (Epstein, 2015; Guala, 2005) suggest that this thought style is popular in Economics as a whole[56].

Lastly, there are some authors in the network who, by the contents of their texts, do not seem to carry either of the three thought styles mentioned above. These authors are mainly philosophers who defend mentalism on RPT (e.g. Hausman, 2000; Dietrich and List, 2016). These texts are mainly concerned with the challenges to EUT and RPT, and their implications, in a very broad revisionist spirit. Unfortunately, because of the limitations of this example implementation, no reaction of social scientists to these texts is shown.

[56] A notable exponent, which unfortunately does not appear in the network, is Milton Friedman.

3.3.2 Relative Evaluation

Now, one useful way of evaluating the development of EUT's account of individual decision-making is in terms of *reactions* of different authors to the challenges posed to fundamental ideas.

Even with the apparent influence traditionalism, as we have seen, in the network both the Allais Paradox and the Preference Reversal phenomenon have been thoroughly discussed. Remember here the idea that there has been an overfocus on artifacts at this point (Camerer, 1995, p. 674; Guala, 2005, p. 126), which seems to have come from the fact that the experiment introducing PRs to Economics (Grether and Plott, 1979) included the much discussed BDM mechanism.

On the one hand, we could argue that the problem was the lack of consideration for psychological experiments on the part of economists, since it had already been suggested that PRs were not related to the BDM mechanism (Lichtenstein and Slovic, 1971/2006b). However, the fact is that, as discussed in the previous chapter (Della Porta and Keating, 2008, p. 23), social-scientific frameworks have different epistemic goals, and also methodological standards. In my experience, this causes very scarce inter-framework communication, and I think that it is instead remarkable that experiments made in a different discipline caused an important discussion in Economics. So, this issue is much larger than the practice around EUT or even Economics as a whole, and hence its proper description and evaluation requires a broader analysis.

On the other hand, we could charge economists of dogmatism. Respecting the texts in the network, however, I do not think this follows. The discussion around the experimental challenges to EUT, at the level of the network, do not show a clearly dogmatic attitude. And the discussion on revealed preferences suggests that many have renounced to reductionism, which is an idea born around the time of the classical axiomatization of EUT. Furthermore, we cannot charge the authors discussed for going against their methodological standards, since they seem to have been largely consistent with them: reactions to the challenges of EUT have been met with either formalized alternatives, in line with the concerns of formalism, or with further experimental inquiry, in line with the empiricists' concerns. I would suggest, then, given the dynamics found in the network, that the dynamics described have been mostly constructive.

But all of this does not mean that there is no relative criticism to make. Consider that even if the network's section of the experimental issues covers four decades after Allais' paper, I found no explicit discussion of this tension. There seems then to be an *unresolved* tension in the form of a lack of a unified concept of systematicity and set of methodological standards. This further suggests that, if it were not for sharing a concern for mathematical rigor and the same domain of social reality, we could even argue that traditionalists and empiricists compose two different social-scientific frameworks. Worthy subjects for future research in this line may be: analyzing in detail if the thought styles characterized are sound, representative, and if there are any others; and if it is a fair characterization, analyze how the traditionalist-empiricist tension has been discussed (if it has been discussed), as well as the current situation surrounding it.

3.3.3 General Evaluation

Certainly, both empiricists and traditionalists consider themselves to be expected utility theorists and social scientists, claiming that at least a part of their scope includes individual decision-making. Both have their reasons to believe they are correct and, considered relatively, it is not clear how should we prefer one over the other (as Hoyningen-Huene argues). However, as discussed in the previous chapter, I think that we can discuss certain points on EUT's historical development from a more general scientific perspective. So, to conclude this chapter, I will further suggest some evaluations of the historical development analyzed in line with the general evaluative part of my scheme.

Two General Values

As discussed in section 2.1, I assumed as non-renounceable two general values that should be respected by any discipline that claims that they are seriously researching an aspect of reality with descriptive aims. By being explicit with this, in my view, reasonable assumption which is compatible with naturalism[57], I think that general, non-relative evaluations of the historical development analyzed in this work are worthy of discussion.

First, let's consider the value of basic logical consistency of scientific framework's systems of ideas. Now, I find it hard to combine this value with carelessness for how ideas are expressed and so, in my view, valuing logical consistency is related to having a healthy amount of logical and mathematical formalization of ideas from which to have a transparent discussion. Together, arguably this set of features pave the way for a constructive accumulation of ideas and constructive criticism, since later authors can easily take previous work and explore its implications, or propose clear alternatives. Consider that it is hard to (logically) evaluate a system of ideas that is expressed with little care for formalization, since many details may be obscured by non explicitly defined terms. Furthermore, being explicit with our definitions and using formal languages to express ideas leave less room for interpretation, arguably making our arguments accountable by having clear implications.

Since the shift around Pareto's time, the formalist regard for logical consistency and transparency notably goes in this line. If anything, this social-scientific thought style deserves praise if we contrast it with others that rely heavily on the use of natural languages (such as English) to express their ideas, and also if we contrast it with most of the pre-Pareto Political Economy in the network. As we have seen, contemporary Economics as a whole has had a constructive accumulation of knowledge since the time of Jevons, and internal criticism has been expressed on clear terms.

Second, let us now consider the value of empirical consistency of a scientific framework's account of reality with the dynamics suggested by empirical observations. Assuming this as a general value goes in line with the idea that *descriptions* is a widely shared dimension of systematicity across scientific disciplines (Hoyningen-Huene, 2013, p. 38). In my opinion, this value implies that any scientific framework with descriptive aims must propose some way to verify that

[57] As discussed in page 2.

their ideas match reality. Otherwise, we certainly have no way of evaluating that the framework's ideas are descriptive of the world. In my view, this is related to the first principle, since formalized and transparent ideas pave an easier way towards obtaining implications that are empirically testable.

Now, while experimental economists and empiricists seem to be an important group in the network, let us bring back here the discussion on the tension between this group and traditionalists. This latter group of authors, such as Gul and Pesendorfer, 2008, argue that inaccurate representations of individuals do not matter as long as economic theories are successful in predicting economic phenomena. Under a first reading, if we accept this approach it does not really make sense to defend the classic representation of preferences (as traditionalist seem to do). As we have seen, observed (and arguably *economic*) individual choices have been shown once and again to be a challenge for EUT, to the point that some suggest that they challenge whole edifice of Economic theory. But even if expected utility theorists abandon the pretension of representing and predicting individual decision-making and focus exclusively on macroeconomic phenomena, prediction at this level is also hardly defensible, as anyone caring to compare economists' predictions with what ends happening even in the short term will attest (see for example Epstein (2015, pp. 2-4)).

But even if this were not the case, I find it hard to defend that the *whole* of Economics should reject the importance of correct descriptions at the individual level and neglect experimental evidence. Of course the purely formal exploration of the implications of a set of axioms has a place in scientific research, but hardly anyone could argue that this directly help us in finding out how to correctly describe social phenomena[58]. And of course that constructive criticism to the way contemporary social-scientific experimentalists gather and analyze empirical evidence is important. But obliviousness to all forms of evidence by claiming that *in principle* it cannot falsify economic theory certainly makes difficult the development of a descriptive project that works better than what we have today. A final point to consider in this line regards the underdetermination of economic

[58] In this line, see for example the criticism of Economics made by Rosenberg (1992)

theory by evidence. If traditionalists simply abandon pretensions of individual description, it is not clear how researchers could decide between EUT and a hypothetical competing theory, if they both have the same predictive power at a macro-economical level but the second has the advantage of describing better individual decision-making.

For all these reasons, as a more general evaluation, I find that the traditionalists' methodological standards characterized here are not defensible in the long run. At most, I think that traditionalists should propose a division between adherents to classical EUT who want to explore its formal implications, and adherents to a project for developing a theory that is both better at describing evidence on individual decision-making and that has, at least, the same amount of predictive success as EUT (which is not one of its virtues in any case). Unfortunately, the limitations of my historical analysis does not allow for an estimation of the popularity and influence of traditionalism in the contemporary Social Sciences. For now, following the second general value mentioned above, I hope that the Social Sciences *learn from* (instead of just ignoring) the late twentieth century's very interesting efforts and debates around EUT and the many experiments surrounding it, and consequently react by trying to produce a superior theory.

Folk Psychology

Lastly, let us now discuss the role of folk psychology and ideas in the historical analysis presented above. Consider first the claim that EUT is a formalization of a part of folk psychology (Guala, 2005, p. 98; Pettit, 1991, §1). Prima facie, without much reflection, some people in our contemporary society may inclined to believe that people's preferences are, for example, asymmetric. On top of this, expected utility theorists additionally commit to further assumptions, such as the possibility of quantifying utility. All of these assumptions are certainly not systematically justified by evidence or solid arguments, and throughout the history I reviewed many authors have relied on an appeal to intuitiveness to justify these choices.

For example, some of these assumptions include the following. The Bernoulli's thesis' existential implication of expected utility being a real entity or phenomenon in individual decision-maker's minds is an assumption held by anyone accepting the classical EUT. The existential implication of preferences existing as a structured entity (in particular, a weak order) with a central role in individual decision-making, which is an assumption of anyone accepting the Pareto-VNM general account of preferences. And, since the discussion on preference reversals in the seventies and eighties, it is clear that many expected utility theorists hold procedure invariance as an assumption related to the basic understanding of preferences behind EUT. But at no point in time empirical arguments are advanced to justify these assumptions, which do seem to come, at least partially, from a folk psychological understanding of individual decision-making.

Now, in subsection 2.2.2 I made an argument on how we should *revise* the role of folk ideas in the development of scientific frameworks, instead of just declaring them as flawed and useless. I think the historical development of EUT in the social sciences is a good example of the importance of my argument, since it is easy from the outside to fall in the eliminativist trap and condemn Economics as a pseudoscience for relying too much in folk psychology, without considering the overall historical context of the discipline. Let us first remember that, as Fleck (1935/1979) illustrates, before the establishment of systematic theories justified by a robust corpus of experimental evidence, scientific and proto-scientific frameworks had nothing else than folk concepts about reality from where to start. Let us then remember, as mentioned in the characterization above, that we can only recognize a more or less unified social-scientific framework since around the time of Pareto, so there has been only a century worth of accumulation of knowledge in a not entirely systematic way, as apparently there is no unified set of methodological standards. It is then understandable that we find more folk ideas in different aspects of EUT's account of individual decision-making than in more mature sciences.

However, EUT's representation of individual decision-making *is not* folk psychology as defined in the previous chapter, since it is *a systematic formalization* of aspects of it. Additionally, it is evident that no one outside the people trained in

Decision Theory uses this formal model to describe individual decision-making, so it can be very hardly understood as ”common sense” or ”folk”. Furthermore, in my view, being a systematic formalization and refinement of folk ideas about the reality of individual decision-making, as some authors have suggested (Guala, 2005, p. 98; Pettit, 1991, §1), is precisely the value of this theory. And as I mentioned above, this systematization brings a transparency that is notable when compared to the work of other social-scientific frameworks, which is what allowed for experimental testing of its implications in the first place. The possibilities that this opens are quite important because, by building on all the accumulated efforts, overall this has the potential of being a constructive project which could approach more and more to a systematic description of individual decision-making (given the experimental evidence). That this account has, at present, a folk-psychological justification and many empirical challenges does not mean that this will continue to be the case in the future.

Furthermore, it is one thing to demand from the outside that Economists should abandon this theory, such as an eliminativist would suggest, but a much more difficult task to provide a working alternative that satisfies everyone’s concept of systematicity plus more general scientific desiderata. This is non-trivial, since social scientists will not simply abandon their working tools without having an alternative that is clearly superior. Demanding from the outside a clearly superior alternative, that can compare to what the more mature sciences posses (such as some well established theories in Physics), ignores the size of this task.

I find if hard to blame contemporary economists for not solving what may be one of the most difficult problems in the whole domain of knowledge (i.e. developing a representation of individual decision-making, with all its nuances, that is descriptively accurate and predictively powerful). At best, the analysis presented in this work suggests that we may criticize non-constructive attitudes that seem to be ignoring that there is much to gain from solving this problem, such as extreme defenses of traditionalism. All things considered, in my view the issue of a ”better alternative” to EUT is bigger than any particular group of social scientists and deserves a systematic study on its own. In this line, and in my view, external a priori judgments from approaches that do not make the effort to understand or contextualize social-scientific practice are, arguably, non-constructive and irrelevant for anyone motivated in finding possible improvements of this practice.

Chapter 4

Discussion and Conclusions

I began this work with an introductory chapter on my general motivations, ambitions, and objectives. I followed, in Chapter 2, with a conceptual scheme to better organize (i) a historical analysis of the development of ideas in a social-scientific framework, and (ii) an evaluation of this development. In Chapter 3, I then presented an example implementation of the descriptive-evaluative scheme, using a restricted set of texts related to the development of Expected Utility Theory's account of individual decision-making.

In this final chapter, I will conclude this work by directly responding to my research question by presenting the lessons learned from this whole process, framed in a series of reflections. These will include, first, reflections surrounding the delimitation of the objects of study and evaluative criteria in Philosophy and History of the Social Sciences. I will then discuss the benefits and limitations of Network Analysis for the analysis of the Social Sciences, as well as some prospects for future research.

4.1 Objects of Study and Evaluative Criteria

In my opinion, Hoyningen-Huene's systematicity is not only useful to think about the practice of scientific disciplines, since we may as well reflect on what we value as the systematicity concept(s) of the Philosophy of Science. And while this subject deserves a dedicated work, the first lesson here regards

1) *the importance of being systematic in defining our objects of study before characterizing and evaluating scientific disciplines*.

The value of this lesson is exemplified in the fact that my descriptive-evaluative scheme is accountable. For example, anyone agreeing or disagreeing with the set of texts included in a *full* implementation of my scheme (rather than the limited example presented in this work), for any given social-scientific framework, may choose to develop further or criticize one or many of the following points: (i) the definition of thought collective (from which we move to an empirical research method such as Social Network Analysis); (ii) the choice of using SNA to operationalize the definition of social-scientific framework given; (iii) the assumptions made to carry out the SNA itself (e.g. the use of citations as a proxy for intellectual interactions) In this line, while my example implementation has many caveats at (iii), following this approach I was able to asses the limitations of my own analysis, in such a way that my conclusions are carefully framed.

For me this contrasts with what some philosophers or historians of science tend to do when very vaguely defining their objects of study as "Economics", and so on. A researcher that agrees or disagrees with the delimitation made to analyze any of the object under question (e.g. Economics), may have to interpret and infer the criteria behind the delimitation (which may be a hard and inaccurate task in itself), before being able to extend what was presented or criticize what was done. Now, I have no evidence that the particular entities studied by the authors I have in mind in this line (Rosenberg, 1992; Hausman, 2018) are not representative of their intended object of study, or that they have not been selected in a systematic way. But the point here is that they are not transparent with their criteria to study certain entities (e.g. texts) and not others to characterize Economics.

For me, the first lesson mentioned so far not only includes the benefit of transparency to define an object of study, but also transparency in the features of this object that will be analyzed. It is in this line that I introduced the concepts of existential implications, methodological standards, and systematicity as features of social-scientific frameworks. This paves the way for an organized way of advancing evaluations, and so I proposed two ways in which this may be done:

relative, in terms of the social-scientific framework's own methodological standards; and general, in terms of (a) defensible scientific values, and (b) the role that folk psychology plays in the social-scientific framework.

As shown in my example implementation, this allows for a nuanced evaluative discussion, which in my view is valuable because these evaluations are more fair and potentially relevant for social-scientific practice. I do not think I would have developed the evaluative aspect of my scheme in this way without my general attitude in Chapter 2, from which I arrive at a second lesson, specific for the Philosophy of the Social Sciences, regarding

2) *the importance of not imposing an evaluative scheme that emits judgments before analyzing actual social-scientific practices and their context*.

As mentioned in the Introduction, I believe that the Philosophy of Science has tools which can be useful for directly or indirectly improving social-scientific practice. An example of this is Dietrich and List's paper (2016), where, using tools from the philosophy of science, attempt to clarify Gul and Pesendorfer's (2008) claims and further challenge them on very carefully considered grounds. Another notable example is Guala's book, where a detailed and nuanced analysis of experimental economics using tools from the Philosophy of Science is presented.

However, some other authors that engage in a broader discussion (Hausman, 2000; Hausman, 2012) are not explicit with their own criteria for evaluation, and have in particular a very unsystematic way of presenting negative evaluations. And, together with other authors (Rosenberg, 1992), they fall in the additional problem of not doing a careful contextualization based on what is important for social scientists, and how their ideas have been developed historically. Thus, a third lesson here is that,

3) *as a whole, my work suggests that a constructive evaluation of the Social Sciences at a general level, using tools from the Philosophy of Science, is possible*.

4.2 Network Analysis and Prospects for Future Research

Now, while I have been overall satisfied with how my descriptive-evaluative scheme works when implemented, most of the difficulties I had during my research stemmed from the particular way I implemented it.

The main difficulty I identified during my example implementation was to make sense of the fact that a citation network may *not* completely overlap with a social-scientific framework as I defined it. The reason is rather obvious, but did not come up to my mind until I was working on the implementation itself. As more history is covered, it becomes more likely that authors of a thought collective intellectually interact with authors of an originally different thought collective, who may carry a different thought style. This is the case of the experimental psychologists in my network who, even if we may argue that they were later ”assimilated”[59], at their initial appearance in the historical description they seemed to participate only from the outside, since their work was ignored for some years in favor of Grether and Plott's. Given the impact of these psychologists' ideas, however, if I had been strict in identifying my object of study with a particular social-scientific framework by filtering these texts out, I would have lost many nuances in my historical description.

I think instead that citation network analysis, in virtue of being oblivious to who belongs to which discipline or scientific framework, allows in fact for an enriched analysis of how different thought collectives and thought styles interact and influence each other. Thus, a fourth lesson to consider is that,
4) *when analyzing the historical development of ideas, we should not filter out content on the a priori grounds that it was not produced by some category of authors (such as institutional disciplines, social-scientific frameworks, and so on).*

Further difficulties I encountered are related to how to fully implement the scheme I suggested using Social Network Analysis. These other difficulties surround the constraints and limits of the example implementation discussed in sec-

[59] Daniel Kahneman, for example, received in 2002 the Nobel Prize in Economics.

tion 3.1, but their potential solutions are the grounds for an interesting reflection on prospects for future research.

A first difficulty in this line surrounds how to systematically include the most recent texts in the citation network. Depending on commentators, as I did for the example implementation, puts us at best a step behind the more current debate, depending on how recent the commentators' texts are. And in general it is not evident how to make sense of the increasing amount of new literature each year and how to systematically define what to include in our analysis. In this line, a prospect for future research offers potentially a way of overcoming or mitigating this difficulty and, in my view, would allow also for solving a difficulty I mentioned in the previous chapter: how to obtain a *complete* network of citations, providing us a much less biased representation of the development of ideas in scientific frameworks.

As suggested in subsection 3.1.1, a rigorous snowballing of enough waves to exhaust the appearance of new nodes may provide us an almost perfectly complete network of texts. It has been shown that under certain conditions, successive waves of snowballing progressively diminish any bias caused by having some nodes and not others in the first wave, to the point that eventually the initial set of nodes does not matter (Heckathorn, 1997). Now, while we may not have an unbiased, systematic up-to-date curated register of all texts being published for the scientific disciplines, what we do have are (very) large and up-to-date databases of citations[60]. While these databases focus mainly on recent literature, a combined approach of using commentator's texts for a first wave of texts plus the use of the citation databases may anyway yield a historically complete citation network.

Following these ideas, I suggest that from the following process we may obtain an almost perfectly complete network of texts.

- First, to compose a first wave of texts that arguably matches a part of the development of the ideas under inquiry, we register the texts mentioned in some set of commentators' texts

[60] Two examples of these databases are Elsevier's Scopus and Web of Science.

- Second, we then follow the citations in successive waves until we exhaust the appearance of new nodes
- Third, we consult the citation databases for the recent literature citing the texts in our partial network, and by controlling for any additional text that we missed, we add them to our network
- Fourth, on the result of the previous step, we start a new process of looking for texts in successive waves until again no new texts appear

Note that here I say an ”almost perfectly complete” network because there is always the possibility of there being texts that influenced some author in the network, but was never cited. That is, a limitation to consider is that some crucial steps in the development of ideas may sometimes be skipped or obscured. An example of this is the description of the previous chapter surrounding the first formulations of the independence axiom as derived from the original von Neumann and Morgenstern's axiomatization of EUT (1944). As noted by Camerer (1995, p. 619), this derivation was done *after* VNM published their text, in two brief symposium papers by Samuelson and Malinvaud (both from 1952). This fact is not evident at all in the network I presented, so my historical description presents a slightly distorted picture of the development of these ideas. Of course, good training in the history of the discipline under inquiry may mitigate this issue by allowing us to identify some of these inaccuracies manually, but the problem may remain for very large networks.

Now, I have not used in my example implementation any *network or node metrics*[61] as is commonly done in Network Analysis. The reason is that the important limitations and biases of my example citation network, in my view, renders them useless: I have no way of estimating how representative my network is, nor of estimating if these values are any close of the values of a hypothetical complete network on the same subject. But regardless of this, if we were to analyze a representative network these metrics could give us objective ways of operationalizing

[61] These metrics are meant to represent different properties of networks or nodes, and their calculations may depend on how the edges connect the nodes, on the position of the nodes in the overall structure, or on other characteristics.

notions such as the *popularity* and the *impact* of texts[62]. This leads me to the fifth lesson, regarding

5) *the relevance of Social Network Analysis as a powerful tool to complement how the Philosophy of Social Sciences produces historical descriptions, providing us with a systematic way of approaching our objects of study*[63].

Then, coming back to the idea of obtaining almost perfectly complete networks, let us consider now that manually analyzing each of those text's mentions, as I have done in my example implementation, would require a large amount of time rendering my approach practically impossible. In this line, a further and exiting prospect for future research surrounds the use of the power of computer programming techniques and computer processing (both of which are growing very quickly), which allow for fast processing and analysis of data. In particular, it is relevant to mention some techniques of the so-called *Machine Learning* (ML).

In short, ML is an umbrella term for programming techniques where an algorithm goes through a phase of "training" by being calibrated with previously curated databases (i.e. by researchers), with the aim of obtaining, in virtue of different techniques[64], a modified algorithm that is precise in categorizing data points in non previously curated databases. An example of this is the application of ML for *sentiment analysis*, aimed to obtain an automatic classification (thanks to calibrated algorithm) of pieces of text with categories that indicate the general "sentiment" or attitude behind[65].

As a proof of concept, ML sentiment analysis is particularly relevant because it suggests that we could automatically identify the *kind of mentions* (e.g. agree-

[62] For example, a form of popularity may be measured with the simple count of mentions, while different forms of impact may be measured using metrics such as the *betweeness* or *eigenvector* centralities of texts.

[63] This conclusion is similar to one of the advantages of the use SNA for analyzing the history of Economics defended by Herfeld and Malte (2018).

[64] Classical Machine Learning relies, for example, on statistical models, while a more recent technique called Neural Networks uses a different approach that intends to mimic how a biological nervous system learns to process information passed by the senses.

[65] For example, sentiment analysis this has been applied to *tweets* (short messages) of the Twitter social-media site to measure "positive", "negative", or "neutral" attitudes towards different topics such as political parties or consumer goods. Twitter sentiment analysis has become so popular that a quick internet search of the term yields numerous results. See, for example, the following article: https://arxiv.org/pdf/1711.10377.pdf

ment, disagreement, neutral) between texts in large networks of (unclassified) citations. This idea remains, however, speculative since to my knowledge there has been no attempt of implementing such a thing. There may be many reasons for this, related to the complexity of the task, the conjunction of skills that need to be put together, and the time it would demand to obtain tangible results. The technological advances of our time are rapidly changing how the work in many areas of our society is done, and I do not think that Philosophy of Science should be the exception. If nothing else, a largely representative, always up-to-date, and rapidly produced network of classified citations sounds like a very valuable tool from which to start a systematic inquiry into any social-scientific topic.

In any case, these final remarks justify an additional lesson, with which I close this thesis.

6) *A further advantage of using Network Analysis as a method is that it provides us, as philosophers and historians of science, of a linguistic framework with which to communicate clearly with the wider scientific community, opening interesting possibilities for interdisciplinary collaboration.*

References

Bricker, P. (2016). Ontological Commitment. *The Stanford Encyclopedia of Philosophy, Winter 2016 Edition*. Retrieved from https://plato.stanford.edu/archives/win2016/entries/ontological-commitment/

Carnap, R. (1950). Empiricism, Semantics, and Ontology. *Revue Internationale de Philosophie*, *4*(11), 20–40.

Churchland, P. M. (1988). *Matter and Consciousness (Revised Edition)*. Cambridge (USA): The MIT Press. (Original work published 1984)

Churchland, P. M. (1989). Folk Psychology and the Explanation of Human Behavior. *Philosophical Perspectives*, *3*, 225–241.

Della Porta, D. & Keating, M. (2008). How Many Approaches in the Social Sciences? An Epistemological Introduction. In D. Della Porta & M. Keating (Eds.), *Approaches and Methodologies in the Social Sciences: A Pluralist Perspective* (pp. 19–39). Cambridge (UK): Cambridge University Press.

Dogan, M. (1996). The Hybridization of Social Science Knowledge. *Library Trends*, *45*(2), 296–314.

Doppelt, G. (1990). The Naturalist Conception of Methodological Standards In Science: A Critique. *Philosophy of Science*, *57*(1), 1–19.

Epstein, B. (2015). *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. Oxford (UK): Oxford University Press.

Fleck, L. (1979). *Genesis and Development of a Scientific Fact. [Entstehung Und Entwicklung Einer Wissenschaftlichen Tatsache]*. Chicago (USA): The University of Chicago Press. (Original work published 1935)

Guala, F. (2005). *The Methodology of Experimental Economics*. New York (USA): Cambridge University Press.

Guala, F. (2016a). Philosophy of the Social Sciences: Naturalism and Anti-Naturalism in the Philosophy of Social Science. In P. Humphreys (Ed.), *The Oxford Handbook of Philosophy of Science*. Oxford (UK): Oxford University Press.

Guala, F. (2016b). *Understanding Institutions: The Science and Philosophy of Living Together*. Princeton (USA): Princeton University Press.

Hallpike, C. R. (2011). *On Primitive Society and Other Forbidden Topics*. Bloomington (USA): AuthorHouse.

Hansson, S. O. (2017). Science and Pseudo-Science. *The Stanford Encyclopedia of Philosophy*, *Summer 2017 Edition*. Retrieved from https://plato.stanford.edu/archives/sum2017/entries/pseudo-science/

Hausman, D. M. (2018). Philosophy of Economics. *The Stanford Encyclopedia of Philosophy*, *Fall 2018 Edition*. Retrieved from https://plato.stanford.edu/archives/fall2018/entries/economics/

Heckathorn, D. D. (1997). Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems*, *44*(2), 174–199.

Heckathorn, D. D. (2011). Comment: Snowball versus Respondent-Driven Sampling: *Sociological Methodology*.

Herfeld, C. & Malte, D. (2018). Five Reasons for the Use of Network Analysis in the History of Economics. *Journal of Economic Methodology*, *25*(4), 311–328.

Hindriks, F. & Guala, F. (2015). Understanding Institutions: Replies to Aoki, Binmore, Hodgson, Searle, Smith, and Sugden. *Journal of Institutional Economics*, 1–8. Retrieved from http://journals.cambridge.org/abstract_S1744137415000120

Hirstein, W. (2009). Confabulation. In B. T., C. A., & P. Wilken (Eds.), *The Oxford Companion to Consciousness*. Oxford (UK): Oxford University Press.

Hoyningen-Huene, P. (2013). *Systematicity: The Nature of Science*. Oxford (UK): Oxford University Press.

Kahneman, D. & Tversky, A. (1996). On the Reality of Cognitive Illusions. *Psychological Review*, *103*(3), 582–591.

Kincaid, H. (2012). Introduction: Doing Philosophy of the Social Sciences. In H. Kincaid (Ed.), *The Oxford Handbook of Philosophy of Social Science* (pp. 3–20). Oxford (UK): Oxford University Press.

Osborne, R. C. (2016). Debunking Rationalist Defenses of Common-Sense Ontology: An Empirical Approach. *Review of Philosophy and Psychology*, *7*(1), 197–221.

Pettit, P. (1991). Decision Theory and Folk Psychology. In M. Bacharach & S. Hurley (Eds.), *Foundations of Decision Theory: Issues and Advances*. Oxford (UK): Blackwell.

Ramsey, W. (2019). Eliminative Materialism. *The Stanford Encyclopedia of Philosophy, Spring 2019 Edition*. Retrieved from https://plato.stanford.edu/archives/spr2019/entries/materialism-eliminative/

Ravenscroft, I. (2019). Folk Psychology as a Theory. *The Stanford Encyclopedia of Philosophy, Summer 2019 Edition*. Retrieved from https://plato.stanford.edu/archives/sum2019/entries/folkpsych-theory/

Rosenberg, A. (1992). *Economics — Mathematical Politics or Science of Diminishing Returns?* Chicago (USA): University of Chicago Press.

Thoma, J. (Forthcoming). Folk Psychology and the Interpretation of Decision Theory. *Ergo*.

Thoma, J. (2019). Decision Theory. *The Open Handbook of Formal Epistemology*. Retrieved from https://jonathanweisberg.org/pdf/open-handbook-of-formal-epistemology.pdf

Wellman, B. (1988). Structural Analysis: From Method and Metaphor to Theory and Substance. *Contemporary Studies in Sociology*, *15*.

Citation Network Texts

Afriat, S. N. (1967). The Construction of Utility Functions from Expenditure Data. *International Economic Review*, *8*(1), 67–77.

Allais, M. (1979). The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School. In M. Allais & O. Hagen (Eds.), *Expected Utility Hypotheses and the Allais Paradox: Contemporary Discussions of the Decisions under Uncertainty with Allais' Re-*

joinder (pp. 27–145). Theory and Decision Library. Dordrecht (Netherlands): Springer. (Original work published 1953)

Bentham, J. (2007). *An Introduction to the Principles of Morals and Legislation*. New York (USA): Dover Publications. (Original work published 1789)

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, *22*(1), 23–36. (Original work published 1738)

Bolker, E. D. (1966). Functions Resembling Quotients of Measures. *Transactions of the American Mathematical Society*, *124*(2), 292–312.

Camerer, C. (1995). Individual Decision Making. In J. H. Kagel & A. E. Roth (Eds.), *Handbook of Experimental Economics* (pp. 587–703). Princeton (USA): Princeton University Press.

Chew, S. H. & MacCrimmon, K. (1979). Alpha-Nu Choice Theory: A Generalization of Expected Utility Theory. *Working Paper of the Faculty of Commerce and Business Administration, Number 686*.

Cox, J. C. & Epstein, S. (1989). Preference Reversals Without the Independence Axiom. *The American Economic Review*, *79*(3), 408–426.

de Fermat, P. & Pascal, B. (1929). Fermat and Pascal on Probability. In *A Source Book in Mathematics*. New York (USA): McGraw-Hill Book Co. (Original work published 1654)

Dietrich, F. & List, C. (2016). Mentalism versus Behaviourism in Economics: A Philosophy-of-Science Perspective. *Economics & Philosophy*, *32*(2), 249–281.

Grether, D. M. & Plott, C. R. (1979). Economic Theory of Choice and the Preference Reversal Phenomenon. *The American Economic Review*, *69*(4), 623–638.

Gul, F. & Pesendorfer, W. (2008). The Case for Mindless Economics. Oxford (UK): Oxford University Press.

Hausman, D. M. (2000). Revealed Preference, Belief, and Game Theory. *Economics & Philosophy*, *16*(1), 99–115.

Hausman, D. M. (2012). *Preference, Value, Choice, and Welfare*. Cambridge (UK): Cambridge University Press.

Herstein, I. N. & Milnor, J. (1953). An Axiomatic Approach to Measurable Utility. *Econometrica*, *21*(2), 291–297.

Hicks, J. R. & Allen, R. G. D. (1934). A Reconsideration of the Theory of Value. *Economica*, *1*(1), 52-76 and 196-219.

Holt, C. A. (1986). Preference Reversals and the Independence Axiom. *The American Economic Review*, *76*(3), 508–515.

Houthakker, H. S. (1950). Revealed Preference and the Utility Function. *Economica*, *17*(66), 159–174.

Hume, D. (1994). Of Money. In K. Haakonssen (Ed.), *Hume: Political Essays* (pp. 115–125). Cambridge Texts in the History of Political Thought. Cambridge (UK): Cambridge University Press. (Original work published 1752)

Jeffrey, R. C. (1983). *The Logic of Decision* (2nd ed.). Chicago (USA): University of Chicago Press. (Original work published 1965)

Jevons, W. S. (1871). *The Theory of Political Economy* (1st ed.). London (UK): MacMillan and Co.

Joyce, J. M. (1999). *The Foundations of Causal Decision Theory*. Cambridge Studies in Probability, Induction and Decision Theory. Cambridge (UK): Cambridge University Press.

Karni, E. & Safra, Z. (1987). 'Preference Reversal' and the Observability of Preferences by Experimental Methods. *Econometrica*, *55*(3), 675–685.

Keller, L. R., Segal, U., & Wang, T. (1993). The Becker-DeGroot-Marschak Mechanism and Generalized Utility Theories: Theoretical Predictions and Empirical Observations. *Theory and Decision*, *34*(2), 83–97.

Lichtenstein, S. & Slovic, P. (2006a). Response-Induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas. In P. Slovic & S. Lichtenstein (Eds.), *The Construction of Preference* (pp. 69–76). Cambridge (UK): Cambridge University Press. (Original work published 1973)

Lichtenstein, S. & Slovic, P. (2006b). Reversals of Preference Between Bids and Choices in Gambling Decisions. In P. Slovic & S. Lichtenstein (Eds.), *The Construction of Preference* (pp. 52–68). Cambridge (UK): Cambridge University Press. (Original work published 1971)

Little, I. (1949). A Reformulation of the Theory of Consumer's Behaviour. *Oxford Economic Papers*, *1*(1), 90–99.

Little, I. (1957). *A Critique of Welfare Economics* (2nd ed.). Oxford (UK): Oxford University Press.

Machina, M. J. (1982). 'Expected Utility' Analysis without the Independence Axiom. *Econometrica*, *50*(2), 277–323.

Marschak, J. (1950). Rational Behavior, Uncertain Prospects, and Measurable Utility. *Econometrica*, *18*(2), 111–141.

Mill, J. S. (1998). *Utilitarianism* (R. Crisp, Ed.). Oxford (UK): Oxford University Press. (Original work published 1861)

Mill, J. S. (2011). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation*. Cambridge Library Collection - Philosophy. Cambridge (UK): Cambridge University Press. (Original work published 1843)

Pareto, V. (1971). *Manual of Political Economy* (A. Schwier, Trans.). New York (USA): A. M. Kelley. (Original work published 1909)

Quiggin, J. (1982). A Theory of Anticipated Utility. *Journal of Economic Behavior & Organization*, *3*(4), 323–343.

Ramsey, F. P. (1988). Truth and Probability. In N.-E. Sahlin & P. Gärdenfors (Eds.), *Decision, Probability and Utility: Selected Readings* (pp. 19–47). Cambridge (UK): Cambridge University Press. (Original work published 1926)

Robbins, L. (1995). An Essay on the Nature and Significance of Economic Science. In *The Foundations of Econometric Analysis* (pp. 98–102). Cambridge (UK): Cambridge University Press. (Original work published 1932)

Safra, Z., Segal, U., & Spivak, A. (1990a). Preference Reversal and Nonexpected Utility Behavior. *The American Economic Review*, *80*(4), 922–930.

Safra, Z., Segal, U., & Spivak, A. (1990b). The Becker-DeGroot-Marschak Mechanism and Nonexpected Utility: A Testable Approach. *Journal of Risk and Uncertainty*, *3*(2), 177–190.

Samuelson, P. A. (1938). A Note on the Pure Theory of Consumer's Behaviour. *Economica*, *5*(17), 61–71.

Samuelson, P. A. (1947). *Foundations of Economic Analysis*. Cambridge (USA): Harvard University Press.

Savage, L. J. (1954). *The Foundations of Statistics*. New York (USA): John Wiley & Sons.

Segal, U. (1988). Does the Preference Reversal Phenomenon Necessarily Contradict the Independence Axiom? *The American Economic Review*, *78*(1), 233–236.

Sen, A. K. (1971). Choice Functions and Revealed Preference. *The Review of Economic Studies*, *38*(3), 307–317.

Sen, A. K. (1973). Behaviour and the Concept of Preference. *Economica*, *40*(159), 241–259.

Sen, A. K. (1993). Internal Consistency of Choice. *Econometrica*, *61*(3), 495–521.

Slovic, P. (1995). The Construction of Preference. *American Psychologist*, *50*(5), 364–371.

Slovic, P. & Lichtenstein, S. (1968). Relative Importance of Probabilities and Payoffs in Risk Taking. *Journal of Experimental Psychology*, 1–18.

Smith, A. (1937). *An Inquiry into the Nature and Causes of the Wealth of Nations*. New York (USA): Random House. (Original work published 1776)

Starmer, C. & Sugden, R. (1991). Does the Random-Lottery Incentive System Elicit True Preferences? An Experimental Investigation. *The American Economic Review*, *81*(4), 971–978.

Tversky, A., Slovic, P., & Kahneman, D. (1990). The Causes of Preference Reversal. *The American Economic Review*, *80*(1), 204–217.

von Neumann, J. & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton (USA): Princeton University Press.

Vredenburgh, K. (2019). A Unificationist Defence of Revealed Preferences. *Economics & Philosophy*, *36*(1), 149–169.

Yaari, M. E. (1987). The Dual Theory of Choice under Risk. *Econometrica*, *55*(1), 95–115.

Appendix A

Citation Network

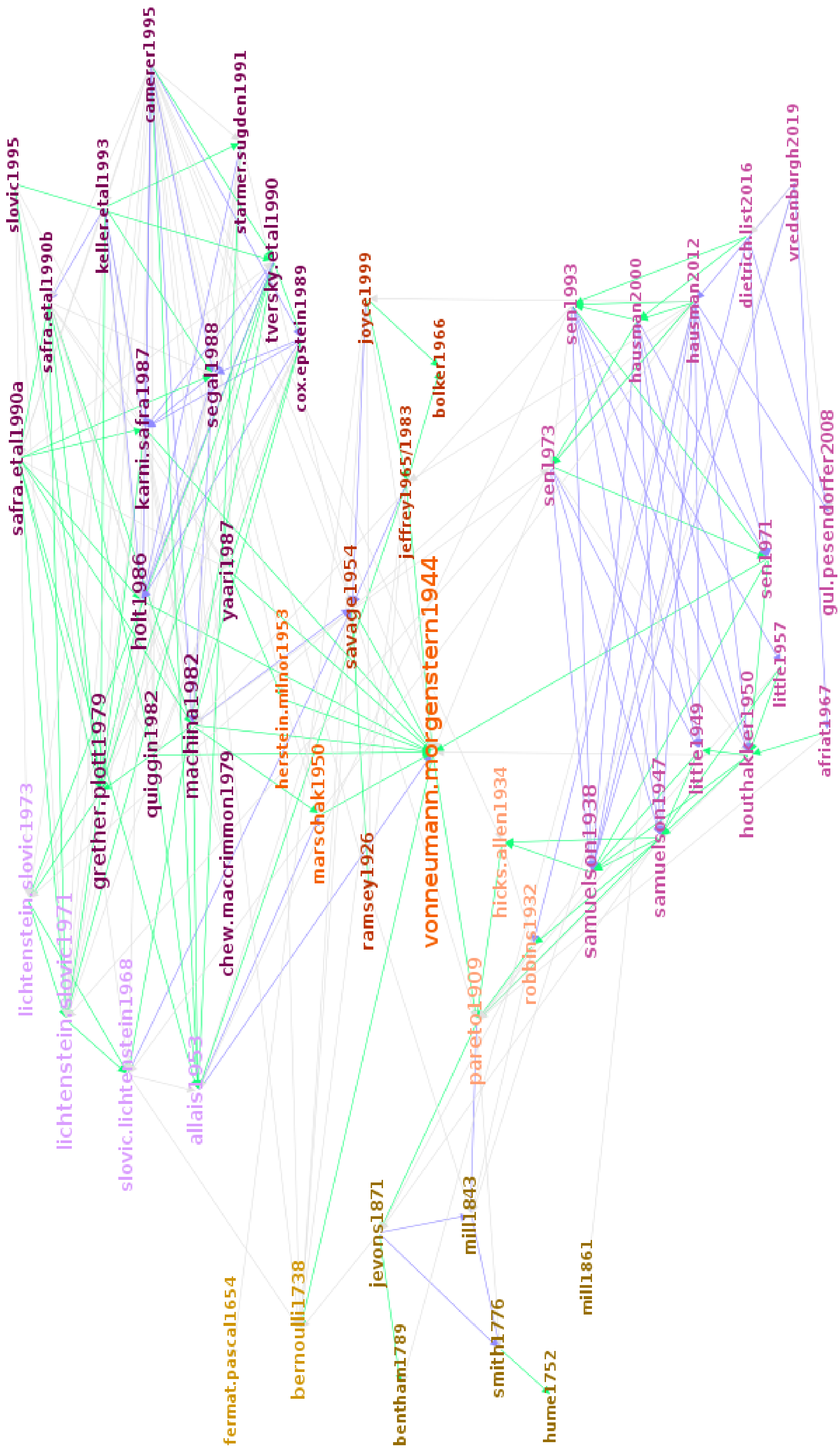
The following page contains the citation network constructed and used as described in Chapter 3.

The nodes represent specific texts registered in the Citation Network Texts section of the References, with the convention ”authorsORIGINALYEAR”. For example, Pareto (1909/1971) is represented as ”pareto1909”, and Safra et al. (1990a) as ”safra.etal1990a”. Each node is colored depending on the subsection of the historical description (section 3.2) they appear in, and their sizes are proportional to how many times they were mentioned by others in the network.

The edges represent citations between texts and their colors represent mention types, such that agreement are in *green*, disagreements are in *light blue*, and neutral mentions are in *light gray*.

I have made publicly available (i) the list of nodes, (ii) the list of edges, which includes a ”reason” column explaining potentially controversial categorizations of mention types, and (iii) the digital image of the network as presented below[66], at Zenodo (https://doi.org/10.5281/zenodo.4005963).

[66] This visualization of the network was obtained using the software Gephi.



slovic1995
safra.etal1990b
keller.etal1993
camerer1995
starmer.sugden1991
tversky.etal1990
cox.epstein1989
segal1988
karni.safra1987
safra.etal1990a
holt1986
yaari1987
machina1982
quiggin1982
grether.plott1979
chew.maccrimmon1979
lichtenstein.slovic1973
lichtenstein.slovic1971
slovic.lichtenstein1968
allais1953
joyce1999
bolker1966
jeffrey1965/1983
savage1954
herstein.milnor1953
marschak1950
ramsey1926
vonneumann.morgenstern1944
hicks.allen1934
robbins1932
pareto1909
sen1993
hausman2000
hausman2012
dietrich.list2016
vredenburgh2019
sen1973
gul.pesendorfer2008
sen1971
little1957
houthakker1950
afriat1967
little1949
samuelson1947
samuelson1938
mill1843
jevons1871
mill1861
fermat.pascal1654
bernoulli1738
bentham1789
smith1776
hume1752