Minding the gap between micro and macro: reductive and non-reductive strategies applied to economics

MULLER-FONSECA, ALEXANDRE

Minding the gap between micro and macro: reductive and non-reductive strategies applied to economics

Alexandre Müller Fonseca

A thesis presented for the degree of
Doctor in Philosophy

Faculty of Arts and Humanities
Department of Philosophy
University of Durham
England

December 2022

Minding the gap between micro and macro: reductive and non-reductive strategies applied to economics

Alexandre Müller Fonseca

Submitted for the degree of Doctor of Philosophy
December 2022

Abstract

This dissertation discusses the application of two reductive and one non-reductive approaches to relate macro-to-microeconomics properties. Reduction is treated as an explanatory strategy where the inquirer transfers the explanans of her target of interest (explanandum) to the next level down. In the context of economics, this condition is not sufficient to satisfy a reduction of a macroeconomic phenomenon/principle to microeconomic principles. It is also required that the explanans (the explainee doers) admits the properties of individuals (beliefs, expectations) only. The other side of this view, the non-reductive relation of macro-to-microeconomics, is positively attained when the inquirer explains how a given macroeconomic phenomenon was eventually possible by appealing to other features that do not stand on individuals' properties. Contrary to the reductionist approach, the non-reductionist contender is the case when other elements included in the microeconomic level share the same causal efficacy to determine the manifestation of the macroeconomic phenomenon. In this dissertation, I argue that both approaches, the reductive and non-reductive frameworks, are feasible according to the context. By assessing a classical economic model as a case study, the Lucas's Phillips curve model, I demonstrate how the selected reductive paradigms, the Nagel's reduction and constitutive mechanism, fit the model, illustrating the Lucasian approach as a positive example of reduction of a macroeconomic principle/phenomenon to a set of microeconomic principles. Conversely, I also detail and identify an example of non-reductive explanation where outer elements inside the microeconomic level are as much efficacious as the individuals' properties for determining the chosen macroeconomic set of properties. I demonstrate that, contrary to Kevin Hoover's belief, the dispute between reduction and supervenience (non-reductive approach), the alternative concept employed to describe how micro and macro sets are related, is not an all-or-nothing matter: we may have both scenarios depending on the context and explanatory purposes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION TO THE NOTION OF REDUCTION

Before outlining the aims and contents of this thesis, I shall in this first section introduce the core notion on which the thesis lies: reduction. In its digital version, *Oxford English Dictionary* defines "reduction" as a noun that denotes "the action of fact making something smaller or less in amount, degree, or size". Metaphorically, Lou Reed's *Voices of Freedom* captures the spirit of this definition: "*The river runs right through those mountains*, *and reduces that rock to stone*, *the ocean reduces that stone to sand*". These senses of "reduction" are not how philosophers typically employ the term. However, the latter part of the Reed's quote offers an example which may be amenable to the philosophical sense of reduction.

Metaphysically, we may construe, in a charitable way, the snipped short piece of Lou Reed's track as identifying a part-whole relation: a rock is made of billions of grains of sand. As an external force, the ocean can decompose the whole (rock) into smaller items (grains of sand), making visible that what is perceived as a unitary object is a sum of grains of sand. We can look at concrete objects such as rocks and bottles or more abstract entities like societies, electrons, and GDP and raise questions about their constitution.

Thomas Hobbes (1968 [1651]), for instance, envisioned society as a compositional (abstract) entity. In their state of nature, individuals live in a permanent sense of vigilance, seeking to preserve themselves against constant competition for scarce resources. As self-interested individuals, at some point in time, people introduced some rules, i.e., forms of interaction that facilitate their coexistence in the same environment. Our complex societies, Hobbes believed, came from a general agreement among individuals – society is a great whole made up of individuals.

In a similar vein, one of the pioneers of the marginalist revolution in economics, Carl Menger, claimed that every observable economic phenomenon is determined by people's actions.[1] Menger, the founding father of Austrian economics, argued that to understand economic phenomena, we must investigate the very elements that constitute them – individuals

[1] Between the end of the 19th century and the opening of the 20th century, through the writings of authors like Carl Menger (1963 [1883]), Stanley Jevons (1911), and Leon Walras (1954 [1926]), the marginalist revolution was born. Together, these three scholars argued in favour of a subjective theory of value: the value of goods is primarily determined by their extrinsic features, i.e., they depend on how much individuals value their utilities according to their needs. At that time, this idea collided frontally with the labour theory of value. According to this paradigm, claimed by Adam Smith (1976 [1776]), David Ricardo (1911 [1811]), and (with a different lens and consequences) Karl Marx (1976 [1867]), things got their value due to the embodied work disposed during the process of creation: more labour, more cost of production, implies more value.

– and seek to uncover "the laws by which the former are built up from the latter" (Menger, 1963 [1883], p. 93).

The views expressed by Hobbes and Menger suggest that we, as humans, have an inclination to systematise the external world into levels of organisation, from huge complex organisations, such as societal groups, to smaller components that make up those greater entities on top: "levels of organization are a deep, non-arbitrary, and extremely important feature of the ontological architecture of the world" (Wimsatt, 2007, p. 203).

In 1958, Paul Oppenheim and Hilary Putnam took this ontological principle for granted and proposed a potential guideline to integrate the various levels of reality. Reality, Oppenheim and Putnam argued, is divided into six great layers: social groups (level 6), multicellular living things (level 5), cells (level 4), molecules (level 3), atoms (level 2), and elementary particles (level 1). These levels are compositionally related: entities of level 5 are parts/components of level 6; entities of level 4 are parts/components of level 5, respectively – this process should terminate at the most fundamental level (Oppenheim and Putnam, 1958, p. 9-10).

Oppenheim and Putnam proposed a picture that those who pursue the unity of science as a plausible working hypothesis may embrace. Tuomas Takho (2021, p. 1) interprets this view of the unity of science as an *ontological ideal*: to replace the descriptions of, for example, a typical tiger with the descriptions of its correspondent components located at lower levels of organisation. In such an account, the unity of science is the *end* which necessarily requires a reductive strategy across various levels as the *means*. As long as levels are compositionally related, inter-level reductions are attained if, for example, reductionists can replace the descriptions of an entity from level 5 (a tiger, a dog) with the descriptions of its components placed at the next level down. When scientists can characterise *x* as nothing more than a sum of its parts, they can say that this whole, this unitary kind, is nothing over and above its structural items.

Oppenheim and Putnam had an ontological eye on reduction: showing that a whole is a sort of mereological entity that is nothing more than its parts, which permits us to decrease the ontology of the universe – where we previously counted two types of things, we discover there was only one (Schaffner, 1967, p. 143). If scientists could systematically decompose higher-level phenomena by redescribing them in terms of their structural features, then we may achieve unification across the various disciplines. But the ontological approach towards reduction wasn't the only approach towards reduction.

Interested in offering a formal recipe for reduction – what "reduction" is and how it could be obtained, Ernest Nagel developed his own approach to reduction. Like Oppenheim and

Putnam, Nagel presupposed the idea that reality is somewhat organised in layers. Nonetheless, he associates reduction with the very explanations couched by scientific disciplines. In this regard, a reduction isn't a direct relation between kinds, properties, and other real-world items that sciences typically describe.

From his first publication on reduction (published in 1935) until his final writings about the topic, Nagel always treated reduction as a special type of *explanation*. For Nagel, scientific reductions are inter-level explanations whereby the regularities of a discipline (e.g., biology) are fully explained by the regularities of a lower-level discipline (for example, chemistry). Nagel took for granted the ontological principle that reality is somewhat organised in layers. In this fashion, each scientific discipline concentrates its theoretical efforts around a specific group of kinds (natural or social) and laws/regularities to forge their explanations. In the spirit of logical positivism, Nagel stressed that the relata of reductions would mainly cover representational items. It's not the very regularities of things that are reducible but instead the statements about those phenomena (Nagel, 1970, p. 119).

Inspired by Nagel's approach, in this thesis, I also treat reduction as an explanatory strategy that can, in different scenarios, be pursued to achieve different goals. Potential reasons why scholars might be interested in reductions are:

> *Ontological parsimony*: To show that where there appears to be two types of entity there is in fact only one, more fundamental type.
> *Decomposition*: Given a feature possessed by a certain set of systems, to show that the systems' parts and their interactions, described at a specified level of detail, are sufficient to explain the feature.
> *Unification*: To demonstrate that a wide array of distinct generalizations and observations can be explained by a small number of more fundamental generalizations.
> *Correction*: To identify and explain exceptions to less fundamental generalizations using generalizations and details drawn from a more fundamental realm. (Steel, 2004, p. 59)

This list doesn't intend to be exhaustive. Its purpose is to show the variety of possible objectives that are involved in reductive research – why reduction may sound like an interesting and important program for the purposes of understanding in some scientific contexts. The question about the *form* of reduction is subordinated to its purposes, what is considered as the acceptable sufficient condition(s) for reduction. To give an example, consider Jaegwon Kim's (2005) functional model of reduction.

Very briefly, Kim argues that reduction should make intelligible how a given phenomenon derives from a more basic one. The answer to a question like "Why does this phenomenon occur?" should be conceived in functional terms. By "functional", he means that the target phenomenon should be explained in terms of its causal role, which means we have to seek a detailed description of its role inside a mechanism (or a system) in which the phenomenon itself interacts with other elements.

So, the above question might be rewritten in the following way: "In virtue of the realization of what functional roles does the phenomenon occur?". Herein, "realization" is a relation involving two orders of properties, one which is the realizer – a first-order property that does the realizing – and the other being the realized property (a second-order property defined by its causal role), whereby both are instantiated in the same thing. The realized property is just the second-order property of having some property or others that play a specific causal role, and the realizers are the first-order properties that play that role. For example, we might consider the property of pain to be a functional property that disposes an individual to behave in certain ways, which is in turn realized by certain first-order properties of the individual's brain. According to this functional analysis, the form of reductive explanation is as follows:

> Step 1 [Functionalization of the target theory].
> Property *M* to be reduced is given a *functional definition* of the following form: Having *M* = (def.) having some property or other *P* (in the reduction base domain) such that *P* performs causal task *C*. [...].
> Step 2 [Identification of the realizers of *M*].
> Find the properties (or mechanisms) in the reduction base that perform the causal task *C*.
> Step 3 [Developing an Explanatory Theory].
> Construct a theory that explains how the realizers of *M* perform task *C*. (Kim 2005, p. 101-102, original emphasis).

It's not necessary to address Kim's theory in detail. I merely aim to illustrate that, in Kim's view, the goal of reduction is to get ontological parsimony, and the best way to get it is via a functionalization at the level of properties. He believes that to simplify the world's ontology, we should functionalise the derivative target properties and uncover their primitive properties that perform their causal task.

As I said before, ontological parsimony is one of the popular goals of reduction. Many critics of reduction across various sciences raise queries about the practicality of that goal. In biology, for instance, authors like the French physiologist Claude Bernard claim that organisms are more than their physiology or physical constitution: "Physiologists and physicians must therefore always consider organisms as a whole and in detail at one and the same time" (Bernard, 1961 [1865], p. 91). In a similar line of reasoning, the biologist Ludwig von Bertalanffy writes:

> Since the fundamental character of the living thing is its organization, the customary investigation of the single parts and processes, even the most thorough physico-chemical analysis, cannot provide a complete explanation of the vital phenomena. This investigation gives us no information about the co-ordination of the parts and processes in the complicated system of the living whole which constitutes the essential 'nature' of the organism. (von Bertalanffy, 1933, p. 64)

In his words, Bertalanffy suggests there's something that biologists may lose in terms of understanding if they insist on looking at organisms exclusively in terms of their constitutions. Importantly for the purposes of this thesis, similar reservations appear in economics concerning a type of reduction that tries to explain all the behaviour of an economic system or phenomenon in terms of its parts.

At the beginning of the 20th century, Lionel Robbins's definition of economics gained tremendous popularity among the practitioners of economic science. In Robbins's words (1935, p. 16), "Economics is the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses". In such an account, economics is inevitably behavioural-driven – it primarily seeks to understand how people make economic decisions under a set of available constraints (e.g., budget constraints, environmental constraints, and the like). As a result, economics has become the science of choice. In other words, economics, in modern terminology, *is* microeconomics (Hoover, 2001b, p. 70).

Historically, since John Maynard Keynes's *General theory* (1936) onwards, economic science has been divided into two major branches: microeconomics and macroeconomics. Typically, the difference between the two is a matter of scale, a contrast between small and large (Janssen, 1993, chapter 1). Microeconomics is generally defined as the branch of the economic discipline that studies "how households and firms make decisions and interact in specific markets," whereas "macroeconomics is the study of economic-wide phenomena" (Mankiw and Taylor, 2017, p. 8) – e.g., inflation, rate of economic growth, changes in unemployment and the like. Based on these definitions, it is often said that the very nature of economics rests upon the microeconomic domain: it seems that every macroeconomic phenomenon demands a microeconomic explanation. In other words, any macroeconomic phenomena require a reductive explanation (Hoover, ibid. p. 70), i.e., an explanation that is grounded in individual decision-making.

In economics, this grounding relation between individuals' conditions and subsequent macroeconomic results is encountered in many mainstream economic models and methods where economists try to provide microfoundations for macroeconomic events: the idea that macro phenomena are directly and exhaustively explained by individuals under a set of constraints – what is also named as “explanatory individualism”.

There's a reason why economists are interested in reduction: *intentionality*. It's safer to conceive economic forecasting models based on individual decision-making instead of trust in large-scale regularities. As Hoover (2009, p. 389) saliently puts this issue, "[…] the economist finds it hard to see any analysis as 'economic' unless it deals in intentional states. The point of

microfoundations is to recapture the teleology that appeared to be missing in macroeconomics". But how could this be possible?

In a series of writings, Kevin Hoover (2001a, 2001b, 2009, 2010, 2015) contends that microeconomics is not exclusively made of people's mental states (their intentions, expectations, and the like) and relative prices. People use the money issued by the government on a daily basis and trust macroeconomic data to alleviate potential misallocations. So, macroeconomic aggregates have an influence on individual decision-making.

Moreover, some macroeconomic entities (e.g., Gross domestic product (GDP) and general price level), according to Hoover, are impossible to decompose into their microeconomic components – they are dimensionally distinct from the components (Hoover, 2001a, p. 113). There's no reason to assume, Hoover contends, that only the properties of individuals are causally efficacious. According to Hoover, economists should once and for all discard the microfoundational program, accepting an alternative non-reductive paradigm between micro-and-macroeconomics. The difficulty of getting a complete explanatory individualism of macroeconomics is what makes reduction in this context impossible.

In this thesis, I challenge this idea. Hoover asserts that reduction is impossible due to some ontological barriers: the fact that we can't eliminate macroeconomic entities, and that macro entities have causal influence over the micro, makes reduction in this context an unattractive program. In the thesis I shall argue that the goal of the microfoundational thesis, the explanatory thesis prescribing that economists should develop their explanations based on individual decision-making, should be assessed on a case-by-case basis.

By employing two philosophical frameworks for reduction, Nagel's view and the constitutive-mechanistic account, I demonstrate how they apply to Lucas's Phillips curve. Scholars like Hoover see the goal of reduction as requiring ontological parsimony. In economics, the problem is obvious: If we can't decompose every macroeconomic entity and explain every macroeconomic event solely in terms of individuals' properties (beliefs, expectations, and actions) under a set of constraints (e.g., relative prices), then there's no chance for reduction. One aspect of this challenge is that the mapping between macroeconomic entities and the agents in the economy is not one to one (Hoover, 2001b, ibid, p 73).

Demonstrating an identity relation between the micro and macro entities involved in reductions is the usual requirement claimed by many reductionists, and the difficulty in many scientific contexts of articulating such a relation in a transparent way raises scepticism among

scholars who, like Hoover, are unconvinced by reductive programmes.[2] The weight of such an objection depends on, as I already indicated, the purposes of reduction. As an explanatory strategy, the goal of reduction in economics is to obtain explanatory parsimony – to demonstrate that macroeconomic outcomes result only from individuals' properties.

The condition that reduction must establish identity between micro and macro entities concerns *one* of the potential goals of reduction – ontological parsimony. But there are other goals for reductions across various sciences and, in economics, its primary aim is to preserve intentionality (i.e., to get explanatory parsimony whereby only properties of individuals are causally efficacious). And to demonstrate that a given macroeconomic principle is fundamentally explained by the properties of agents doesn't require that we establish a one-to-one relation between explanans and explanandum.

Using the case study method, I will show that if the macroeconomic phenomenon or principle is explained by individuals' properties exclusively, then there's a reduction. The fact that there exists an irreducible macroeconomic entity, and that the microeconomic level includes other features beyond individuals' properties (e.g., macroeconomic inputs and relative prices), does not raise barriers to reduction necessarily. For reduction, it is sufficient that inside that explanans, only the individuals' properties count as causally efficacious for the rise of the explanandum.

I will argue that in his representative-agent model, Lucas provides us with this type of reductive example, where the macroeconomic phenomenon, the Phillips curve, derives from individuals' properties. All the extra elements inside the microeconomics are non-efficacious. In the course of the discussions, I will also show that reduction is an explanatory strategy with various targets which may adopt distinct forms. My approach in this thesis is distinctive in the way that I utilise and develop ideas from the philosophy of science literature and elsewhere to enlighten the debate about reduction in economics. Among other things, I highlight the importance of explanation as central to reduction attempts in economics.

I treat the goal of reduction in economics, the ambition to privilege the intentionality of agents in economics, as *explanatory parsimony*: where there appears to be more than one type of entity performing the explanation of a given macroeconomic phenomenon, there's, in fact, only one. When the economist, via their models, explains how a macroeconomic

[2] Jerry Fodor's influential paper in favour of the disunity of the sciences stresses identity relations as a necessary condition to endorse reduction. Accordingly, to demonstrate satisfactory correlations between the reducing and reduced terms is not enough. We may have reduction, for instance, of a psychological kind to a corresponding neuroscience kind "[…] so long as we are able to show that the type of distinct neurological events paired with a given kind of psychological event are identical in respect of whatever properties are relevant to type-identification in psychology" (Fodor, 1974, p. 106).

phenomenon/principle is eventually possible based only on the properties of individuals, then we have the main ingredient for reduction. This is a necessary but not a sufficient condition. To get a successful reduction, it must also be the case that any other existing elements inside the microeconomic level (macro data, relative prices) and the irreducible macro entities should be causally inert with respect to the rise of the macroeconomic outcome. For the purposes of explanation of the macroeconomic target, they are mere passive signals – they are the objects of inquiry of individuals' beliefs and expectations. In contrast, if these extra non-individual elements were causally active for the manifestation of a given macroeconomic phenomenon/principle, there would be no reduction inasmuch other features would be as equally important as individuals' properties.

The structure and aims of the discussion

This thesis comprises five main chapters and a conclusion. In the first two chapters, I lay out two reductive frameworks that may fit reduction in economics. As defined in this thesis, the main goal of reduction in economics is explanatory parsimony. The relata of reduction covers (theoretical) mind-dependent representational items (theories, statements, models) and real-world items – events, regularities, natural (and social) kinds, and properties. We may get reductions in purely theoretical contexts. For instance, we may represent the manifestation of a macroeconomic regularity in a fictional world. In such a context, the chosen macroeconomic regularity expresses a theoretical item (e.g., an equation) that does not represent any actual event in the world, even though it mimics its target in the relevant respects.

In economics, a model builder who seeks explanatory parsimony might articulate a fictional world modelling that respect the purpose of reduction in economics. In this case, we may qualify such a theoretical enterprise as a type of theoretical reduction. Contrarily, when scientists step further, moving from a pure theoretical domain to a world domain, they should provide empirical support for their claims. In this context, reductions between real-world items must be empirically grounded. When this is the case, we have ontological reductions. In this thesis, I explore two possible ways to get a reduction: Nagel's theory and a constitutive-mechanistic account.

In the first chapter, I introduce Ernest Nagel's theory of reduction. As I already briefly indicated, the general message of Nagel's reduction account is that reduction is a special inter-level explanation between distinct disciplines. When a reducing science, for instance, neuroscience, via its axioms, principles, and regularities, can explain the regularities of some

higher-level science, such as psychology, then the latter is reduced to the former. This special type of explanation follows the Deductive-Nomological (DN)-model of explanation where the explanans (the elements that perform the explanation of the target phenomenon) consists of regularities plus initial conditions, from which the explanandum is deduced.

In a similar fashion, reduction has the same logical pattern according to Nagel. In Nagel's view, scientific reductions are satisfactorily attained when the laws from the reduced science are subsumed by the laws of the reducing science. However, as long as autonomous disciplines have their own vocabulary, the structure of reducing and reduced theories may not be commensurable. When the relevant terms involved in reductions are commensurable, this could be qualified as a lucky example of homogenous reduction. But most of the interesting cases of reductions across the sciences don't share the relevant vocabulary. For these cases, additional statements are required to map the terms of the reduced domain that are absent in the reducing domain. These bridge principles or reductive functions permit that the reduced regularities might be inferred from the reducing ones. The nature of such bridge principles is the focus of much discussion among philosophers who debate Nagel's view of reduction. I discuss two of these interpretations in chapter 1, Colin Klein's (2009) and Kenneth Schaffner's views (1967, 1993, 2012).

In brief, for Klein, Nagelian reduction is satisfied if the reducing science, through its own theoretical apparatus, can represent the reduced laws. To materialise this goal, the reduction function can be a mere definition responsible for pairing the terms of both theories. Inasmuch as this definition permits the reducing theory to explain the facts in the reduced domain, this will count as a sufficient condition to get a reduction in Nagelian terms. On the contrary, Schaffner claims that reduction functions should express empirical facts – a mere inferential rule isn't enough to satisfy scientific reductions. The difference between Klein and Schaffner is obvious: while the former allows us to identify Nagelian reductions in pure theoretical contexts, the latter doesn't allow this as a permissible interpretation of bridge principles. Consequently, Schaffner's view can only be applied in empirical contexts. It's worth noticing that in this exegetical dispute, there's no contradiction between these two interpretations. Historically, Nagel was a pluralist about the nature of such bridge principles. So, we may perfectly have Nagelian reduction in both contexts, theoretical and empirical. We shall see that for the case study under analysis in chapter 3, it’s Klein’s interpretation, and not Schaffner’s proposal, that is suitable to the purposes of Lucas’s model.

Moving on, I set forth the second reductive contender, the constitutive-mechanistic account of explanation, which I address in the second chapter. This framework is popular in the

contemporary philosophy of science. Crudely put, the idea is that when a phenomenon can be identified as a type of mechanism – a heart, molecule, a ringing bell, political movement, or an economical phenomenon, these exemplars of mechanisms permit mechanists two sorts of analysis: etiological/causal and constitutive analysis. For the purposes of reduction, it's the second aspect that is of interest: it is the constitutive analysis of the mechanism that permits a constitutive explanation and, consequently, a reductive explanation. Constitutive explanations link parts and wholes. They intend to illustrate how the whole has gotten certain features in virtue of its structural items at lower levels of organisation. Unlike causal explanations, which form another class of explanations included in the mechanistic account that express relations of dependence between events, constitutive mechanisms trace relations of dependence between causal properties: powers, dispositions, and tendencies of the structural items inside the mechanism.

Contrary to Nagel's view, I focus on James Woodward's counterfactual theory of explanation as a way to understand the connection between the phenomenon of interest (explanandum) and the elements that explain its occurrence (explanans). Good explanations, in Woodward's view, are those capable of demonstrating counterfactual dependences between explanans and explanandum. As a result, for the purposes of reduction, there's a constitutive-mechanistic reduction if and only if the explanandum (the whole) is explained via counterfactual dependence by its structural items from the next level down (explanans). I am to show that both frameworks, Nagel's view, and a constitutive-mechanistic approach under the Woodwardian counterfactual analysis, can be applied in a case study in economics, namely Lucas's Phillips curve.

In chapter 3, I discuss Lucas's model. In that model, Robert Lucas argues that the correlation between output and inflation – a macroeconomic principle – is determined by suppliers' expectations about the possible values of the general price level. As maximisers who don't know the demand conditions and have only access to relative prices and some macroeconomic data, suppliers can't perfectly discriminate how much the variation in the price of their inputs and outputs represents the conditions of demand or a rise for all the prices in the economy (inflation). The Phillips curve emerges when suppliers produce more or less than is desirable by their customers, misguided by the available cues in the economy. When suppliers can perfectly anticipate inflation, there's no Philips curve. In case they don't, there's a Phillips curve.

I shall present Lucas's model in more detail and consider it as an example of reduction. Lucas's goal is to obtain explanatory parsimony: despite other existing entities included in the

microeconomic level (relative prices and macroeconomic data as inputs), and even though the general price level hasn't been decomposed to its microeconomic components, only the properties of individuals, their beliefs and expectations, are causally efficacious for the manifestation of the correlation at the macroeconomic level. While the inflation-output correlation is the explanandum, only individuals' properties figure in the explanans – other existing entities in the microeconomic structure are causally inert for the rise of the Phillips curve.

Moreover, the irreducible macroeconomic entity, the general price level, isn't included in the explanans, only in the explanandum. The economist doesn't need to address every ontological issue (e.g., decompose every macroeconomic entity) to forge her/his explanations – this is a matter of choice. As this case shows, it's possible to obtain explanatory parsimony even if the ontological obstacles described by Hoover have not been strictly addressed. In this chapter 3, I illustrate how Nagelian reduction, as defended by Colin Klein, and constitutive-mechanistic account under the Woodwardian counterfactual view can be correctly applied in Lucas's model.

Contrary to Hoover, who denies the nearest chance of reduction, I don't think we can justify a categorical statement that there's no chance for reducing macroeconomic principles to microeconomic ones. As explained, I advance the debate about the reduction in economics by offering a detailed response to Hoover's scepticism in this domain by providing a detailed case study that is interpreted through the lens of Nagel's reduction and constitutive-mechanistic reduction. As I said before, the prospects for reduction depend on how we define it, its purposes, and how it can be possibly attained. I think the importance of reduction is primarily explanatory, and the same is true for a non-reductive analysis.

As philosophers, our natural tendency to push the debate towards a more abstract level may risk obscuring several details and specifications of individual cases. Rather than priorly fixing the ontology and then developing a (presumably) most accurate method, the analysis of reductive and non-reductive explanations should be focused on the explanatory relevance of those variables included in the explanans in the context of specific economic models. In the opposite of stating there is no chance for reduction of macroeconomics, we should be more modest by saying, "Well, in this particular case, the sufficient conditions for a positive example of reduction are not matched".[3] Thus, we should be local instead of adopting a global all-or-

[3] In the context of scientific realism debate, Asay (2019) offers a local defense in favour of realism: there is no *one* realism, but many realisms that can only be accounted in the context of individual scientific practices. Asay (2019, p. 603) underlines "[…] we should attend to the details of individual cases". This localist approach suggests that the success-to-truth inference that always permeate the scientific realism positions should adopt a (qualified)

nothing approach. Despite the conclusions of chapter 3, we may still agree with Hoover that, in some economic cases, the explanans involves more than the properties of individuals – non-individual properties count as much relevant as the individuals' features.

Contrary to reduction, in a case study that bridges micro-to-macroeconomics, a non-reductive explanation is identified when the explanans involves individuals' properties plus other extra elements inside the microeconomic level – i.e., a case where the properties of individuals do not prevail over the other elements inside the microeconomic structure. In such cases, both individual and non-individual properties within the microeconomic structure have relevant causal powers which contribute equally to the macroeconomic result. My approach to this discussion is pluralist – sometimes, according to the case studies, we can be reductionists, but in others we should embrace a non-reductionist approach. So, one of the take-home messages of this thesis is that when we are interested in understanding how micro-and-macroeconomics are possibly connected, there is unlikely to be a one-size-fits-all approach.

In chapter 4, I offer a positive example where the macroeconomic phenomenon is unequivocally explained by the individuals' properties plus the extra non-individual elements inside the microeconomic domain. In the philosophy of economics literature, Kevin Hoover is the main advocate in favour of global supervenience: although macroeconomics isn't reducible to microeconomics, every microeconomic arrangement generates macroeconomic outcomes. When the microeconomic level is set, the macroeconomic level is also set. So, if we duplicate the properties in the micro level, we automatically duplicate the macro level. I reconstruct Hoover's defence of supervenience and discuss two criticisms.

Julian Reiss (2004) claims that the influence of macroeconomics upon individuals precludes the application of supervenience. He reminds us that supervenience requires bottom-top determination instead of the other way around. More recently, Brian Epstein (2014) goes further, arguing that macroeconomic facts are autonomous with respect to microeconomic facts. By constructing an economic thought experiment, Epstein claims that it's possible to have two indiscernible microeconomic worlds with dissimilar macroeconomic properties. In this chapter, I answer these two criticisms, arguing that both problems, the downward causation problem and the supposed autonomy of macroeconomics, do not undermine the applicability of global supervenience.

Against Reiss, I show how downward causation and supervenience don't necessarily conflict with one another. These relations, I argue, manifest themselves at different points in

localistic approach about the truth-inferences (Vickers, 2019). I think this idea is also applicable to the context of reduction v. non-reduction debate in economics.

time. By exploring one economic example, Taylor's principle, I demonstrate how the two relations can live in harmony. In response to Epstein, I accept that the direction of causation flows from the top level towards the microstructure and subsequently back up to the top (macroeconomics). However, I argue that Epstein's argument makes sense only if we exclude the institutional features (legal rules) from the microeconomic set. When we re-insert them, we reveal the supervenience relation between macro and micro. Interestingly, this chapter also demonstrates that it's not every supervenience relation between micro and macroeconomics that is explanatorily informative.

In my response to Reiss's concern about downward causation, I retort that, in Taylor's rule example, supervenience and causation occur at different times, but only the latter, not the former, is epistemically interesting. In the example, the monetary committee announces the value of interest rates that is informative for economic decisions of individuals. However, identifying a supervenience relation between the committee's announcement and the members' individual decisions is not important for economic agents.

Conversely, in the case of Epstein's criticism regarding the autonomy of macroeconomics, I claim that the introduction of institutional features inside the microeconomic level is indispensable to explain what is going on at the macro level. If we set aside this institutional aspect from the micro, we lose track of the entire direction of causality. My argument in favour of supervenience reveals its application in economics, showing that supervenience may appear in economics with different features and, most importantly, it's not every supervenience case that is explanatorily illuminating. Sometimes, as Taylor's rule will reveal in chapter 4, supervenience is just a modal relation. So, it's not every supervenience relation between micro and macro that entails explanation. For this reason, I dedicate the last chapter to the topic of explanation.

In chapter 5, I discuss one theory of explanation, namely the (quaternary) contrastive account of explanation. Some scholars have discussed the topic of contrastive explanation in the context of economics.[4] In this chapter, I try to show its importance in this context: what we can learn from this explanatory syntax and how flexible such a framework is, capable of covering a variety of explanations (causal and noncausal ones) and informing us about possible predictions. Contrastive accounts claim that explanation is a matter of demonstrating contrasts among alternative explanatory scenarios for the questions at hand. Contrastive theorists claim that saying that *X* explains *Y* is to say that when we compare *X* against alternative explanatory

[4] Among these are Lawson (2001, 2009), Lewis (2013), Marchionni (2006), Tseng & Elsaesser (2011), Morgan & Patokämi (2017), and Kaul (2022).

scenarios for the occurrence of *Y*, only *X* appears as the unique reasonable possible answer to *Y*: it's *X* rather than *X**, that explains *Y* rather than *Y**.

The contrastive view stresses a crucial point about explanation: we never offer a tout court explanation of events; we only explain certain aspects of events. And what is explained depends crucially on what is being questioned. To disambiguate the questions, contrastive theorists suggest we should insert contrasts in the questions, such as "why has Liz Truss, rather than Rishi Sunak, become the actual Prime Minister?". In response to this question, we can answer, "Because she won the final election rather than the former chancellor". For someone who isn't strongly acquainted with British elections, they might be interested instead in the more general question, "why did Liz Truss, rather than someone else from Conservative Party, become the actual Prime Minister?". Then we should answer, "Because other candidates who were running for Prime Minister were eliminated in the ballots, rather than Liz Truss". In short, according to the contrastive model, causal *and* noncausal explanations demand contrasts on both sides –explanans and explanandum.

In this final chapter, I revisit most of the examples discussed across the thesis to demonstrate that this contrastive framework is consistent with Woodward's counterfactual view and can perfectly cover causal and noncausal cases of explanations. I apply the contrastive model in a new way by showing how we can portray the Lucasian model in such terms, showing in a more transparent way how his model provides a complete explanation of how the Phillips curve is possible by virtue of suppliers' expectations. Moreover, this form of explanation provides an interesting epistemic guide to predictions. Finally, in the case of supervenience, this syntax makes visible that, in Epstein's case, it's only the microeconomic base that provides a direct explanation for the manifestation of the corresponding macroeconomic sets. Unlike causal explanations, supervenience explains, in Epstein's fictional world, by providing constraints, which make predictable the macro phenomena, even though they do not causally fix it. Supervenience, in this case, imposes the limits, the set of conditions for which the external events, like an environmental event, may produce its causal effects on macroeconomics. However, without the institutional apparatus, this would not be possible.

To briefly summarize, the thesis show how reductive models are appropriate in some economic cases while acknowledging that supervenience models may be more applicable in other cases. So, a one-size-fits-all approach to micro-macro economic relations is not likely to work. Nonetheless, all these different models can generate explanations if certain conditions are met, revealing a plurality of inter-level explanations to bridge micro-to-macroeconomics. My approach is also optimistic in terms of the epistemology of economic models: I believe that

formulating models like Lucas does can provide genuine explanations even if such models operate at the level of abstraction. In such cases, we can learn about some modal relations among target entities inside the model that may mimic corresponding relations in the real world. In this thesis, this last issue isn't addressed in great detail, but I shall explain this point in more detail in the concluding chapter of this work.

CHAPTER 1: DIFFERENT ROADS LEAD TO ROME: ERNEST NAGEL'S PATHWAY TOWARDS REDUCTION

1. Introduction

In this opening chapter, I introduce the first candidate for reduction in economics: the well-known account of Ernest Nagel. Nagel's ambition consisted of setting out formal (and informal) criteria whereby a theory – or fragments of a theory – could be reduced to a more fundamental theoretical level of inquiry. Reduction is, in Nagel's view, a special type of explanation that is attained when the empirical laws (the regularities) of the target science can be successfully explained by the empirical laws of primary science. If the target laws from the higher-level science are deducible from the laws of the reducing science, then the latter explains the former. Therefore, reduced laws would be nothing but conclusions extracted from subsumptions whereby the premises are the reducing laws plus initial conditions.

Nonetheless, just as in life, things are not so simple in science: since different theories do not necessarily share the same relevant vocabulary, further tools (reduction functions/ or bridge laws) are typically needed to make reductions possible. Very often, reductions can only go through when the reduced laws are inferred from a set of statements that combine the reducing lawlike statements, boundary conditions, *and* reduction functions – when required. This is the overall message of Nagel's model of reduction.

Reactions rapidly surged in the literature, especially after the publication of *The structure of science* (1961), when Nagel provided a more detailed qualification of what he meant by "reduction". The first wave of criticisms pointed out multiple problems surrounding Nagel's approach (such as (a) the incommensurability problem, (b) the narrow range of application, (c) the impossibility of correction or replacement, and (d) the problem of symmetry).[5] From this initial reaction onwards, new approaches to scientific reduction tried to work out the limits of Nagel's approach in an attempt to overcome Nagel's original boundaries and (potential) failures.[6]

[5] The Incommensurability problem was raised just one year after the publication of Nagel's book by Paul Feyerabend (1962). Challenges over (b) were stressed by philosophers of science (but especially those from philosophy of biology) such as Wimsatt (1972), Hull (1976), and Sarkar (1992) who argue that their disciplines do not fit into the picture painted by syntactic theories. Contentions against (c) were stressed by many writers, including Feyerabend (1962), Churchland (1986), Schaffner (1993), and Bickle (1998). Regarding (d), that criticism proposed that once Nagel's reductive approach is committed to a DN-model of explanation, then Nagel's reduction would also fail. That criticism was latterly responded to by Schaffner (1993).

[6] Scholars such as Churchland (1979, 1985), Hooker (1981), and Bickle (1998) provide similar reductive approaches as Schaffner's (1967). The revival of discussions about reduction became known as new-wave reductionism. Unlike Nagel, these approaches, specially Churchland (1986) and Bickle, have an eliminativist

More recently, the first two decades of the 21st century have witnessed a new wave of debates surrounding Nagel's reduction. But at this time, a new generation of scholars have stepped into Nagel's shoes, preserving the nucleus of his legacy: reduction is a special sort of inter-level explanation that, whenever possible, appeals to logical inferences between reducing and reduced domains with auxiliary adjustments when required.[7] Although those initial criticisms have identified multiple problems with Nagel's view, they didn't challenge the sufficient conditions (derivability and connectability) for scientific reductions according to Nagel's recipe.

The resurrection of Nagel's view demonstrates its actual relevance to the contemporary subject of reduction. Aside from the general agreements, there are also disagreements on the table: (i) which kinds of things Nagelian reductions should accept as their relata, (ii) the nature (ontological status) of reduction functions which, in turn, leads to two distinct versions of reduction, and which conditions are necessary and sufficient to get reductions through according to Nagel's recipe (iii).[8]

In the context of economic science, to evaluate whether the Nagelian framework may successfully be applied to reduce macroeconomic theory (or some of its fragments) to the microeconomic level in some cases, a good strategy is to consider a study case (i.e., involving some economic model) that logically infers the behaviour of a macroeconomic generalisation due to microeconomic laws and principles and then judge if, with the auxiliary of the right aggregation procedures, the macroeconomic phenomenon follows from a particular set of microeconomic laws and principles:

> Microeconomics is the reducing theory, macroeconomics is the reduced theory, and the aggregation procedures serves as "the reduction function" which provides the connections necessary for the deduction of the latter theory from the former. If one suspected that some theory were reducible to another in this way then, of course, one would have two theories on the table and the salient question would be: What are the bridge laws? (Nelson, 1984, p. 581)

To understand the world we live in, economists create economic models, some pictorial worlds, and thought experiments to assess relations of dependence that supposedly exist in the real world. Usually, economists generate theoretical and empirical models: the former are

perspective when applied to the boundaries between psychology and neuroscience. So, neuroscience will nearly replace every psychological explanation. Nagel's approach is not committed to eliminativism (see section 3).

[7] Among the recent defenders of Nagel's approach are Fazekas (2009), Klein (2009), Dzadji-Bahmani et.al. (2010), Needham (2010), Butterfield, (2011a, 2011b), van Riel (2011), Schaffner (2012), and Sarkar (2015).

[8] Van Riel (2011) treats Nagel's nonformal conditions for reduction as much necessary as the formal conditions to enact reductions. Herein, I treat nonformal conditions as epistemic ambitions. In this regard, we may have a Nagelian reduction even if the five nonformal clauses are identifiable in the case under assessment. For a response to van Riel's argument, see Sarkar (2015).

idealised pictures of the world, whereas the latter employ econometric tools to test their theoretical models (Boland, 2017). For the quest of reduction analysis in the context of economics, the interesting point is to investigate whether the aggregation functions associating the chosen micro-to-macro quantities in the model capable of enacting reduction *a la* Nagel are theoretical associations (theoretical models) or if they represent something else like material or factual claims (empirical models).[9]

Similarly, Nagel wondered whether the connections between reducing and reduced theories are possible at the representational level – i.e., the relata of reductions would be limited to the theoretical assumptions/concepts/statements across reducing and reduced theories – or it may step further towards the observational level where the relata of reductions depend on what the world is like.

If Nagel's account can be successfully applied to some economic models that bridge micro-to-macroeconomics, two scenarios are possible. One where the case of study articulates an entailment of a macroeconomic phenomenon from a set of microeconomic principles with the auxiliary of reduction functions whose nature expresses conventions, i.e., some definitions adopted by the economist grounded in economic theory.

In the chosen study case I shall discuss in chapter 3, Lucas's Phillips curve model, the bridge between the two economic layers, the micro-and-macroeconomics, is channelled via a *definition* adopted by Lucas. Therein, I show that the general price level, a macroeconomic entity, is treated as a summary of microeconomic components – an average of prices of consumer goods and services relative to a given period. From an empirical point of view, the relationship between the general price level and relative prices aren't taken for granted as far as the price of particular goods and the general price level may have independent causes. But this isn't a concern for Lucas: his modelling is a theoretical exercise designed to demonstrate which economic conditions may possibly determine a macroeconomic phenomenon such as the Phillips curve, the correlation between inflation and output.

Alternatively, if we choose an empirical economic model to evaluate the connection between micro and macroeconomics, then the macroeconomic assumption, the outcome of the model, should express a factual statement logically deduced from a set of microeconomic

[9] Historically, many economists believed that the best (and perhaps, the *only*) way to provide solid microfoundations of macroeconomics is to solve the problem of aggregation: which relationship exists between a macro entity and its parts or, more generally, between microeconomic theory and macroeconomic theory (Rosenberg, 1976, p. 181). In an influential article, Lawrence Klein (1946, p. 93) argues that the relationship between aggregative theories [macroeconomic theories] and individual behaviour is, to a large extent, a problem of proper measurement. In chapter 3, I demonstrate why the aggregation problem is not a problem for the application of Nagel's reduction in economics.

assumptions. In such an account, every reduction function and the bridge principles used to aggregate the micro-and-macro quantities across the two layers must exhibit factual claims. This dichotomy between empirical v. pure theoretical approach also appears in debates around the nature of Nagel's reduction functions.

Colin Klein (2009) believes that Nagelian reduction functions, i.e., the set of auxiliary assumptions used by the reductionist who attempts to explain the experimental laws from the higher-level (target) science via the theoretical apparatus from a more fundamental level of inquiry (the reducing science), don't need to express factual claims necessarily. It's sufficient for Nagel's reduction, Klein argues, that the reducing theory can represent the target reduced laws through its own tools.

Kenneth Schaffner (1967, 1993, 2012), on the contrary, rejects this view, arguing that reduction functions, the connection between reducing and reduced theories, should express empirical claims. From a scientific point of view, relevant reductions are empirically grounded, and there's no reason to accept that a mere definition or convention should count as a robust scientific reductive function. By their own nature, conventions are arbitrary, and all that matters is that they are respected after a choice has been made (Dzadji-Bahamani et.al., 2010, p. 403). Therefore, unlike Klein's view, the Schaffnerian view of Nagel's reduction has a proper application to the context of empirical economic models designed to channel objective relations of dependence between macro events and microeconomic principles. As such, these two approaches entail different versions of Nagel's reduction: Klein's is happy with epistemic reductions – reductions are fine at the representational (theoretical) level. Schaffner, in his turn, demands a more stringent condition: it depends on what goes on in the world, entailing an ontological requirement. We shall discuss this debate between Klein and Schaffner in greater detail in section 4.

All this philosophical discussion is a prelude to a later examination of the classical Robert Lucas's Phillips curve (in chapter 3). As we shall see in chapter 3, the case study, Lucas's model, is an inter-level type of explanation where the model's outcome represents a macroeconomic phenomenon that is logically inferred from a set of microeconomic principles plus reduction function assumptions. In this initial chapter, I concentrate on Nagel's view so that we can then apply this reductive recipe to Lucas's model.

This chapter is structured as follows. First, I set forth Nagel's formal requirements for reductions in more detail (Section 2). Next, I present the informal conditions that may make reductions attractive to scientists and which every reductive enterprise should pursue in order to be a relevant one, according to Nagel (section 3). Then I present the exegetical disputes on

the nature of reduction functions (Section 4). I first introduce the epistemic interpretation of the nature of reductive functions, as argued by Colin Klein (Sub-section 4.1). Finally, I lay out Kenneth Schaffner's account of reduction functions (Sub-section 4.2). As I shall later argue in chapter 3, it's Klein's interpretation of the nature of Nagel's reduction function, not Schaffner's requirement, that permits us to identify Lucas's model as a genuine example of Nagelian reduction in epistemic terms. Nonetheless, I shall also lay out Schaffner's alternative view by way of contrast.

2. Ernst Nagel on reduction: the formal conditions

To Nagel, reduction belongs to the class of explanations.[10] He defines it as a special type of inter-level explanation in which a secondary science or its set of experimental laws is explained by a set of laws and axioms from a more fundamental theoretical domain (Nagel, 1961, p. 338). Schaffner (2012, p. 535) tells us that Nagel's motivation was mainly triggered by the rapid advances in statistical mechanics. In the early decades of the 20th century, the achievements in statistical mechanics promised that this novel science could explain the laws of thermodynamics. If this new emerging science could explain and represent the complex phenomena that were formerly explained exclusively by thermodynamics, then we may have an empirical example of scientific reduction between different branches of inquiry that overlap the same scope of application. What is therefore required is a proper logic to accommodate this and other potential cases of scientific reduction across the sciences.

In the spirit of logical positivism, Nagel conceived reduction as an (inter-level type of) explanation based on the logic of a deductive-nomological account of explanation (Huttemann and Love, 2011, p. 525). The idea, exhaustively discussed by Carl Hempel, states that to explain an event E_2 by reference to a former event E_1 requires an appeal to empirical laws (or general propositions) capable of correlating events of the type to be explained (the *explanandum* sentence) with events of the type cited as its causes or conditions (the *explanans* sentence).

According to Hempel, explanans and explanandum are the major constituents of scientific explanations. In this account, scientific explanations are arguments whereby we explain the occurrence of a singular event (the explanandum) by describing the antecedent conditions of that event with the auxiliary of (at least) one lawlike sentence:

[10] Nagel stresses this issue, i.e., reduction consists in a special inter-level explanation, in several writings across his career. In his first paper dedicated to discussing the logic of scientific reduction, he introduced that notion (Nagel, 1935, p. 46). A more careful definition is cashed out in his classical debate over reduction in 1961. Later on, he again treated reduction as a special sort of explanation (Nagel, 1970, p. 119).

Premises: $L_1, L_2, \ldots L_m$

$C_1, C_2, \ldots C_n$

Conclusion: E

In the scheme above, $C_1, C_2 \ldots C_n$ represent the initial conditions, $L_1, L_2 \ldots L_m$ represent the potential laws (general lawlike statements) applicable to the case in question, and E is the event we are seeking to explain. Essentially, explaining E consists in combining general statements plus boundary conditions that entail E's occurrence. Assuming the explanans are true, the explanandum arises as a logical consequence: if the premises are true, then the conclusion shall necessarily be true:

> A *DN* explanation answers the question "Why did the explanandum-phenomenon occur?" by showing that the phenomenon resulted from certain particular circumstances, specified in $C_1, C_2, \ldots, C_k$, in accordance with the laws $L_1, L_2, \ldots, L_r$. By pointing this out, the argument shows that, given the particular circumstances and the laws in question, the occurrence of the phenomenon *was to be expected*; and it is in this sense that the explanation enables us to *understand why* the phenomenon occurred. (Hempel, 1965, p. 337; italics in original)

Hence, we may say the explanandum (the event to be explained) was not merely explained; it's also expected to happen. In this account, explanations are connected with predictions – while explanations look backward at why the explanandum occurred, the latter look forward:

> the same formal analysis [...] applies to scientific prediction as well as explanation [...]. An explanation is not fully adequate unless its explanans, if taken account of in time, could have served as a basis for predicting the phenomenon under consideration. (Hempel and Oppenheim, 1948, p. 138)

In this regard, explanations would be later predictions (Douglas, 2009, p. 448), what has been traditionally named the symmetry thesis – as far as explanations and predictions hold the same logical form, the two would be symmetrically related. Hempel assumed that his mode of explanation could safely guarantee an asymmetrical relation between the explanans and the explanandum as long as the explanans cites causal laws: if A causes B, then B cannot count as the cause of A.[11]

Likewise, Nagel envisioned reductions in the form of arguments. He suggested that, *ideally*, all the axioms and experimental laws of the target (the reduced) sciences involved in reductions should be conceived in terms of statements (Nagel, 1961, p. 345). Nagel had in mind a syntactic view of what a theory is. This view of scientific theory, also popularly known as the

[11] Schaffner (1993, p. 267) suggests that, pragmatically, nomological DN-explanations might appeal to metaphysical presuppositions to preserve the asymmetry in reductions: saying the evidence supports that (1) C causes E and not the other way round or affirming that (2) C, in comparison with E, belongs to a lower-level of organisation (ibid. p. 431), may explain, semantically, the direction of causality.

received view, considers a theory as an axiomatic system of propositions formulated in first-order logic and a set of (theoretical, logical, and observational) terms.[12] Once this (desirable) pre-condition is satisfied, then the formal basis to warrant deductions could carry it out more properly. Thus, a reduction shall be effective when

> […] the experimental laws of the secondary science (and if it has an adequate theory, its theory as well) are shown to be the logical consequences of the theoretical assumptions (inclusive of the coordinating definitions) of the primary science. (Nagel, 1961, p. 352)

Basically, a target theory T_t is reducible to a lower-level theory T_b *iff* T_t is derivable from T_b. This is the first formal condition to get reductions: the *condition of derivability*. If the relevant laws from the target theory are derivable from the laws from the primary theory, then we may get reductions.

The greatest advantage of deduction is the preservation of truth. In this regard, we would have the lawlike statements from reducing sciences figuring in the explanans, and the conclusion must be the reduced lawlike statements (thus, the explanandum). We transfer the truth-value of the explanans to the explanandum. In satisfactory exemplars of reduction, the reducing statement replaces the reduced one for the purposes of explaining the facts from the very scope of the reduced domain. Nagel calls this type of reduction homogeneous: when the reduced and reducing sciences share the relevant vocabulary necessary for deduction. Apparently, homogeneous reductions – at first glance – should not impose great obstacles.

In homogeneous cases, their premises are composed of the assumptions that represent the reducing laws, while the conclusions are the reduced ones. As far as there is no missing term in the reduced theory that imposes additional challenges to deducing higher-level laws from the laws of the reducing theory, then reductions might go through. One example of homogeneous reduction is the explanation of Kepler's laws of motion for celestial bodies using the theoretical framework initially conceived by Newtonian mechanics. Both theories share significant vocabulary (e.g., equivalent notions of crucial terms such as "distance", "time", and "acceleration") and even their metrics, admitting thus that depending on which sort of phenomena physicists are aiming to explain about the behaviour of celestial bodies, they can, in principle, deduce the laws that govern these celestial bodies from the laws of Newtonian physics.

[12] Several authors (Wimsatt, 1972; Suppe, 1977; Waters, 1990) contended that the idea of theory reduction according to the syntactic view raises an unjustifiable barrier for several disciplines (e.g., biology, neurosciences) which don't follow the syntactic view. However, the formalisation of theories in terms of first-order predicate logic was not a *necessary* pre-condition for reductions; it's a *desirable* pre-condition. Dzadji-Bahmani et.al. (2010) provide an inter-level explanation which fits Nagel's formal requirements that doesn't require the formalisation of first-order predicate logic. Thus, the syntactic approach is not strictly required for reductions in Nagelian terms.

Nevertheless, in his later writings, Nagel acknowledged the limits in achieving reductions from direct inferences even for successful homogenous cases of reductions. In practice, what is normally deduced in satisfactory homogenous reductive examples is not the exact laws of the reduced theory. The sine-qua-non condition that defines reduction as an inter-level explanation via deductions imposes a very stringent requirement that would rarely be fulfilled in practice (Needham, 2010, p. 168). Instead, the reduced laws are, at best, *approximations*:[13]

> It is undoubtedly the case that the laws derivable from Newtonian theory do not coincide exactly with some of the previously entertained hypotheses about the motion of bodies, though in other cases there may be such coincidence. For it is a widely recognized function of comprehensive theories [...] to specify the conditions under which antecedently established regularities hold, and to indicate, in the light of those conditions, the modifications that may have to be made in the initial hypotheses, especially if the range of application of the hypotheses is enlarged. Nevertheless, the initial hypotheses may be reasonably close approximations to the consequences entailed by the comprehensive theory, as is indeed the case with Galileo's law as with the Kepler's third law. (Nagel, 1970, p. 120-121)

Nagel relaxed his original formulation: reductions are deductions of reduced laws from the reducing domain that is approximately the same as the laws being targeted. But Nagelian reductions have another formal challenge: the incommensurability of terms. Different scientific theories do not normally share the same vocabulary. Theoretical terms are theoretical-laden, which may make deductions, and hence reductions, impossible. So, Nagel sorted reductions into two types: homogenous and heterogenous reductions.

Heterogeneous reductions are the usual exemplars of reductions in science. Unlike the homogeneous reductions, these present plural typologies and theoretical terms across the primary and secondary sciences. To the extent that reduction is a special sort of explanation that mainly follows the footprints of the DN-model of scientific explanation, adjustments are required to derive a logical conclusion that encompasses absent terms in one of the theories involved. For those cases, they must attend to a *condition of connectability*:

> [...] when the laws of the secondary science do contain some term "A" that is absent from the theoretical assumptions of the primary science, there are two necessary formal conditions for the reduction of the former to the latter: (1) Assumptions of some kind must be introduced which postulate suitable relations between whatever is signified by "A" and traits represented by theoretical terms already present in primary science. (Nagel, 1961, p. 353-354)

The requiring adjustment claimed by Nagel is the introduction of an auxiliary assumption that may be required to make deductions possible. These are the reduction functions, the inter-level statements that bridge the primitive terms across reducing and

[13] Nagel is silent on what he meant by "approximation". I think that Norton's (2012) definition, where he takes the term as an unprecise description of a target system, is a good characterisation of what Nagel had suggested.

reduced domains. In the quotation above, Nagel uses the verb "introduce", suggesting that if the reducing science has the right resources to represent "A" from the reduced science, then the two are connectable.[14]

Historically, the nature of these auxiliary assumptions raises the most controversial topic that surrounds Nagel's theory. Nagel was a pluralist about the status of these connectors. He believed that "[...] there appear to be just three possibilities as the linkages postulated by these additional assumptions" (Nagel, 1961, p. 354). He didn't deny that perhaps there could be other options. The natures of such assumptions are (a) logical or meaning connections, (b) conventions/stipulations or (c) factual/material relations.

In (a), the links would be established by virtue of the meanings involved in the derivation. In this form, the connectability assumption connecting "A", a term from reduced science, to term "B" from the reducing science might express a type of meaning equivalence ("A" and "B" can be synonyms) or presents some sort of entailment. When applied to a gas, the expressions "temperature" and "mean kinetic energy" designate the same referent; nonetheless, they originally belong to different theoretical branches.

The second candidate (b) admits that different theoretical terms across reducing and reduced theories could be paired via stipulations/conventions by deliberate fiat. Definitions can be satisfactorily claimed for (reduction) cases whereby the primary science has the appropriate theoretical vocabulary to introduce the reduced term via its own tools. If, for instance, a term "*A*" that belongs to the reduced science is an observational one, the connectability assumption shall assign the observations of the properties of "A" to a particular theoretical assumption from the reducing science.

Finally, the third contender (c) takes reductive functions factually as experimental hypotheses that must be empirically grounded. In this case, the reduction function is a statement grounded on evidence that indicates the occurrence of a state of affairs signified by the reducing theory's expression "B" constitutes a sufficient condition for the occurrence of the state of affairs "A" in the reduced theory (Nagel, 1961, p. 354). Among those candidates, Nagel believed that option (c) was the most promising candidate:

> [...] such bridge laws are empirical hypotheses concerning the *extensions* of the predicates mentioned in these correspondence rules – that is, concerning the classes of individual things or processes designated by those predicates. An attribute of things connoted by a predicate in a reduced law may indeed be quite different from the attribute connoted by the predicates of the reducing theory; but the class of things possessing the former attribute may nevertheless coincide (or be included in) the class of things which possess the property specified by a complex predicate in the reducing theory. (Nagel, 1970, p. 126-127)

[14] The reducing theory's capacity of representing the absent terms from the reduced theory is taken by Klein (2009) as the sufficient condition to get Nagel's reduction. You will see his argument in detail in Section 4.

If reductive functions can bridge reducing and reduced theories extensionally, we may infer that those functions link the very ontologies of primary/secondary sciences, respectively. In conclusion, what is actually being bridged is not the theoretical terms or assumptions only, but also the reduction functions would channel the very properties of the terms included in reduction functions. As indicated in the introduction, in the disputes around the nature of such reductive functions, Colin Klein opts for the second candidate as the sufficiency condition to get Nagelian reductions: definitions or conventions that provide a rule of mapping between reducing and reduced theories. On the contrary, Schaffner adopts a more stringent criterion, arguing for a reductive function that is empirically grounded. We shall discuss each of these arguments in more detail in section 4, but for now, I shall layout the desirable epistemic virtues that Nagel proposed as the main guide of every reductive enterprise.

3. Desirable epistemic virtues in scientific reductions: the informal conditions

Besides the formal conditions, Nagel highlights some informal requirements to consider an exemplar of reduction as particularly good. These requirements are ambitions, a set of ideals to which every reductive scientific enterprise should aspire. Without these requirements, Nagel argues, reductions could be otherwise condemned to mere formal exercises without expanding scientific knowledge. The list below includes five of these nonformal conditions that should be (ideally) met in every reduction:

> (I) The empirical laws from the reducing science cannot be *ad hoc* assumptions. To figure in the explanans of scientific reductions, theoretical assumptions from reducing sciences should have some degree of probative force, i.e., be somewhat supported by empirical evidence (Nagel, 1961, p. 358).
>
> (II) The evidence of the reducing lawlike statements must be gathered *independently* from the reduced theory (ibid., 358).
>
> (III) Relevant reductions should augment our knowledge of the reduced theory (ibid., p. 359).

For instance, the mere fact that the Boyle-Charles' law could be derivable from the laws of kinetic theory might not, *per se*, count as a significant finding for physicists. They may argue this deduction can solely be applied to ideal gases. We cannot extrapolate this idealised deduction of the Boyle-Charles' law for gases at lower temperatures than their points of liquefaction. Alternatively, what physicists may count as a theoretical breakthrough is when the

ideal assumptions from the kinetic theory can be replaced by other kinetic laws while still preserving the inferences across reducing and the reduced theories, increasing our knowledge of the connections across both theories:

> […] instead of the stipulations with the aid of which the Boyle-Charles' law is derivable from the theory, we can assume that the dimensions of gas molecules are not negligible when compared to the mean distance between them, and that in addition to forces of impact there are also cohesive forces acting upon them. It is then possible to deduce from the theory employing these more complex special assumptions the van der Waals' law for gases, which formulates more adequately than does the Boyle-Charles' law the behavior of the ideal and nonideal gases. (Nagel, 1961, p. 360)

Despite their formal merit, deductions don't necessarily enhance scientific knowledge; not every inter-level inference can lead the scientific community towards new directions or novel promising research agendas. Reductions are meaningful when they somewhat enlighten reduced theories and augment or even correct them. Interestingly, this point suggests that Nagel's theory didn't intend to eliminate the reduced theories (Sarkar, 2015, p. 53). Reducing sciences are in the business of empowering our knowledge over the secondary sciences, not exterminating them, which leads us to the two final informal conditions:

(IV) Significant reductions provide brand-new patterns of dependence between the target laws from reducing and reduced sciences, supplying a unified explanation across reducing and reduced domains. As such, any evidence for the laws of the secondary sciences shall count indirectly as evidence for the laws of primary sciences (Nagel, 1961, p. 361).

(V) Reductions should lead the scientific community towards more powerful predictions (Sarkar, ibid., p. 53). After all, the pursuit of more reliable predictions is one of the greatest targets of scientific inquiry:

> The purpose of scientific inquiry is not only to describe and catalog, or even explain, that which is present to everyday experience, but to facilitate prediction, intervention, control, or other forms of action on and among the objects of nature. (Longino, 2002, p. 124)

To the extent that reduction, according to Nagel, holds the same logical form as predictions, fruitful reductions are those that permit accurate expectations.

In summary, relevant reductions are those that (a) have some probative force, some evidential support to endorse them and avoid *ad hoc* assumptions; where (b) the truth of reducing statements is established independently from the reduced domain; they (c) enhance knowledge about the reduced science; they (d) shed light on new patterns of dependence

between the target theories; and, most importantly for economists, they (e) boost our forecast knowledge, amplifying our preditive power.

4. The nature of bridge laws

In the contemporary philosophy of science, there are various accounts of reduction.[15] Plainly put, these multiple views on reduction can be typically sorted into two types: (a) epistemic and (b) ontological reductions. To characterize one or another type of approach, the account of reduction must answer what types of things it links (Van Gullick, 2001, p. 2). Depending on the types of things we could admit, we may have different sorts of reductions. I adopt the following classification to distinguish ontological from epistemological types of reduction:

(i) Ontological reduction relates real-world components such as objects, kinds, substances, properties, events, or individuals.
(ii) Epistemic reduction relates representational items such as theories (or fragments of theories), concepts or models.[16]

In his interpretation of the nature of Nagel's connectability assumptions, Klein (2009) has in mind an epistemic approach regarding the status of reduction functions. Reduction functions can be seen as instruments, i.e., they don't have a lawlike status. In such an instrumentalist version, reduction functions are like rules, i.e., inferential tickets for inferring statements from the reduced science (Nagel, 1970, p. 123). In such a view, reduction functions are judged by their representational power: if the reducing science, using its own apparatus, can introduce the missing term from the reduced science, then the two are connectable. Kenneth Schaffner, on the contrary, argues that reduction is only successful when the reduction functions bridge the very extensions across reducing and reduced theories. Although Schaffner himself does not qualify his adaptation of Nagel's approach as being an ontological version of reduction, his interpretation paves the way towards an ontological reduction since the channel across the theories is channelled extensionally (Fazekas, 2009, p. 316). In the next sections, I lay out the arguments of Klein and Schaffner. As we shall notice in chapter 3, it's Klein's interpretation that fits best for the purposes of Lucas's model, instead of Schaffner's view that would require an empirical economic model.

[15] For a summary of these theories of reduction, see van Riel and van Gulick (2019).
[16] A similar scheme has been coined by van Riel (2014, p. 19).

4.1 The epistemic status of bridge laws: Colin Klein's view

Colin Klein (2009) reads Nagel's condition of connectability epistemically. Nagel's connectability, according to Klein, is satisfied when the reducing theory can fairly *represent* or *introduce* terms from the reduced theory utilising its own theoretical apparatus. This condition does not strictly impose an identity relation between the very theoretical terms from the reducing and reduced sciences. To satisfy connectability,

> […] the condition of such connection is co-reference: connection between the two representational items across the two theories is enough. Nagelian bridge laws should pair theoretical expressions from different theoretical domains in a way that once they have been paired, they would both pointing out the same state of affairs. (Fazekas, 2009, p. 305)

For instance, terms like "heat" or "temperature", which pertain to thermodynamics vocabulary but are missing in statistical mechanics, are admissible into the scope of this latter science if, through its lexicon, statistical mechanics can represent them (Nagel, 1961, p. 353-354). Accordingly, both theoretical terms are connectable if the very referents of "mean molecular kinetic energy" and "heat" can both *amount to* the same state of affairs. However, as we shall see below, not all cases have this simple structure, whereby a single term is co-referential with another single term. In some cases, the connectability relation will be many-one, meaning that the reduction cannot be a straightforward case of identity, as Klein argues.

The metaphysical interpretation of Nagel's connectability condition comes from an apparent natural consequence of relating the representational target items (referents) of reductions. Once the representational items have been paired, this mapping would automatically point out the very extensions of the terms included in the reduction function statements. Connectability can be attained at the representational level without unlocking an automatic identity relation across reducing and reduced theories at the *observational* level: "[…] it is possible to introduce a referring term using complex descriptions in such a way that no metaphysical relationship need hold between the referent of the terms employed in the description and the referent of the term introduced" (Klein, 2009, p. 42-43).

Consider chemical salts. Sodium chloride (NaCl), potassium dichromate ($K_2Cr_2O_7$), and calcium chloride ($CaCl_2$) are all examples of salts. They are composed of an acid and a base in which the acid possesses a positive ion (cation) bounded to a negative ion (anion) holding in the base. Despite all the special properties that each of these salts has (distinct colours, potential implications for the human body under consumption, etc.), they all share the common property of *being a salt*, a semantic predicate (or semantic value) that, according to chemistry, we can confidently ascribe to any instantiation of these chemical compounds. Whenever the predicate

being a salt truly applies to a particular – for instance, a table salt placed alongside the oven, the exemplar shall instantiate the corresponding property.

In principle, contemporary physics can fully explain every generalisation about salts, making them explainable and consequently reducible to physics. For example, we can explain why salts are soluble in water, pointing out the polar character of water capable of breaking salt's ionic bonds. To reduce "salt" into physics vocabulary, according to Nagel's recipe, it requires that the typology of physics can fruitfully provide a co-referential expression using its lexicon to represent the term "salt". A potential strategy is to cash out a reductive function, a *salt bridge definition* (SBD):

(SBD) Something is salt *if and only if* it is a cation ionically bound to an anion.
(SBD) is an expression (or a statement) which allows physics to represent the set of salts, and so is a candidate for introducing a term "salt" into physics. (Klein, 2009, p. 43)

We have three elements to represent the term "salt" to physics: (1) the association of a cation ionically bounded to an anion (the theoretical expression that comes out from the reducing theory), (2) the salt bridge definition – the reducing function that may pair theoretical expressions from distinct theories – plus (3) the theoretical expression from chemistry (*being a salt*). Assuming that the assumption "Salt is nothing but a cation ionically bound to an anion" is true, the expression "being a cation ionically bound to an anion" would point to the same state of affairs of *being a salt* via the auxiliary of SBD – (1) amounts to (3) via (2).

Nevertheless, it is worth noting that the reducing theoretical expression is at best referring to an assemblage of properties bounded to a relation: there is the property of (a) *being a cation*, and (b) *being an anion*, being channelled through an ionic bonding – a *relation* bridging these two properties. In this regard, we have a many-to-one relation in terms of properties: a pair of properties plus another relation to a single property (*being a salt*). As far as we take a 1:1 proportional property identity relationship as a necessary condition to achieve reduction between properties, then we cannot say we reduce "salt" to physics vocabulary, despite the ample evidence that acknowledges that contemporary physics can successfully explain several facts about salts.[17] As a consequence, a stringent metaphysical identity criterion to make reductions go through would blind our eyes to what scientists are doing.

[17] Robert Causey (1972) was one of the first authors who claim that predicate or property identities are necessary for reductions.

Nagel's connectability, Klein argues, is a less stringent criterion. It can be satisfied with a co-referential relation among theoretical expressions that allow scientists to explain facts from the higher-level theory through the instruments of more fundamental sciences.[18] However, even when strict 1:1 property-to-property identity is impossible to obtain, reducing sciences can satisfactorily refer to the very properties of reduced domains, as this case illustrates. If from the laws of physics, plus the auxiliary of SBD, we can describe the behaviour of salts, then we achieved a reduction in Nagelian fashion. A set of statements from a reducing theory being capable of representing the terms from the reduced domain is sufficient for reduction.

As Nagel claimed, the *role* of connectability is operative: it is conceived to make reductions possible, to build bridges, to make sciences like physics capable of talking about chemical compounds like salts. According to Klein, Nagel's reduction is, in this regard, an epistemological approach as far as the connectability is concerned with the reducing theory's capacity to represent the reduced theory's laws or statements. The reducing science should have the power to express the properties of the reduced science: "N-connectability [*Nagel's connectability*] is [...] a relationship between the reducing science apparatus and the world, while metaphysical connectability is a relationship between two (overlapping) parts of the world" (Klein, 2009. p. 42).

A common criticism of Klein's interpretation is that correlations fail to explain. In his defence, Klein retorts that this would conflate the syntactic form that Nagel's connectability normally assumes (biconditionals) with their actual role in reductions:

> But this is to confuse the hypothesis about the syntactic *form* that bridge definitions should take (i.e., biconditionals) with the role that bridge principles play. Again, bridge principles are there to *introduce* a term into the reducing science which is coreferential with some term in the reduced science. This is already a stronger requirement than mere correlation, whatever the syntactic form of the bridge principle might suggest. (Klein, 2009. p. 49)

Klein's view is primarily contemplative. He recommends that it is worth looking at what scientists are doing instead of potentially excluding reductions by requiring identity between properties as necessary conditions. There is no mandatory prescription or stringent criterion about what sciences (in general) should be doing amid inter-level research investigations. The permissiveness aspect does not exclude those special sciences that might consider the syntax of bridge laws solely, while biconditional statements are the necessary criterion for reductions. On the other hand, a robust stringent requirement might exclude successful reductive

[18] One example which reinforces this idea was given by Lawrence Sklar (1993, chapter 9) where he argued why temperature cannot be considered as identical to mean kinetic energy.

enterprises across the sciences. As we shall see in chapter 3, this is what would exclude cases like Lucas's model from being seen as a positive example of reductive explanation.

4.2 Bridge laws as synthetic identities: an ontological consequence

Kenneth Schaffner embraces the general message of Nagel's recipe: reduction is an inter-level explanation that appeals to deductions and, very often, to bridge principles (Schaffner, 1967). However, unlike Nagel, he took some concerns surrounding Nagel's account seriously, especially issues regarding the incommensurability of terms. Since his first article on reduction in 1967, Schaffner's view continues to be discussed as a "sophisticated Nagelian schema" (Batterman, 2016, section 1). He believed that – as Nagel himself later acknowledged as well – the commensurability issue undermines the subsumption of statements with different vocabularies. In heterogeneous reductions, the theoretical terms that make up the reduction functions are non-commensurable.

Incommensurability means non-common measure. In the last century, the incommensurability thesis gained its popularity mostly due to the work of Thomas Kuhn. In his classical *The Structure of Scientific Theories* (1962), Kuhn contends that theoretical terms gain their meanings according to the theory they are embedded in. In the history of ideas, some concepts appear in different theoretical apparatuses (e.g., the concept of "motion" that is present in Aristotelian physics and other theoretical paradigms like Newtonian physics), but this doesn't mean they necessarily signify the same thing in each of these different theories. Therefore, the acknowledgement that scientific terms are fundamentally theory-laden suggests that, very often, scientists can't translate terms from one scientific paradigm to another one. That's why heterogeneous reductions impose a great challenge for Nagel's reductions.

Schaffner believes we can't pair reducing and reduced domains before we *correct* the target domain. Summarising, the target reduced theory T_2 and its corrected version T_2^*, belong to a set K of the same theory:

$$K = \{T_2, T_2^*, T_2^{**}, \ldots\}$$

For each potential pair of theories from set K, there's approximate equality, close analogy, and close agreement. These semantic/qualitative features impose some informal requirements for analogue systems to be accepted as a member of K. With this strategy, Schaffner tried to respond to the problem of incommensurability: it's the corrected version of the reduced theory

which is commensurable with the original reducing theory T_1. According to Schaffner, the first condition for reduction is that, in homogenous reductions, all the primitive terms of the corrected reduced theory T_2^* appear in the reducing domain T_1 or, as it happens in heterogeneous reductions, are associated with one or more terms from the reducing domain *if and only if*:

(1) It's possible to set up a one-to-one correspondence between the kinds, terms, or predicates across the reducing and the corrected reduced theory.
(2) All the primitive predicates of T_2^* must be effectively associated with an open-sentence, a statement that contains one blank or unknown value that becomes true-or-false after the blank is filled, from the reducing theory. For instance, to reduce some fragments of biology to chemistry, it's required that whenever the biological predicate *B* has been inserted, *B* must be replaced by a chemical predicate *C* (or combination of terms), which preserves the truth value (of course, assuming *B* and *C* have the same extension).
(3) Conditions (1) and (2) must have empirical support, representing referential identity that are syntactically channelled via the relevant biconditionals. This condition diverges from Klein's view, in which the co-reference doesn't require empirical support for reductions to go through. When the open sentences that make the bridge between the corrected version of reduced theory and reducing theory are closed, they must express empirical statements. This correspondence between distinct kinds from different theoretical domains should represent a synthetic identity (a posteriori) statement in the same fashion as the statement "The morning star is the evening star".
(4) The correct version of T_2 should indicate in which way(s) the original theory is wrong and must provide more reliable predictions (Schaffner, 1967, p. 144).
(5) As already indicated before, any pair of theories from set *K* must hold close analogy, similarity, or relevant agreement. Schaffner doesn't provide a formal characterisation of such semantic qualifications. Nevertheless, he indicates some vague characterisations that impose effective constraints on what "closed analogy" suggests: (i) the original reduced theory and the analogue system must share all relevant terms (Schaffner, 2012, p. 546) and (ii) the analogue system must hold, at least, the same empirical adequacy as the original theoretical system (Schaffner, 1993, chapter 5).

In Schaffner's account, there's only one possible candidate for good reductions: synthetic identities associating terms across two different theoretical domains and vocabularies. To the

extent that, in heterogenous reductions, the incommensurability of terms between the two sciences inevitably demands (at least) the correction of the reduced theory, we can't expect without further work to obtain a direct inference of experimental law belonging to the top-level science from the lower level reducing science.[19] Thus, this protocol, i.e., correcting the original theory instead of insisting on deriving the original reduced target theory from the reducing one, is the best positivist reductionist program of reduction: it's the corrected theory that serves as the commensurable link between the original reduced and reducing theories, respectively (Winther, 2009, p. 123). Then we can finally employ the derivability condition, fulfilling all Nagel's requirements.

Klein and Schaffner have very different views on what reduction requires. In one view, Klein underlines the plurality of forms that Nagel's original connectability condition demands as a sufficient requirement: the capacity of a reducing theory for representing the relevant terms from the reduced domain using their own tools. In this regard, the incommensurability problem can be avoided if, and only if, the reducing theory can positively introduce the terms from the reduced theory. By doing so, the reducing domain may explain some facts from the reduced science. Klein, I believe, is more worried about the right exegetical interpretation of Nagel's view. Schaffner, in turn, is more concerned with expanding Nagel's view, identifying some crucial problems that should be properly addressed.

5. Concluding remarks

In this chapter, I present the first theoretical framework capable of enacting reduction in economic science: Nagel's theory. This account suggests that reduction is a typical inter-level explanation across two different theoretical branches where the reducing science explains the experimental laws from a top hierarchical domain of inquiry. The plurality of reduction functions (or bridge principles) admits at least two possible links: essentially a representational one, permitting an inter-level connection focused on epistemic conditions, or another possibility that demands empirical counterparts between the two layers that may allow the inter-level explanation and consequently reduction. Although we shall primarily employ the former representational approach in chapter 3, we have in this chapter briefly explored the latter approach for the purposes of bringing out the contrast.

[19] We shall see that, in the Lucasian model explored in chapter 3, it's possible to obtain a reduction in Nagelian terms even in cases where the reduction functions involve incommensurable terms across micro-and-macroeconomics.

Although Nagel suggests a preference in favour of empirical bridges across reducing and reduced sciences, he acknowledges that these links may acquire various forms, and no exhaustive classification is available (Nagel, 1970, p. 126). This is natural to expect as far as he couches a general account of reduction that, in order to be functional, should adjust itself according to the inter-level connections under assessment (e.g., biology/chemistry or micro/macroeconomics).

As I shall argue in chapter 3, according to the explanatory purposes at stake, the model builder can abstract away some aspects he considers non-relevant for his target. Among these abstractions and idealisations, the very reduction functions could be among them: a bridge between micro-and-macroeconomics could be linked via a conventional theoretical definition. If the economist explains how a macroeconomic phenomenon is eventually possible due to the very microeconomic assumptions and principles that are in place with the auxiliary of a reduction function which is, in that usage, a mere definition, this could be qualified as a positive Nagelian type of reduction.

CHAPTER 2: MECHANISMS AND THE RISE OF A NEW ERA IN THE PHILOSOPHY OF SCIENCE

1. Introduction

As noted in the introduction, this thesis explores two different kinds of reduction that philosophers have offered: Nagelian and constitutive-mechanistic-based reductions. This chapter is about mechanisms as a prelude to discussing the constitutive-mechanistic reduction in the next chapter. Mechanisms have been under much discussion in the philosophy of science lately, especially as means for providing causal explanations. I will explain how two different kinds of mechanisms (or, more accurately, mechanistic models) have been proposed – causal mechanisms and constitutive mechanisms and give a brief account of each. It is essential to keep the difference between them in mind because it is, I shall explain, constitutive mechanisms, not causal mechanisms, that can provide reductions.

Temporal asymmetry is one such feature: it is supposed to hold for causal process mechanistic explanation but not for the constitutive mechanism and not for mechanistic reduction. So, in this chapter, I lay the groundwork for the second contender for reducing macroeconomic phenomena to microeconomics, mechanistic reduction, by providing a general introduction to mechanisms in the contemporary philosophy of science.

Since their inception in the 20th century, mechanistic accounts have become a hot topic in the philosophy of science. Historically, they grew from the ashes of logical positivism. Until the mid-1950s, philosophy of science was heavily influenced by logical positivism, which, at that time, was more like a research program rather than a coherent compilation of doctrines. Scholars like Carnap and Reichenbach thought that philosophical inquiries, and likewise scientific enterprises, should be constrained by the tools of logic and mathematics. They conceived philosophy as mostly a study of the logic of natural sciences (Carnap, 1937). Logical empiricists reasoned that their views might accurately cover every sort of scientific discipline. However, their ambition was found wanting. Many philosophers of science highlighted that the positivist view of science does not bode well for scientific practice. Since the late 1960s, in an attempt to overcome the limitations of logical positivism, an alternative framework was put forth based on mechanisms.

There is no consensual definition of what a mechanism is. Different authors have forged distinct definitions. Unlike “chair”, mechanisms cannot be defined ostensively. They do not have a clear set of defining traits. Instead, they are characterised qualitatively and contextually. Mechanisms, it is claimed, can be a powerful tool for unveiling various relations of dependence

(e.g., causation, constitution), integrating multiple levels of organisation that aims to describe why a target phenomenon was possible, and overcoming the limits of the DN-model of explanation. By "mechanism", I adopt the following definition: "A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations" (Glennan, 2002, p. S344).[20] As such, a mechanism is inherently a causal notion where the mechanist tries to describe how a given target phenomenon, identified as a mechanism, has happened, putting it into the causal structure of the world, articulating the sequence of events that, combined with the underlying structural features of the phenomenon at stake, make possible its manifestation.

Mechanisms seem a good candidate for the role of enabling reduction insofar as they appeal to parts and underlying components of a target system when couching their explanations. To get a mechanistic reduction, one must demonstrate that the phenomenon under analysis depends on its components. If this is the case, then a reductive explanation can be mechanistically possible: it's just a matter to demonstrate that relevant features of the target phenomenon is counterfactually dependent on some of its relevant internal components or structural features (Ylikoski, 2012, sections 2 and 3).

The structure of this chapter is as follows. I first outline two types of explanation that mechanistic accounts typically contain: constitutive and causal explanations (Section 2). Then I identify some general traits that constitutive mechanisms normally have (Section 3). Moving on, I present the account of (causal) explanation hereby explored – James Woodward's counterfactual-interventional-theory of explanation (henceforth, CITE). Other notions of causation have been put forth in mechanistic explanations as well.[21] I have chosen to discuss James Woodward's view because of its consistency with how many economists have envisioned the nature of causality in their field, and it also respects the definition of "mechanism" adopted in this thesis.[22]

Woodward's account is usually included in the class of interventional theories of causal explanation. Although interventions count as a decisive mark for acknowledging causal (and,

[20] Although different mechanistic accounts provide different definitions of the term, those definitions are not all incompatible. I echo Tabery (2004), who suggests that the available definitions of "mechanism" are better construed as complementary rather than irreconcilable competitors. A list containing this diversity of definitions can be founded in Hedstrom and Ylikoski (2010). The list illustrates the consistency between Glennan's definition with other widely popular definitions of the term.

21 Other accounts include conserved-quantity accounts (Salmon, 1984), activity-based accounts (Bogen, 2005; Machamer, 2004), and purely mechanistic ones (Glennan, 1996, 2009).

22 See, for instance, Hoover (1988), Heckman (2005, 2008). In a recent paper, Henschen (2018) argues in favour of Woodward's account for tracking causality in macroeconomics.

according to the case, also applicable for noncausal) relations, in Woodward's theory, interventions do not per se play the explanatory role. It's the counterfactual information that mainly provides explanations for causal *and* noncausal explanations in the heart of the mechanism. That's why, following Saatsi and Pexton (2012), I qualify Woodward's view as a counterfactual-interventionist view (CITE). Other versions of interventionist accounts of causation construe human action as causal as well, but Woodward's theory is not committed to that claim.[23] We can still embrace interventions as the best guide for tracking causal and noncausal dependencies, even in contexts where factual manipulation is virtually impossible, by invoking counterfactual reasoning that channels the explanans to explanandum (Section 4).[24]

Moving on, I describe the role that the microfoundational thesis plays in sociological explanations, especially how it integrates the mechanistic framework, and how this precondition claimed by various methodologists in social sciences may enact a constitutive mechanistic reduction. I argue that it's the microfoundations requirement – the thesis that states that information about individuals' conditions and what leads them to bring about a macro phenomenon are indispensable in sociological explanations (Little, 1998) – that can supply relevant information from lower levels of analysis in causal and constitutive terms capable of enacting a reductive explanation. The thesis can be construed as the attempt to preserve the canons of Methodological Individualism (Janssen, 2008, p. 1).

In the context of a social mechanism, we have a reductive explanation of a macrosocial phenomenon, according to the CITE, if the explanandum placed at the top level is counterfactually dependent on people's decisions. It's the principle of methodological individualism that raises the prospect of constitutive explanation and hence a constitutive reduction. When the most visible property (or properties) of a target macrosocial phenomenon is mainly explained by consequences of individuals' choices under a set of constraints, we may have a reductive explanation.

At the same time, mechanistic-based explanations may subscribe to the explanatory autonomy of higher-level explanations. Reduction consists, in a mechanistic-based explanation, in a stage of the analysis of the entire behaviour of the mechanism, how multiple levels of inquiry are harmonised across multiple layers. So, attaining reduction, in mechanistic terms, does not eliminate the upper level. Mechanistic explanations may acknowledge the explanatory

23 See Menzies and Price (1993).
24 Other interventional writers such as Hausmann (1998) and Pearl (2000) share the core of the message advocated by Woodward. Although they disagree in somewhat in detail, mostly in virtue of these authors have different focus and motivations.

autonomy of events at the macrosocial level (Section 5). If a macroeconomic phenomenon can be reduced to microeconomics mechanistically in an interventionist fashion, it shall be necessary to demonstrate a counterfactual dependence between the chosen target variable the at macro level and the correspondent variables representing individuals' behaviour (beliefs or expectations). All these nuances of mechanistic explanations will appear in the following chapters when I present positive examples of reductive and non-reductive explanations in economics.

2. Causal & constitutive explanations

Generally, a mechanism envelopes two different classes of explanations. Following Salmon (1984), mechanisms provide constitutive and causal (or etiological) explanations. A mechanism can be analysed and described both constitutively and causally. Depending on what the inquirer is investigating, she may concentrate her attention on (mainly) two different aspects of the phenomenon: either (a) she looks at previous processes and interactions that produce it or (b) to their components and structure that might explain its causal features/powers observable at the surface:

> If we want to show why *E* occurred, we fill in the causally relevant processes and interactions that occupy the past light cone of *E*. This is the etiological aspect of our explanation; it exhibits *E* as embedded in its causal nexus. If we want to show why *E* manifests certain characteristics, we place inside the volume occupied by *E* the internal causal mechanisms that accounts for *E*'s nature. This is the constitutive aspect of our explanation; it lays bare the causal structure of *E*. (Salmon, 1984, p. 275)

These two classes of explanations differ over the type of relata that figure in the explanans of a phenomenon (explanandum) *E*. The explanans of constitutive explanations is causal powers (tendencies, dispositions). Suppose someone asks you, "What makes the glass become fragile?".[25] An adequate answer does not describe the situation of any glass in particular. It aims to shed light on the components that make any glass breakable, which features make any glass possess the capacity of being fragile – rather than robust. A proper response should therefore include in explanans information about molecules and bonds that keep glass molecules grouped under ideal conditions and why they get unbounded under a huge impact, for instance, of a hammer. By saying this to our inquirer, we affirm that glasses have a tendency (or disposition) to be broken due to their composition. We offer information about the relevant parts responsible for particular features of the phenomenon at the surface.

25 For more details about this example, see Ylikoski (2013).

Unlike causal explanations, constitutive answers do not claim timing distinctions between explanans and explanandum – the relation between the two is synchronic. When the volume of an ideal gas is altered due to variation in pressure, its molecules must also vary. In the same vein, if the temperature of a system *S* is somehow modified, it must have had an alteration at the underlying molecular level. In the case of temperature, we might assume that its causal dispositions can only be displayed if the causal powers of its components are also active, revealing the asymmetrical relationship between components and the whole in terms of existence: the parts constitute the whole but not the other way round. How many underlying entities should be active in displaying their causal powers depends on their explanatory relevance; the context and the inquirer's demands determine what is relevant for explaining the phenomenon.

Causal explanations, by contrast, describe relations between actual spatio-temporally distinct events. Unlike constitution, causation is asymmetrical under interventions: since the cause is constitutively independent of the effect, in modifying the cause, we do also thereby alter the effect. Although relations of constitution and causation are metaphysically dissimilar, they are pretty similar when we forge constitutive and causal explanations:

> Although metaphysically the relations of constitution and causation are quite different, in terms of explanation the basic principles are quite similar. Both explanations attempt to track networks of counterfactual dependence. A causal explanation tells us how the antecedent event and their organization (timing and location) bring about the event to be explained. In contrast, a constitutive explanation describes how properties of the components and their organization give rise to the system's properties. (Ylikoski, 2012, p. 34)

The means of elucidation of our causal and constitutive queries can be clarified by invoking various counterfactual dependencies between antecedent processes that causally determined *E*, or how the relevant parts and components of *E* somewhat interact that making it exhibit the manifested behaviour. When our queries involve timing distinctions, sequence, and stages of processing of distinct items, the investigation shall focus on sequences of events (explanans) that somehow might causally bring about the phenomenon at issue (the explanandum). On the other hand, when the inquiry asks for the inner engines that could fire the causal powers that figure in the explanandum, one has to look for parts whose absence produces an impact on the behaviour of the whole. Both constitutive and causal answers must withstand counterfactual analysis: if the relevant components of the target phenomenon are inactive, then the explanandum *E* would be different (constitutive explanation). In case of causal answers, if the antecedent event had not been occurred, the explanandum event would have been different.

3. Constitutive mechanisms: an overview

First, a central task of mechanists is to identify the phenomenon as such. Hearts (Bechtel, Abrahamsen, 2005), biological systems (Boogerd, et.al., 2007), brain cells (Craver, 2007), Rube Goldberg apparatus (Woodward, 2013), and electric bells (Shapiro, 2017) can all be viewed as exemplars of mechanisms. We can look at these exemplars in terms of their functionalities: they can produce or modify some target phenomenon chosen by the mechanist. If the phenomenon is due to a mechanism, the inquirer can open its black-box and map out its internal compositions, which, by its functions, contribute to what goes on at the surface. What is observed at the surface is the manifestation of the inner workings of mechanisms.

Second, there are hierarchies of mechanisms. Once the phenomenon and underlying mechanisms are identified, those mechanisms can be assigned to specific levels within the mechanism.[26] We can usually go on breaking down mechanisms into further components. Any phenomenon can, in principle, be explained by a more fundamental phenomenon. In that sense, mechanisms overlap, which means that they span across multiple levels. In other words, mechanisms can be decomposed; a legitimate mechanism can be formed by dozens of sub-mechanisms in their composition. Consequently, the idea of "level" in terms of mechanisms is essentially *perspectival*: a mechanism can simultaneously be placed at higher and lower levels, depending on which sort of item is being contrasted with. For example, microeconomics, in comparison with macroeconomics, represents the lower level. By contrast, when microeconomics is compared with neuroeconomics, microeconomics corresponds to the higher level. In this sense, micro-and-macro or small-and-large is a matter of contrast (Ylikoski, 2012, p. 25).

Third, mechanisms have a structure. Each exemplar has its mode of organisation: parts, entities, properties, and a web of connections. The underlying structure is spatially organised: each part has its place, shape, size, matter, motions, and even the things they are made of are vital for the mechanism's global functionalities. For instance, an electronic bell can be identified as a mechanism: its primary function depends entirely on its structural components to perform

[26] To assess reduction in the context of mechanistic-based explanations, I opt for Carl Craver's view of "levels of mechanisms". According to Craver, every exemplar of a mechanism is an active compositional (part-whole) phenomenon whereby the whole is at the higher level and its structural units belong to the lower level that perform some function, a function that is vital for the entire performance of the mechanism at the higher level. So, the higher-level entity is an active mechanism performing some function, while the lower-level elements are the components that contribute to the behaviour of the mechanism. Craver develops a local approach to levels that is applicable only in the context of case study analysis: "Levels of mechanisms are locally, within the context of a given type of mechanism" (Craver and Bechtel, 2007, p. p. 550). He continues: "How many levels there are, and which levels are included, are questions to be answered on a case-by-case basis by discovering which components at which size scales are explanatorily relevant for a given phenomenon" (Craver, 2007, p. 191).

its proper operation. Its armature should be iron made. An armature made of wood would not attract electromagnets across the circuits of units that compose the device. In the same vein, the armature cannot also be placed at a far distance from the iron core; otherwise, the attraction of the magnet might not be sufficient to pull the armature toward it, compromising the ringing of the bell (Shapiro, 2017). Thus, the disposition of the structure is crucial for the correct operation of the entire mechanism.

Finally, a mechanistic-based explanation describes causal processes of a mechanism selectively. Once mechanists open its black box, they shall have to recompose it, showing how the organized parts orchestrating their operations could generate the phenomenon (Bechtel, 2017). They track the relevant processes, phases, and entities serially. If an entity within the structure of the system is explanatorily irrelevant to the behaviour of the mechanism, it can be disregarded. This does not mean that the entity is irrelevant in all respects. Instead, it means that that entity is not significant according to what we are aiming to explain. Although it figures within the structure of the mechanism, for specific explanatory purposes, its role can be disregarded.

4. Woodward's counterfactual-interventionist account of explanation

In this section, I present Woodward's theory of explanation. In *Making things happen* (2003), Woodward mostly explores examples of explanations that exhibit causal relations under factual interventional scenarios, although he truly believes his theory can cover (i) non-interventional cases and (ii) noncausal examples of scientific explanations. It's just a matter of successfully demonstrating a counterfactual dependence of the explanandum from the explanans:

> […] the common element in many forms of explanation, both causal and noncausal, is that they must answer what-if-things-had-been-different questions. […] When a theory or derivation answers what-if-things-had-been-different question but we cannot interpret this as an answer to a question about what would happen under an intervention, we may have a noncausal explanation of some sort. (Woodward, 2003, p. 221)

In what follows, I start by laying out an example of a factual intervention in a noncausal explanation (Sub-section 4.1).[27] Then I introduce an example from an empirical science where factual interventions are impossible: history (Sub-section 4.2). As I intend to demonstrate, what

[27] More recently, Woodward (2018) explicitly relaxes the interventional criterion, arguing that explanations should still answer what-if questions, but not necessarily under interventions. In so doing, his counterfactual view potentially contemplates different sorts of noncausal examples, showing the crucial point of Woodward's view is the counterfactual aspect.

unifies different exemplars of explanation into a single core and constitutes an imperative condition of causal and noncausal explanations is, in Woodward's view, the counterfactual aspect.[28] These two types of explanation are important to keep in mind once I employ them in the next section (section 5) to assess Weber's explanation for the rise of capitalism and, most importantly, in the next chapter where I apply Woodward's account to evaluate the merit of his proposal to unlock reduction in economics in mechanistic terms.

4.1. Interventional condition in the context of noncausal explanations

Think about what causation is or looks like. We daily experience multiple spatio-temporal contiguous events: (a) a glass that had been suddenly broken after someone had accidentally hit it, or (b) a slice of roast pork after one hour in the oven. Common sense naturally associates causal interpretations for these phenomena: they represent a sequence of events that can be read counterfactually: under different antecedent conditions, the results would have also been different. If we could modify the antecedent conditions, the outcome would have also been modified. Putting it differently, if we could *intervene* upon the antecedent conditions, then we expect to obtain a different result.

Outside philosophy, the idea that causal relationships are potentially exploitable under interventions is extremely popular. Several practitioners, from life sciences to social sciences, endorse that if *X* can correctly be viewed as the actual cause of *Y*, then we can say that *Y* also varies under the direct intervention of *X*. To guarantee that *X* causes *Y*, there must be some values that *X* might assume that also provokes different values of *Y*:

> *The paradigmatic assertion in causal relationships is that manipulation of a cause will result in the manipulation of an effect. [*…] *Causation implies that by varying one factor I can make another vary*. (Cook and Campbell, 1979, p. 36, emphasis in original)

Woodward (2003, p. 98) lists four necessary and sufficient conditions that interventional causal claims must have:

> (I) *I* causes *X*.
> (II) *I* acts as a switch for all the other variables that causes *X*. That is, certain values of *I* are such that when I attains those values, *X* ceases to depend on the values of other variables that cause *X* and instead depends only on the value taken by *I*.
> (III) Any directed path from *I* to *Y* goes through *X*. That is, I does not directly cause Y and is not a cause of any causes of Y that are distinct from X except, of course, for those causes of *Y*, if any, that are built into the *I*-*X*-*Y* connection itself: that is, except for (a) any causes of *Y* that are effects of *X* and (b) any causes of *Y* that are between *I* and *X* and have no effect on *Y* independently of *X*.
> (IV) *I* is (statistically) independent of any variable *Z* that causes *Y* and that is on a directed path that does not go through *X*.

[28] For a more detailed defence in favour of a monist counterfactual account of explanation, see Reutlinger (2016).

Interventions *I* must be the unique causal factor on *X*. All the previous values that *X* might have assumed should be disrupted after the intervention. *I* does not cause *Y* directly; it triggers a circuit of causal dependencies that flows through *X* until *Y*. Using an example, a simple textbook syllogism below illuminates the contrast between Woodward's account and the traditional DN-model of scientific explanation:

(I) All ravens are black.
(II) *a* is raven.
(III) *a* is black.

This is a valid deductive argument consistent with the DN-model: (I) is a generalisation (nomological assumption) whereas (II) is a particular instantiation of (I). The conclusion follows due to the meaning of the terms involved. In this example, the explanandum (III) cannot have been different since it is a logical consequence from prior assumptions. The individual raven had to be black. However, the argument is silent about *why* any raven is black, what makes a raven black. To provide an appropriate why-answer for this question, the explanans should list relevant information that makes a difference to whether a bird is black or not, which, for instance, should include the biological trait that ravens possess such that in its absence, their colour would have been different.

A proper explanation must reveal that if we modify the explanans, the outcome would also be altered. In this case, to what extent can we safely say that modifying the specific biological trait of ravens would therefore modify their colour? Relevant counterfactuals can successfully describe the outcomes of satisfactory interventions (Woodward, 2003, p. 196). Thus, a proper explanation for a raven being black should ensure that if we intervene over their specific biological traits, then we can change their pigmentation:

> […] when *Q* asks for an explanation of why some raven is black, he wishes to know what it is about that raven that makes it black. *Q* is puzzled because he is unable to identify those features of a raven on which its pigmentation depends and is unaware of the laws or generalisations that describe this dependency. When *Q* is in such a situation, he will not be helped by being told that all ravens are black. This generalisation tells *Q* that all other ravens have the feature he finds puzzling about this particular raven, but it does not tell him what the feature depends on, which is what he wishes to know. (Woodward, 2003, p. 203)

In Woodward's approach, his theory can cover type-causal and token-causal claims, once the primary relata for causal explanations are variables that may assume different values in different contexts (Woodward, 2003, p. 122). In this example, a proximal explanandum is the

ravens' *colour* while the explanans is the *biological trait*. If we could map out the right DNA sequence responsible for determining raven's pigments, then we can change their appearances. In this case, the explanans is performed by biological properties, which DNA sequence has the capacity to determine ravens' pigmentation. In this regard, we may say that a most visible trait of ravens is determined by some internal biological component of raven.

In this example, we are talking about the causal powers that a biological property may manifest, its intrinsic capabilities responsible for printing their marks on the aesthetics of ravens. We are not describing a chain of processes responsible for determining the colour of ravens. The relation between biological and aesthetic trait is purely synchronic, aligned with the purposes of constitutive explanations where the relata figuring in the explanans is the causal powers, tendencies, and dispositions of a particular biological trait. But notice this is not a reason for challenging Woodward's view. It reveals instead that, in principle, the interventional element of his theory, initially envisioned to cover causal explanations mostly, can be preserved and extended to constitutive (noncausal) explanations as well.

The raven's case reveals a necessary condition to represent an explanatory relationship between explanans and explanandum: invariance. This notion is closely tied to interventions and counterfactual explanations. It shows that a generalisation, presumedly explanatory, must demonstrate an objective invariant/stable relationship such as if the value of *X* were to change (the DNA's sequence) then the value of *Y* (raven's colour) would change as well. In this example, *X* is the supposed putative factor that prints its features into the aesthetic of ravens. So, invariance is evaluated by directly considering the stability between *X* and *Y* – new values in the explanans variable *X* (x_1, x_2, … x_n) must provoke novel values in the explanandum variable *Y* (y_1, y_2, … y_n).

If scientists have identified and individuated the right DNA sequence responsible for printing the black colour of ravens, then a modification in the right DNA must alter a raven's pigmentation, otherwise we may conclude that scientists did not discover the right sequence of DNA responsible for ravens' colour. Invariance ensures three crucial marks of objectiveness required for any scientific inquiry: "accessibility from different angles, intersubjective agreement, and independence from the observer" (Nozick, 2001, p. 85). Very often, the stability (or invariance) requirement involves the putative causal factor and the target variable. But in others, the invariance is observed – and mostly claimed – in the background conditions (Woodward, 2018, p. 121), a topic I address in the next chapter when I discuss the Phillips curve.

4.2. A (counterfactual) causal explanation in a context of non-factual intervention

The counterfactual rationale is an appealing way of thinking. We often query ourselves wondering how our lives would have been if different attitudes had been previously taken. For example, if Archduke Frank Ferdinand had not been assassinated in June 1914 in Sarajevo, i.e., if different events had happened, would World War I have been sparked just one month later?[29] In the philosophical literature, counterfactuals were born as a variation of the regularity view of causation from the words of David Hume (1777 [1902], section 7; emphasis original):

> *we may define a cause to be an object followed by another, and were all the objects, similar to the first, are followed by objects similar to the second. Or, in the other words, where, if the first object had not been, the second never had existed.*

Imagine that during a pool game, I have hit the red ball into the corner pocket. You agree that the red ball has entered the corner pocket due to its last push, right? There is a sequence of events where a type-*B* event is a consequence of its preceding event of type-*A*. As the player, I directly observe the occurrence of a particular event *a* of type-*A* (cause) followed by an event *b* of type *B* (effect), allowing me to infer that *a* (the hitting) has caused *b* (the rolling ball). Causation, it seems, involves the instantiation of regular conjunction of actual events. To affirm that some event explains the occurrence of another event, three conditions must be present, according to Hume:

> *c* is spatiotemporally contiguous to *e*,
> *e* succeeds c in time, and
> All events of type *C* (i.e., events that are like *c*) are regularly followed by (or are constantly with) events of type *E* (i.e., events like *c*). (Psillos, 2012, p. 112)

Hume originally thought that counterfactual reasoning was a different way to construe causal regularities. The reason is pretty tempting. If *X* is the cause of *Y*, then *Y* would not have happened in the absence of *X*. However, regularity accounts aim to explain causation not in terms of the events themselves but in terms of an underlying regularity of nature conjoining events of similar types. The attempt to explain causation in terms of the events themselves is the target of counterfactual accounts of explanation.

The analysis of causation in counterfactual terms became popular after the 1970s, mainly due to David Lewis's contributions. Although forged in philosophy to dispute the nature of causation in metaphysics, counterfactuals were later adopted by plenty of disciplines (e.g., criminology, history, economics) to develop their causal explanations. For Lewis,

> […] we think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its

29 That imaginary possible world is the subject of Richard Lebow's *Archduke Franz Ferdinand lives*! (2014). He claims that the cause for WWI was ignited by the murder of Franz Ferdinand. Without this event, Lebow argues for, WWI would have not been fired at that time.

> effects – some of them, at least, and usually all – would have been absent as well. (Lewis, 1973, p. 161)

In Lewis' account, there are two sufficient conditions for causation: (I) the cause and the effect are distinct and actual events, and (II) non-C entails non-E ($\backsim C \; \square\!\!\rightarrow \backsim E$). By its nature, counterfactuals are contrary to facts, and thus they cannot be directly verified. That is why Lewis invokes the possible-world semantics to shed light on counterfactuals. Scholars and practitioners who followed Lewisian footsteps do not appeal to metaphysical resources to assess causal claims (Reiss, 2012, p. 160). Instead, they develop specific desiderata to track causal relations but still preserve the fundamental message of counterfactuals:

> Here is my proposal: explanation is a matter of exhibiting systematic patterns of counterfactual dependence. Not only can the generalisations cited in (5.1.3) – (5.1.4) be used to show that the explananda of (5.1.3) – (5.1.4) were to be expected, given the initial and boundary conditions that actually obtained but they also can be used to show how these explananda would *change* if these initial and boundary conditions had changed in various ways. (Woodward, 2003, p. 191)

The connection between the explanatory target (explanandum Y) and the invoking features (explanans X) that may explain the outcome demands that the inquirer demonstrates a difference-making dependency from X towards Y. Counterfactuals preserve asymmetry: they inform us that the explanans makes a difference to the explanandum, but they don't suppose that changing the explanandum would thereby change the explanans. It is the counterfactual dependence that assures the explanans is causally relevant to the explanandum.

Consider a case where actual intervention is impossible: Weber's explanation for the origins of capitalism. Weber was not the only scholar who hypothesised that the Protestant religion exerted a decisive role in the rise of capitalism. There was a common acceptance among scholars of that period that both phenomena were closely correlated. Countries where capitalism had consolidated its supremacy were those in which the Protestant doctrine was heavily influential. There was a strong association between the new cultural environment with an emergence of a novel economic system. Nonetheless, correlation is not the same thing as causation.

To evaluate whether Protestantism can be seen as the legitimate cause for the rise of capitalism, we should modify the antecedent conditions and ask ourselves whether the outcome would remain unmodified. Surely, we cannot require a factual intervention of the initial conditions for this sort of query. In such contexts, interventions are purely hypothetical. But we can envision hypothetical ideal tests. Let's consider that the antecedent conditions for the consequent outcome Y may assume two values, (1) or (0):

(1) = [*under the influence of protestant doctrine*]

(0) = [*without the influence of protestant doctrine*]

To further "test" whether (1) can be construed as the causal event of *Y*, we should turn (1) silent by replacing it with (0) and ask ourselves whether *Y* would have happened under those different antecedent conditions. To evaluate whether a macrosocial event has causally influenced another, proper intervention is impossible to investigate causal queries like this, but this form of analysis is heuristically valuable (Woodward, 2003, p. 11). For this case in particular, the causal explanation must answer what-if-things-had-been-different questions such as: (i) Would capitalism have emerged in the USA in the absence of the Ascetic Protestant cultural environment? Or (ii) would capitalism have surged in Spain under the cultural presence of Protestantism?

Causal answers, Woodward argues, are explanatory in virtue of exhibiting how the explanandum counterfactually depends on the explanans. In cases where manipulation is factually possible, interventions are more than welcome. But when it does not, we should still be looking for what-if answers through hypothetical reasoning that exhibit counterfactual dependence between explanans and explanandum.

In this section, I briefly laid out one example where Woodward's account can be implemented in (fictional) theoretical models of explanations in social sciences. Historians and social scientists can (theoretically, not factually) manipulate the supposed putative causal factor in some ways, which allow them to raise many what-if-things-had-been-different queries and (perhaps) challenge Weber explanation of which antecedent condition was the crucial mark for the rise of capitalism.[30] Nonetheless, to comprehend the entire story of capitalism, mechanist scholars must open its black-box, uncovering how the crucial micro level components of this system (individuals) interacted in order to generate that novel type of macrosocial economic system. In the next sections – particularly in sub-sections 5.2 and 5.3 – I revisit Weber's hypothesis, showing how the same criterion used to evaluate causal explanations (CITE) are applied for constitutive explanations. In this example, it's the constitutive explanation that provides the reductive explanation of the explanandum.

[30] In chapter 5, I re-examine this example under a different framework, the contrastive account of explanation. Unlike this section where my interest is focused on reduction, therein I move my attention to explanations only. It shall be clear that when we subject Weber's hypothesis to the contrastive view, the contrast syntax serves to test different aspects of Weber's hypothesis.

5. The quest for microfoundations in social sciences

Since the birth of sociology as an autonomous discipline, a methodological contention has divided sociologists. Two of the founding fathers of sociology, Max Weber and Emile Durkheim, had distinct views about which social entities, institutions, and features hold causal powers and can figure in the explanans of sociological explanations. Weber held that social facts, the object of inquiry of sociology, result from human actions, i.e., observable attitudes of individuals together with their relations with others. Good sociological explanations should refer to the individuals' actions and interactions; facts encompassing institutions and other social phenomena are consequences of individual people's attitudes. That characterisation is what has been defined as Methodological individualism: only information about individuals' properties (their actions, intentions, purposes, and the like) can figure in the explanans of social scientific explanations (hereafter "MI").

Durkheim, in turn, contends that social scientists must track down prior social features exhibiting relevant aspects that can display social outcomes: the "[⋯] determining cause of a social fact should be sought among social facts preceding it and not among the states of the individual consciousness." (Durkheim, 1938 [1895], p. 110). Norms, institutions, and other social arrangements may also determine social facts. This is called Methodological holism ("MH"): the idea that large-scale social phenomena, not just features of individuals, may have their causal powers, and may thus affect the very constituents that make them up. For methodological holists, in opposition to MI, information about social groups, states, institutions and other social units can perfectly figure in the explanans of social scientific explanations. For those who endorse MH, or more neutrally *non-reductionism*, there are disputes whether the microfoundations requirement is indispensable:

> [...] the microfoundations thesis holds that an assertion of an explanatory relationship at the social level (causal, functional, structural) must be supplemented by two things: knowledge of what it is about the local circumstances of the typical individual that leads him or her to act in such a way as to bring about this relationship and knowledge of the aggregative processes that lead from individual actions of that sort to an explanatory social relationship of this sort. (Little, 1991, p. 196)

In the sections below, I demonstrate how the microfoundations thesis and mechanisms relate. First and foremost, I lay out an overview of MI plus the fundamental insight of MH (subsection 5.1). Then I introduce James Coleman's (1990) diagram, which illustrates a mechanistic-based explanation consistent with the microfoundations (Sub-section 5.2). Finally,

I demonstrate how this method reconciles the explanatory autonomy of upper-level facts with reductive explanations (Sub-section 5.3).[31]

5.1. Methodological individualism

Historically, the genesis of MI can be traced back to the 16th century – although it has even older ancestors across the centuries.[32] Thomas Hobbes argued that in their original positions – in their "state of nature", individuals live under hostile conditions: a permanent sense of alertness, insecurity, a long-term fight for their survival. As rational beings, people have strong incentives to look for alternative methods to overcome their difficulties. To protect themselves against eventual menaces, they ended up introducing prudential measures to minimise the effects of likely harmful situations. In their pursuit of peace and calmness, people created rules, legal institutions, and constitute someone as their sovereign to encourage the entire community to play by the rules.

The birth of national states, Hobbes argued, resulted from cooperative behaviour undertaken among citizens, assuming (ideally) that all prior procedures for establishing that general agreement were fair and impartial (Cudd and Efthekari, 2017, section 1). In the Hobbesian view, society is a great machinery built by individuals. Its existence is *derivative*. That is why Steve Lukes (1968, p. 119) associates Hobbes with MI: he aimed to explain the emergence of society via its fundamental units (people). Following this path of reasoning, the man who introduced the label of "MI", Joseph Schumpeter, followed a similar line of thought.[33]

To Schumpeter, MI was a line of demarcation of pure economics as a self-contained science (Hodgson, 2007). Economists, he argues, must depart from the level of individuals to elucidate economic regularities and phenomena so as "[…] to describe certain economic relationships" (Schumpeter, 1908, p. 91). Schumpeter takes for granted the utilitarian assumption that (i) individuals pursue wealth and aim to diminish their pains and that (ii) only people have needs and desires. Stanley Jevons, one of the founding fathers of the marginalist revolution in economics, defines economics as the science which is "[…] entirely based on a calculus of pleasure and pain; and the object of economics is to maximize happiness by purchasing pleasure, as it were, at the lowest cost of pain" (Jevons, 1911, p. 23).

---

[31] The explanatory autonomy of higher-level events will be revisited in chapter 4, when I present a non-reductive reading of how micro-and-macroeconomics are possibly channelled.
[32] See, Udehn (2002).
33 The expression "Methodologischer Individualismus" was firstly introduced in 1908's article "*Das Wesen und der Hauptinhalt der theoretischen Nationalökonomie*", wrote and published by Schumpeter. The English translation was released one year after.

Economists do not draw utility curves to entities other than individuals. People value how many resources are daily needed, taking their disposable income into account. Assuming that prices of goods are determined via market interactions between people's needs and disposable goods, we could predict potential large-scale regularities. From the point of view of pure economic science, "[…] it is irrelevant why people demand certain goods*:* the only important thing is that all things are demanded, produced, and paid for because individuals want them" (Schumpeter, 1909, p. 216). The task of pure economics is to understand the *consequences* of individuals' preferences, not to respond to *why* or *how* they form them. Understanding what guides individuals' economic behaviour is relevant, although it does not make up the object of inquiry of pure economic theory. This is the task to be undertaken by other social sciences like psychology or sociology. Pure economic theory must confine its explanations and predictions assuming the utility theory as the golden landmark of analysis:

> At the outset it is to emphasise the individualistic character of the methods of pure theory. Almost every modern writer starts with wants and their satisfaction and takes utility more or less exclusively as the basis of his analysis. Without expressing any opinion about this *modus procedendi*, I wish to point out that […] it unavoidably implies considering individuals as independent units or agencies. For only individuals can feel wants. (Schumpeter, 1909, p. 214).

Accordingly, the products of knowledge in economics shall inevitably be derivative. Individuals in the pursuit of their own goals generate multiple economic phenomena like supply/demand curves, Engel's curve,[34] and the like. Pure economic knowledge must be behaviour driven; reference to individuals' economic attitudes is indispensable in economic analysis. As the science of scarcity and human decision-making, subverting this order – from individual to economic regularities – would lead us to humanise entities devoid of consciousness, ascribing intentional behaviour to other units but individuals:

> It now becomes clear that the same method cannot be directly applied to society as a whole. Society as such, having no brains or nerves in a physical sense, cannot feel wants and has not, therefore, utility curves like those of individuals. (Schumpeter, 1909, p. 215)

Some of Schumpeter's message was absorbed by Weber, but not all since Weber sees sociology as studying the motives of individuals' actions. To Weber, the object of inquiry of sociology is *social* action. Actions presuppose teleological dispositions (intentions and motives). Sociologically meaningful actions are those carried out by individuals *taking others into account*. This emphasis is significant to distinguish social actions from mere behavioural reflexive responses. If someone says something that embarrasses you, you could blush. But this is a simple physiological reaction of your body with no *intentional* content before an external

[34] Engel curves, named in honour of German statistician Ernst Engel (1821-1896), describe how a household's expenditure on a specific good or service varies with income – the more they earn, the more they consume that good.

stimulus. Physiological outputs that individuals can potentially trigger under embarrassing circumstances can be subject to an empiricist approach. We can replicate ideal scenarios where other individuals are exposed to the same shameful conditions and assess how often people blush under the same conditions, tracking regularities (and even reducing observations to psychological laws).[35]

Social action, in turn, requires intentionality, the subjective motives of *why* an individual does something towards others. When, for instance, Karl avoids colliding against a pedestrian on a pathway, he takes the pedestrian's trajectory into account. He generates a social action without being aware of it. He saw someone coming in his direction and took a different route, avoiding collision. A social scientist will describe how Karl aimed to deviate from somebody based on what he could directly observe plus an ideal supposition assumed by the social scientist: this is what a reasonable person would have done in that context. Or consider an environmental offence.

Mr Jones, a fictional car factory's owner, has deliberately disposed of oil on the River Wear. According to criminal science, knowing why Mr Jones has done what he supposedly did is a necessary condition for anybody being prosecuted and charged as guilty. Bringing to light the criminal's circumstances and Jones' intentions are crucial elements to assess his culpability. Consider that, during the investigation, prosecutors hinted that Jones's decision to dispose of oil on the rivers' surface was purely economic. He would have had to spend £100,000 pounds with adequate means of disposal (packing, transport, recycling, etc.). Alternatively, in the case of discarding residuals in the river's waters, he knew that, according to environmental legal guidance, he would be charged at a maximum a £50,000 fee. Prosecutors conclude that financial reasons likely motivated Mr Jones's behaviour. Weber would probably have followed the investigators. As a rational economic agent, Jones tried to diminish the factory's losses, despite the legal implications and the indirect damages imposed to dozens of people that might be affected.[36]

In both stories, the rational actions ascribed to Karl and Mr Jones are ideal types: instruments to aid social scientists in approaching social phenomena. The assumption in these two accounts, that people maximize their welfare in the face of the predicted actions of others, does not claim a full correspondence to reality. It is just a tool to organise better the inquiry of certain aspects of the social realm. Assuming that (rational) people (ideally) optimise their

35 In his best-seller *Thinking*, *fast and slow*, Daniel Kahneman (2011) assembles various examples of unintentional psychological outputs of that character.
36 Gary Becker (1968) became famous for, beyond many other contributions, introducing rational decision-making and microeconomics' framework to analyse the reasons behind criminal offences.

decisions is helpful. We ascribe to our typical individual some basic attributes that may represent their reasonings. And, in a further step, the social scientist tries to interpret the observable fact and speculate, taking the ideal type of rationality into account, what might explain the intentions behind Karl and Jones's actions, contrasting the ideal type with the observable actions. To fully uncover subjective motives of social actions, two steps are required:

> It depends on the separation of two levels, the practice description and attribution of motives, which is done largely by the application of "rational" ideal type, and the level of explanation and epistemology, at which every interpretation is credentialed as "causally adequate" and every descriptive terminology is presumed equally valid, because each is equally "ideal-typical". (Turner, 1983, p. 514)

As soon as we uncover people's intentions towards others, we may understand the reasons for their social attitudes, and understand the reasons is to understand *causes*. Causality and intentionality walk altogether. According to methodological individualists, this is the task of the social scientist: to unveil subjective intentions behind individuals' actions. Social groups or institutions, like armies, countries, or firms adopt specific courses of action due to their fundamental components: people. Otherwise, we would have been deputising to social institutions, devoid of consciousness, teleological agency as much as individuals. On the contrary, it is people with their actions *and relations* who supply reality with its content and format:

> Collectivities must be treated as *solely* the resultants and modes of organisation of the particular acts of individual persons, since these alone can be treated as agents in a course of subjectively understandable action. Weber, 1978 [1922], p. 13, emphasis in the original)

To better elucidate any macro-social phenomena, a careful understanding of people's motivations brings information that helps our interpretation of these multiple macro social phenomena. To Weber, reference to individuals' actions is *indispensable* for sociological explanations. Aggregate descriptions are only to be accepted as explanatory if they have been decomposed into discernible patterns of individual action (Turner, 1983, p. 518). Hence, social macro facts are not explicable on their own – they are not causally *autonomous*: any sociological explanation must be derived from individuals.

Despite their different goals, targets, and particularities, Hobbes, Schumpeter, and Weber share a common insight. It does not matter which subject you are investigating. A proper method of inquiry must start with solid foundations, from more specific pieces of knowledge involving basic common-ground assumptions that we can trust to less-specific knowledge – an insight that recalls Descartes, at least in spirit. MI is an orientation of how social scientists should address their object of inquiry. As Udehn (2002, p. 499) illustrates below, several

variations of MI were couched so far, but they are generally constrained by and should pay attention to the following questions (Table 1):

Table 1: Different dimensions of methodological individualism

	(a) Definition of social concepts	(b) Explanation of social phenomena	(c) Reduction of social laws
(1) Methodology	Social concepts should be defined in terms of individuals, their physical and psychological states, actions and interactions, social situation, and physical environment.	Social phenomena should be explained in terms of individuals, their physical and psychological states, actions and interactions, social situation, and physical environment.	Social laws should be reduced to laws about individuals, their physical and psychological states, actions and interactions, social situation, and physical environment.
(2) Epistemology	Social concepts can in principle be defined in terms of individuals, etc.	Social phenomena can in principle be explained in terms of individuals, etc.	Social laws can in principle be reduced to laws about individuals, etc.
(3) Ontology	Social phenomena are made of individuals, etc.	Social Phenomena are caused by individuals.	

Source: Udehn, (2002)

In terms of explanation, John Watkins, a prominent MI defender, suggests that the pathway to explain social macro facts could be done via deductions. Following the DN-model of explanation, he argues that social outcomes must be logically deduced from individuals' conditions plus general principles that govern their decisions: "[Methodological individualism] states that social processes and events should be explained by being deduced from (a) principles governing the behaviour of individuals and (b) descriptions of their situations" (Watkins, 1953, p. 729). The trajectory towards the explanandum would make the explanans combine at least two sorts of information: (a) actual circumstances individuals are surrounded by and (b) the laws of individuals' behaviour. These will be the *necessary and sufficient* conditions for providing sociological explanations; information over institutional facts shall not figure in explanans.

The standard alternative to these various versions of MI is various versions of methodological holism (MH). Briefly, methodological holists claim that sociological explanations are about facts encompassing institutions, religions, culture, and other large-scale social phenomena.[37] Consequently, for a better account of their object of analysis, sociology should muster information about institutions – a sort of information compiled from a higher level of social inquiry (institutions). Information about institutions would be indispensable at

[37] For an overview, see Zahle (2016).

the heart of several types of sociological explanations, which makes impossible, or at best unintelligible, to decompose their explanans to the states of minds of individuals.

Consider this breaking news: "China suspends imports from two additional meat plants amid coronavirus fears".[38] That fact does not attribute to any economic sector or group of individuals the decision to suspend meat supply from the Brazilian food industry. If the Chinese government can monocratically block importations based on health issues despite consumers' desires, then the Chinese State is causally responsible for that decision. It is a decision that did not directly emerge from the states of minds of those who were directly affected by that determination. Nor is it a decision of any single individual. Perhaps you can think of it as a decision made by, for example, each of a group of ruling individuals saying "yes". But the methodological holist would add that the decision only has effects on individuals' work.[39]

A collective entity (e.g., nation, state, social institution) can figure as the explanans for the occurrence of the explanandum as far as it can effectively determine social outputs.[40] In some cases, purely holistic explanations may stand on their own. However, most methodological holists supplement sociological explanations with information from individuals and many consider the knowledge that stems from individuals as indispensable. Ultimately, the methodological dispute between individualists and holists stands on the sort of information that must be included in sociological explanations. A conviction among individualists is that to be explanatory, sociological explanations must encompass information about people's conditions and intentions. This is what the microfoundations thesis endorses.

The thesis is an explanatory claim associated with social theories that attempt to provide individualistic foundations of various social facts and phenomena – also known as explanatory individualism (Epstein, 2009, p. 188). It doesn't necessarily commit to the ontological idea claimed by Watkins that social facts/properties just are compositions of individual-level events/properties (Sugden, 2016, p. 1379). As I intend to demonstrate in these next sub-sections, when we adopt mechanistic-based explanations, MI and MH can live in peace altogether – we can be individualists in some respects and non-reductionists in others (List and Spiekerman, 2013, p. 631-632).

By integrating different explanations to understand different dimensions of the same phenomenon, mechanistic explanation can accommodate reductive explanation with the explanatory autonomy of the higher level. We may have a reductive explanation of a social

[38] See: https://www.thepigsite.com/news/2020/07/china-suspends-imports-from-two-additional-meat-plants-in-brazil-amid-coronavirus-fears.
[39] You will see this kind of view argued by Galbács (2020) in chapter 4. Look there for more details.
[40] In chapter 4 I offer a concrete example where a social institution, placed at a higher level of organisation, is the putative causal factor (e.g., it figures in the explanans) of an economic effect.

property *S* if this property is exhaustively realized by the properties of individuals – what List and Spiekermann (ibid., p. 634) qualify as the *type individualism* thesis. Conversely, when we look at capitalism more globally as a historical process (as an event), when historians are willing to unveil the entire collection of previous events that contribute to the birth of capitalism, especially those elements placed at the macrosocial level, then we surely adopt a non-reductionist point of view. In the next section, I demonstrate how this is possible.[41]

5.2. Mixing the micro and the macro in mechanistic-based explanations

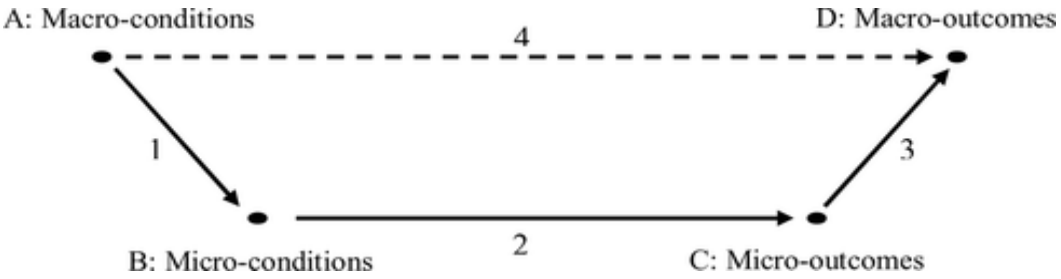
Consider, once more, Weber's explanation for the rise of capitalism. In *The Protestant ethic and the spirit of capitalism* (1976 [1904]), Weber argues that capitalism consolidated its supremacy in nations where citizens had widely assimilated the Protestant doctrine.[42] He suggests a strict correlation between two large-scale social phenomena where the former (the Protestant culture) decisively contributes to triggering the latter (the birth of capitalism). The *ethos* of capitalism with its logic – rationalization in production, economic efficiency, the continuum accumulation of wealth for its own sake – could only develop in a cultural environment dominated by the ascetic Protestantism rationale ethics (Weber, 1976, p. 27).

In that environment, the entrepreneurship virus was inoculated into people's veins, moulding their minds, displaying entrepreneurial behaviours. Weber attempts to provide a causal explanation for the rise of a large-scale phenomenon by reference to another macro-social phenomenon. But if, as Weber argued, macro social facts are consequences of human intentions/actions, how should we articulate two different scales of analysis – individual and the social – into a common core?

James Coleman (1987) criticises Weber's explanation for its failure to successfully demonstrate the web of connections between the individual and the social transparently. Even though we may agree with Weber that the Protestant doctrine was a *sine qua non* condition for the birth of capitalism, his arguments could not satisfactorily demonstrate the sequence of events from macrosocial events to individuals, and then to macro-outcomes. Coleman forges a scheme that, he believes, can better accommodate Weber's goals:

[41] A non-reductive approach is the entire point of defenders of global supervenience as the anti-reductionist antidote to bridge micro-and-macroeconomics. In chapter 4, I address this topic in detail.
[42] Tawney (2015, p. 113) stresses that Calvinism was, like the socialist movement in the early 20th century, an urban movement which helps to explain its wide range of influence in British society.

Figure 1: The Coleman's boat



Source: Coleman (1990)

In the diagram, macro-conditions *A* represent the Protestant religion, its cultural shock plus other concrete conditions from the environment (e.g., institutional facts, scientific achievements of that age that enabled more efficient modes of production, and prices of goods that might have awakened people's sense of opportunity). Arrow 1 depicts the influence of those macro-conditions upon individuals. So, the psychological states of individuals are explained in terms of the social structure – there's a downward causation step: there's a difference-making of people's beliefs explained in terms of the social setting conditions (*A* and *B* are counterfactually dependent).

Arrow 2 depicts a transformational step happening inside the micro base (the level of individuals), from people's beliefs (mental properties) into actions (behavioural properties), where the behaviour is explained in terms of mental states. Then we have node *C*, a second transformational step from micro-level behavioural properties of individuals into the macrosocial outcome – capitalism. Arrow 3 depicts this latest transformational phase where individuals produce unintentional social outcomes through their actions. And finally, arrow 4 portrays an indirect (causal) link between macro-conditions and macro-outcomes.

According to CITE, *A* (causally) explains *D* so long as in modifying *A*, *D* would have had a different result. When the explanans and explanandum exhibit robust counterfactual dependence that remains invariant under (ideal or factual) alterations, then we may say both events are causally channelled. *A*, a macrosocial event, constitutes a decisive condition for the occurrence of the macrosocial event *D*. In this regard, macro social events can be included in the explanans of sociological explanations. *A* explaining *D* is an example of a higher-level explanation in the social sciences: the explanandum, a macrosocial phenomenon, is an event explained in terms of an actual and distinct previous macrosocial event (the Ascetic culture plays the explanans role). But in order to understand how *D* came into being completely demands information from the bottom micro line. Now, we are interested in how the

components of the capitalist system have engendered it – a sort of constitutive knowledge about the target of interest.

Unlike the pair (*A*-*D*), the relation between nodes *C* and *D* does not exhibit temporal asymmetry. While arrow 4 relates two macrosocial events, arrow 3 relates properties from different levels of organisation – the individual and the social. When we look at the pair (*C*-*D*) only, we are searching for constitutive relations, how the properties of the system (node *D*) are dependent on the properties of individuals.

For this case, we may say that capitalism (node *D*) is a social system that has the property *S* of *being capitalist*. Among its general characteristics, capitalism mainly consists of a mode of production oriented for profit organised around independent, privately owned, labour-hiring firms (Little, 1991, p. 44). A country is capitalist when its mode of production of wealth is fundamentally grounded on the widespread private ownership of the means of production, and people sell their labour force to get salaries in compensation. In this regard, capitalists and the labour force are the fundamental units of the capitalist system.

I earlier argued that constitutive explanations relate the properties of the whole to the properties of its components. Constitutive answers must exhibit that the system *S* has a given property due to its structural features. Thus, following the CITE, we must show that the realisation of the social property-*S being capitalist* is counterfactually dependent on two selecting individual properties *I*: if properties *I* were absent or inactive, the whole – the social property *S*, would have been different. For instance, we may take, as Marx did, the private means of production in the hands of capitalists as a crucial mark of the capitalist system. If this condition was not satisfied, for example if the state/public sector controls or monopolises the means of production, then another sort of social system would have been born in England in the 16th century. So, the two lower-level individual properties of the capitalist system – (1) private means of production and (2) labour force – are the necessary (and arguably sufficient) properties of the system. Thus, the social property-*S being capitalist* is counterfactually dependent on (1) and (2).

So far, this classification does not inform what are the causal powers of (1) and (2) as constitutive explanations typically require. Unless we specify exactly what (1) and (2) do, and in which ways these two lower-level properties altogether realize the social property *S*, a constitutive explanation and a consequently reductive explanation of *S* remains superficial. One way to satisfy Weber's ambitions and follow his footsteps is to acknowledge that these lower-level properties amount to social groups – capitalists and working-class – and not to individuals directly. At this point, I propose to cut off the historical ontological disputes over the causal

powers of social groups, the nature of social classes, how could we possibly decompose these wholes into their components, and so on.

Following Weber, and in attempting to provide a solid constitutive explanation of the property *S* consistent with MI's principle, I shall functionalize (1) – (2), rewriting them in terms of ideal types. On one hand, there is a property (1) *being a capitalist*, whose causal features are related to what a typical individual capitalist (e.g., a fabric owner) in the Britain of the 1700s looked like (how they normally behaved, what were their motivations, and so on). On the other hand, there is the property (2) *being a labourer* – which refers to a typical labourer/household of that age. This idealisation has one main role: it functions as a benchmark, i.e., something that is not directly used as a description of anything real but may serve to isolate a causal mechanism or a single dependence relation (Maki, 2020, p. 219-220).

In the Weberian mechanistic explanation, the notion of rationality is crucial for understanding the behaviour of individuals, as we shall now see. In his typology of rationality, Weber distinguishes four types of rationality: (a) formal, (b) theoretical, (c) substantive, and (d) practical rationality. For the rise of capitalism, Weber believes, (c) and (d) were present in the actions of individuals (capitalists and households). Substantive rationality manifests when the values of individuals (e.g., morality, religious beliefs, family values) orient their daily lives, particularly how they choose the proper means to a given end. For example, I always make a call to my old friends on their birthdays. No matter where I am and what my circumstances might be, I never forget to make the calls. Friendship, as I see it, demands loyalty, mutual assistance, and compassion. So, when I call some of my old friends, these values orient my action.

Practical rationality, in turn, manifests when individuals regard only their egoic interests as the end and try to employ the most suitable means to achieve their goals like a calculus, comparing the pros and cons. Unlike substantive rationality, practical rationality does not involve any values impregnating the action. This type of rationality exists as a universal capability of *Homo Sapiens*, a type of rationality that transcends any societal and cultural arrangements.

Suppose someone asks God (or Saints) for a blessing. In return, she/he promises to behave accordingly or do something in exchange. This exchange relation between the believer and the holy, the sacrifice and the prayer, is similar in relevant respects to the case of a firm deciding which is the best means to allocate its resources to maximise its profits (Weber, 1968, p. 424). Both, the believer, and the firm, are looking after their returns in the future.

According to Weber, these two sorts of rationalities, the substantive and the practical, were heavily present in the actions of individuals who engendered the capitalist Britain as such.

First, the cluster of Calvinist values provided a sense of direction, a guide of life that puts individuals (especially capitalists) into action. The doctrine gave a new mindset for those who became capitalists, where the new forms of production sounded like the most efficient way to achieve the end – the creation and accumulation of wealth for its own sake. Substantive rationality is intrinsically grounded in this process, so long as it involves the choices of means towards the ends guided by some system of human values such as Calvinism. In Weber's view, the Calvinist principles pushed people towards a set of actions.

Labour men, the working class, in their turn, considered that selling their labour force was one of the most suitable actions to be done under those conditions. Equipped with the most efficient means of production, products like cotton became gradually manufactured by the cotton industry, destroying the local businesses that could not compete against the efficiency of the machines. So, while the actions of capitalists are mainly guided by substantive rationality, households were mostly *reacting* to that novel pattern of production, adjusting themselves and finding the best suitable mean towards their goals. With all this in mind, let's move back to Coleman's bathtub.

In Coleman's diagram, node *C* represents the properties at the level of individuals, the positions each of the individual units occupy in the capitalist system – the typical capitalist as a member of those who possessed the means of production and the labourer/household as a member of those who sells their labour. When we add the concrete macroconditions *A*, plus the rationality assumptions (node *B*), we may finally make sense of what might be a good explanation for the birth of capitalism in Great Britain. The rational assumptions – substantive and practical – provide a massive plausibility for the occurrence of the explanandum *D*. In that context, under the heavy influence of Protestantism, new scientific achievements, what is the most likely explanation for that new form of sociability?

Freed from the straitjacket of Catholicism and guided by a novel ethos that does not condemn the pursuit of wealth, entrepreneurship behaviour was the most likely rational decision to be made under that context, producing unintended social facts on the societal level. Local suppliers, in return, could no longer compete against the machinery of capitalists for the same markets. In this new setting, selling their labour force seemed the most suitable way to pursue their objectives.

CITE framework is applied to assess the plausibility of alternative imaginary scenarios, allowing us to evaluate objective relations of dependence across different levels of inquiry. In this case in particular, the interesting question for reductionists is – would the social property-*S being capitalist* exist in the absence of lower-level properties *being a capitalist* and *being a*

labourer? An interesting reductive explanation puts in check the plausibility of Weber's rational types. If we cannot change the explanans – *being a capitalist* plus *being a labourer* – without modifying the explanandum (*being capitalist*), then the top-level property is arguably counterfactually dependent on its lower-level properties.

The rise of capitalism is counterfactually dependent upon its fundamental units and partially explained due to the rational principles: we cannot make sense, Weber argues, of the explanandum without conceding the plausibility of these two (substantive and practical) types of rationality manifested in the behaviours of typical individuals (capitalists and households). If Weber's explanation sounds the most plausible reason for the rise of capitalism, then the machinery of mechanistic-based explanations gives to him a great tool to illustrate satisfactorily his original ambition: how one may follow MI's principle and acknowledge the causal efficacy of macro social events simultaneously. Both individuals and macro social events can equally perform the explanans role in sociological explanations. Reduction is just a stage in the entire investigation of what is going on in this social mechanism. When we successfully demonstrate a counterfactual dependence of the selected macrosocial fact/property upon individuals' properties, we have a reductive explanation. It's clear that, as it stands, reduction does not cover the entire story of this social mechanism; other preceding macrosocial events were also crucial for the rise of the mechanism. In this regard, we may say reduction, sometimes, provides partial explanations only,[43] while other cases, as we will see in the next chapter, it provides complete explanation – as the Lucas's model reveals.

5.3. Making sense of autonomy and reduction in mechanistic explanations

So far, I have argued that constitutive mechanistic explanations resort to parts so as to explain the behaviour of a mechanism. Information regarding the underlying entities involved and their structure is indispensable to characterise a mechanism in its plenitude. If a mechanism *Y* is placed at a given level of analysis, then some of its most immediate parts ($X_{1....}X_{n}$) will belong to the next level down. If we can explain what goes on at the surface by appealing to its parts, then we may get a reductive explanation: we would therefore be explaining the top level functionality or property in terms of lower-level functions or properties. At the same time, to reconstruct the entire explanation for the rise of a mechanism, prior stages and processes that happened on the top level prior to our target phenomenon are also crucial to enlighten *why* and *how* the occurrence of the phenomenon at stake was possible.

[43] To see a defence in favour of partial explanations in social sciences, see Northcott (2012).

Coleman's boat illuminates the compatibility of these two kinds of explanations, causal and constitutive ones, demonstrating how significant they are to elucidate our queries and examining Weber's hypothesis along different dimensions. When we look at the Coleman's boat, we observe a circuit of relations throughout the nodes that can be tied via a series of counterfactual dependencies. Following Woodward's account, the relations across the nodes, to be explanatory, must respond to our what-if interrogations. On top, we have the macro-conditions *A* and the macro-outcome *D*. Both are macrosocial events standing at the same level of organisation where the former temporally precedes the latter.

To endorse Weber's view about the causal roots of capitalism, CITE stipulates the explanandum (the birth of capitalism) should be counterfactually dependent on its explanans (influence of Protestant doctrine). Weber believes it was the rise of a new culture with novel values that displayed new social arrangements, capable of putting in check people's old values. If *A* was absent, then, Weber argues, *B*, *C* and *D* would have also not been brought about. In CITE, contiguity between two actual and distinct events does not constitute a necessary condition to get causality. What is required instead is to exhibit the counterfactual dependence between explanans and the explanandum: in changing the explanans, the explanandum must also change. To say *A* (causally) explains *D* is to acknowledge the causal powers that macro social events may exert in sociological explanations.

In the history of *D*, *A* was a difference-maker for *D*'s rising. So, higher-level explanations, in mechanistic-based explanations, can be explanatorily autonomous so long as the dependence between explanans and explanandum respects the counterfactual criterion. They (higher-level explanations) can be perfectly seen as causally efficacious as far as the hypothetical scenarios considered by mechanists who appeal to CITE *do not* go down to the level of individuals. In other words, the potential autonomy of higher-level explanation requires that the explanans evaluate alternative scenarios that ignore information about individuals.

Alternative plots can only list hypothetical *macro*social aspects that turn the presence of *A* inactive such as (a) the absence of Protestant doctrine in England, or (b) the massive presence of other religious beliefs in capitalist countries (e.g., Islamism, Catholicism etc.) rather than the Ascetic cultural dominance. If any of these imaginary worlds where we change the explanans makes the explanandum also change, then we may bet the explanans may count as a legitimate cause for the rise of the explanandum.

In this sense, macro social events can perfectly figure in the explanans of sociological explanations. Macro level explanations are not necessarily what John Watkins (1968, p. 271) calls "half-way explanations", cases where a large-scale social phenomenon is explained with

reference to another large-scale phenomenon. Now it becomes clear that when we adopt mechanistic explanations, many disputes around the holism-individualism debate over which social features hold causal powers and could perform the explanans' role in sociological explanations are pointless:

> […] once we give up the outdated deductive model of theory reduction, many of the traditional fixations of the methodological individualism debate simply become meaningless. For example, there is no need to provide individualistic acceptable redefinitions of macro-social notions because the explanation of macro facts is no longer conceived as a logical derivation. (Ylikoski, 2012, p. 25)

The benefit of adopting mechanistic-based explanations is to break the explanandum into different pieces of knowledge. While some pieces amount to different stages and processes for the rise of the explanandum placed at the same level, other relevant information amount to structural units, components that should be active for the proper behaviour of the mechanism.

When mechanists couch their explanations, it comes in a piecemeal fashion, focusing on special targets at different phases in the causal history of the mechanism (e.g., event *X* that contributes for the manifestation of *Y*), and different structural aspects that contribute to the manifestation of mechanism as such (structural item *W* that is active for the manifestation of *Z*'s behaviour). That's why, sometimes, information from the macro-level performs the explanatory role, while other times the explanans role is granted by information from the micro-level. It depends on which features the mechanist is seeking to explain at the time. In conclusion, we are entitled to say that higher-level explanations can be explanatorily autonomous, an autonomy that may only stand on a case-by-case basis.[44]

Consequently, it is not mysterious how different explanations about the same phenomenon can accommodate the autonomy of higher-level explanation with a reductive explanation. It's just a matter of changing the explanans and explanandum to understand different *aspects* of the same target. In the case hereby addressed (the rise of capitalism), a full comprehension of this mode of economic production and sociability surely demands we step into the micro-level (the level of individuals).

In Coleman's diagram, arrow 3 relates the properties of individuals to the properties of the system. As Weber tells us, it was a new pattern of economic behaviour of individuals that set the novelty of capitalism as such. The properties that make up the system are entirely

[44] Recent example of higher-level explanations in economics is the latest Nobel Laureate Robert Shiller's book *Narrative economics* (2019). In this book, Shiller argues that the explanation of several economic events lies in the proliferation of narratives, stories that went viral and drove the consequence of major economic events. In Shiller's examples, the explanans are events placed at macro-level, an idea that is entirely consistent with mechanistic accounts.

dependent on the properties of their components. Could we then say that we offer a reductive explanation for the rise of capitalism? In mechanistic-based explanations, the answer is yes.

More precisely, pair (*C*-*D*) exhibits an inter-level part-whole relation between components and a system under analysis. In so doing, this node represents the prescriptions of microfoundations the requirement where the macrosocial property *S* is explained by the properties of individuals. It relates their properties that are synchronically related. The qualitative properties of the system were fixed by the properties of the components. Unlike Nagelian reductions, mechanisms do not appeal to syntactic or semantic bridges between different levels of inquiry to achieve reduction via logical inferences. Adopting CITE, we should demonstrate instead that the relation between the properties of the system and properties of the parts respect the counterfactual criterion: if the properties of the parts were different, the system would have been also different.

When searching for the reasons for the emergence of the capitalist system, reductive queries pay attention only to information from the bottom line (the micro level) of the society. In Coleman's boat, node *C* represents the social position of capitalist components: at one side, there are the owners of the means of production; at the other side, there is the labour force. Node *B* provides the theoretical assumptions, the rationality commitments that give intelligibility to the behaviour of our typical capitalist and household. In this regard, if *C* (constitutively) explains *D*, i.e., if the properties of *D* counterfactually depend on the properties of *C* and accepting the plausibility of the theoretical assumptions assumed at Node *B*, then we have a reductive mechanistic explanation: we may finally understand how the constituent properties of the capitalist system determined it.

6. Concluding remarks

In this chapter, I laid out a contemporary influential explanatory framework used across various sciences, namely, the mechanistic-based approach. Contrary to what we perceived in Nagel's view introduced in the first chapter, there's no explicit demand for bridge principles to bridge distinct layers to enact an inter-level explanation (therefore, reduction). By adopting the Woodwardian view, the various layers within a mechanism are articulated in a harmonised sequence of counterfactual dependences. Reduction consists in a stage, an explanatory phase of this entire machine in operation when the mechanist relates the explanandum top-level property to the next lower-level properties (explanans). Moreover, a constitutive-mechanistic reduction recognises the explanatory power of higher-level properties or events in the

mechanism – this is just a different relation of dependence between different nodes inside the mechanism (with different relata). With our reductive frameworks in place we shall now, in the next chapter, examine a case study from macroeconomics, namely, the Phillips curve.

CHAPTER 3: THE PURSUIT OF REDUCTION: TWO VIABLE PATHWAYS TO REDUCING MACROECONOMIC PHENOMENA TO MICROECONOMICS

1. Introduction

In this third chapter, I assess the applicability of the two contending frameworks introduced in the first two chapters to unlock reduction in economics: Nagel's modelling and the constitutive-mechanistic approach. To accomplish this goal, I analyse a canonical example, the Phillips Curve model in Robert Lucas's path-breaking article "Econometric policy evaluation: a critique" (1983 [1976]). Traditionally, the Phillips curve is an equation that represents the relation between rising inflation with falling unemployment. In his model, Lucas supposes that this relation is a consequence of an association between a rise in inflation and a rise in global supply. Lucas derives the correlation between the aggregate supply (global output) and inflation as an effect of the law of supply – a microeconomic principle. In other words, the manifestation of a macroeconomic phenomenon is inferred from the fundamental microeconomic units: individual agents and their features.

The mechanism behind this relation at the macroeconomic level results from an asymmetry in information at the disposal of the suppliers. Summarising, in their attempt to maximise profit, firms cannot accurately discriminate how much an elevation in the price of their inputs and outputs corresponds to an increment in the demand in their own sector or a rise in all the prices in the economy – i.e., inflation. The macroeconomic Phillips curve relation holds in this context where firms elevate their production in virtue of their misinformation about what is going on in the economy.

In a scenario of limited information about all the supply and demand conditions, agents interpret any variation in the general level of prices (a macroeconomic variable) as expressions, at least partially, of transitory variation in (individual) market-specific prices. If the rise in the price of a firm's outputs was known to be mainly economy-wide inflation, suppliers will rapidly adjust their productions, returning to their normal level of production.

In this model, Lucas argues that changes in the money supplied by central banks can circumstantially have impact on real output. However, its effect on the output is short-lived – it disappears as soon as suppliers revise their next supplying decisions. Therefore, inflation can have real impacts on the economy, where the Phillips curve is a consequence of agents' surprise. The model shows why the output-inflation trade-off is not an autonomous macroeconomic regularity: it is a derivative phenomenon that arises because of people's expectations: only unanticipated inflation by economic agents can affect the output. Consequently, the Phillips

curve isn't a trustworthy macroeconomic relation to be exploited by policymakers since people won't be regularly surprised – one cannot systematically do the unexpected.

Phillips curve is an interesting example. It helps to remind us that, in the social world, relationships (regularities) are fragile. So, we should not be amazed that every pure macroeconomic explanation is only accidentally correct and easier mistaken. Pure macrosocial explanations are often short-cut forms of explanations:

> There may be unfinished or half-way explanations of large-scale phenomena (say, inflation) in terms of other large-scale phenomena (say, full employment); but we shall not have arrived at rock-bottom explanations of such large-scale phenomena until we have deduced an account of them from statements about dispositions, beliefs, resources, and inter-relations of individuals. (Watkins, 1968, p. 271)

It was following this line of reasoning argued by a prominent methodological individualism apostle, John Watkins, that Lucas suggested what he envisioned as a more promising agenda: to avoid the sin of fragile higher level/macroeconomic explanations, economists should revise their models and move back to the classical principles of economic theory. From then onwards, economic forecasting should base its models according to the classical postulates that (a) economic agents act in their own interests (based on rational choice theory) and (b) markets always clear (i.e., there is no excess of supply and demand).

Advocates of this kind of microeconomic modelling argue that the development of economic models based on these principles offers more reliable instruments to track economic relations than untrustworthy dehumanised macroeconomic statistical relationships that have no connection with what real people are actually doing (Lucas, apud De Vroey, 2016, p. 188). According to Lucas, any macroeconomic phenomena should be explained as a consequence of individuals' behaviour. Macroeconomics is not a particular branch of the economic discipline with its own laws and principles, contrary to what Keynes originally thought (Lucas, 1987, p. 107-108). Any macroeconomic model without choice-theoretical microfoundations is, according to Lucas, sub-standard (De Vroey, ibid., p. 176).

In opposition, some post-Keynesians argue that if there is a sub-discipline in economics that needs a foundation, it is not the macro but the very domain of microeconomics: macroeconomics should have priority because it is only in the macroeconomic context that (micro) economic decisions are undertaken. We should therefore look after the macrofoundations of microeconomics (Chick, 2016, p. 99).

By contrast, microfoundational advocates, including Lucas, argue in favour of the primacy of microeconomics before macroeconomics. All macroeconomic phenomena are underpinned by millions of individual actions of economic agents – it's people who buy, trade, and sell goods in the market. Only people can count as the immediate cause of macroeconomic

events. In other words, macroeconomic effects can only have immediate microeconomic causes. Macroeconomics is inevitably grounded on microeconomics (Mankiw, 1997, p. 456).

Neoclassical scholars agree that macroeconomics surely affect people's decisions. As rational agents, macroeconomic data provide powerful informational sources that help to diminish uncertainty. However, macro data enter the microeconomic world as *exogenous* types of *constraints*: bits of information that may influence individuals but are independent of people's behaviour.[45] Among those exogenous variables, the most relevant one is economic policy. Central Banks can determine the supply of money in the economy regardless of individuals. Agents' expectations, in contrast, are endogenous: as rational agents, they react to new information available in their environment.

On Lucas's view, economic policy can only fix a set of macroeconomic constraints across the economic setting. The causal powers of macroeconomic entities manifest their effects on macroeconomic phenomena only indirectly: it's people who count as the immediate and direct cause of macroeconomic outcomes. Lucas's form of individualism represents a radical individualistic approach

> […] where the actions of individuals are seen as resulting from (a) his/her psychology, (b) the physical surrounding, and (c) the actions of other individuals. […] The rule guiding this form of methodological individualism says that no economic explanation is considered successful until all exogenous variables have been reduced to psychological states of individuals and natural constraints. Social institutions may appear in the models of neoclassical economics, but only as endogenous variables. (Udehn, 2002, p. 483)

It's noticeable that Lucas's individualistic view finds echo in literature. Popular economic textbooks such as Mankiw and Taylor (2017, p. 8) define microeconomics as the branch of economic discipline that studies "how households and firms make decisions and interact in specific markets," whereas "macroeconomics is the study of economic-wide phenomena".[46] While microeconomics is more about how people make decisions under a set of constraints – which also has its own regularities (the law of diminishing returns, the Engel's law), macroeconomics pays attention to the regularities among their aggregates (GDP, rate of inflation, etc.) where the Phillips curve is one of them. From these definitions, it is not at all surprising that there is a potential reductive impulse among many economists:

> The reductionist methodological predisposition that economists of almost all persuasions share to some degree, according to which no economic explanation of

[45] Exogenous variables are those that are determined by factors outside the system whose values are imposed by external forces of the model. Endogenous variables are those components that are determined in the model. See, *Oxford Dictionary of Economics* (2017).

[46] Historically, this view remains widely accepted across different generations of economists. Examples of identical definitions of "microeconomics" and "macroeconomics" are found in Allen (1967), Henderson and Quandt (1980), Hall and Taylor (1986), Gordon (1993), and Hoover (2010).

> economic phenomena is truly satisfactory if it does not reduce the phenomena to a question individual actions by basic decision-making units. (Howitt, 1987, p. 6107).

Kevin Hoover (2001a), (2001b), (2010) resists this temptation. He highlights that crucial macroeconomic entities like GDP and the general price level cannot simply be decomposed to their microeconomic counterparts – the general price level, for instance, is not even analogous to the units of any individual price (Hoover, 2010, p. 344). Moreover, people unequivocally refer to macroeconomic entities to diminish their chance of committing mistakes as part of their decision-making processes (Levy, 1985; Hoover, 2001b). Apparently, there's no chance to eliminate macroeconomic entities, turning reduction impractical, which makes Hoover advocate in favour of a non-reductive pathway to bridge micro-to-macroeconomics – global supervenience.[47] So, reduction is a dead end, right? Not really. It depends on what "reduction" means and which goals the inquirer is seeking to achieve with it.

In this chapter, I evaluate this possibility. The impossibility of decomposing some macroeconomic entities does not level any serious barrier to reduction necessarily. In this dissertation, I treat reduction as a explanatory strategy: a special sort of explanation whereby the inquirer transfers the explanans from one level of inquiry to the next level down. The goal of reduction, in economics, is to obtain *explanatory parsimony*: to limit the explanans of macroeconomic phenomenon solely to the properties of individuals. If this is possibly to be achieved, we then may have a reduction and the selected frameworks introduced in this dissertation might do the work.

According to our explanatory interests and the target under assessment, we may have reduction in cases where these irreducible macroeconomic entities do not play a relevant causal influence on the manifestation of the chosen macroeconomic outcome. In a context of an economic phenomenon, reduction is attained by demonstrating that the explanans at the microeconomic level is performed mainly by individuals' properties rather than any other outer features inside the microstructure. I argue that if the explanandum, a macroeconomic phenomenon *E* is primarily determined by microeconomic principles and passive inputs (i.e., other features inside the microlevel without relevant causal efficacious for the target macroeconomic outcome), then we may have reduction according to two different schemes: Nagel's account and constitutive mechanisms.

To illustrate how this is possible, I present Lucas's Phillips curve explanatory model as a paradigmatic case of microeconomic reduction. He does so by constructing a small-world economy where the correlation between output-inflation appears in its purity by virtue of an

[47] I discuss the application of supervenience in the next chapter.

optimisation problem undertaken by a typical economic agent under a set of constraints.[48] Macroeconomic inputs enter the micro level as a source of information, i.e., a constraining component.

The correlation on the top-level results from people's beliefs and expectations under a set of constraints where past and actual data of inflation (thus, macroeconomic data) are among the informational sources of individuals. Although these macroeconomic inputs are present in the base, they only appear as a passive constraining pieces of information: people's beliefs and expectations prevail in determining the trade-off between output and inflation. Only the states of minds of individuals are causally significant for determining the macroeconomic trade-off. It is in favour of this sort of psychologism, one of the many faces of methodological individualism, that Lucas argues for.[49]

Here is the road map. I initiate by presenting the Lucas's Phillips curve model (Section 2). Then I layout the type of knowledge produced by the Lucasian modelling – to teach us about economic capacities (Section 3). Moving on, I fit this model into the first reductive framework: Nagel's reduction (Section 4). Finally, I assess the secondary reductive paradigm: I characterise the model under the mechanistic framework (Section 5).

2. Lucas's "surprise" aggregate supply

The birth of the Phillips curve is dated: it was born in 1958 when the economist A. W. Phillips published an article where he found a stable relationship between money wages and the rate of unemployment. Analysing the data collected from Great Britain between 1861-1913, Phillips identified a nonlinear, negative correlation between those two macroeconomic variables. Historically, the Phillips curve hypothesises a negative stable relationship between inflation and unemployment. Accordingly, a higher inflation is associated with lower unemployment. Over decades, the curve represented a causal exploitable trade-off in which policymakers trusted and explored politically.

[48] Representative agent is a class of economic models. It refers to a typical single economic agent (a typical household, a typical firm) that abstract away qualitative differences among firms or households – they are all alike. Firms aim at maximizing profit while households want to maximise utility. The rules that govern the typical agent are assumed to hold in the aggregate as well.

[49] Psychologism is one of the many faces of methodological individualism. It advocates that social explanations must be ultimately based on (or reduced to) the psychology of individuals. John Stuart Mill (1974 [1843-1872]) is famous to endorse this approach.

The Traditional Phillips curve, popularised by Samuelson and Solow (1960), suggests an explicit relationship between the wage-inflation rate $\frac{\Delta w}{w}$ and the level of unemployment U, of the kind:

$$\frac{\Delta w}{w} = \alpha U^{z} + \beta$$

wherein α, β, and z *are* fixed parameters. In the Lucas framework, given the story he tells about this equation, when it holds, a rise in inflation generally causes lower unemployment. In a scenario where the economy is overheated, e.g., when the demand is growing faster than supply, firms tend to raise their prices and hire more workers to follow the increment in demand. Econometric estimations of this relationship aim at inferring the numerical value of the parameters from observed paths of U_t and w_t. However, this equation, Lucas argues, will break down if the government tries to use monetary policy to improve employment.

Lucas (1983 [1976]) suggests that the actual relation between output and inflation, which drives the relation between unemployment and inflation, must be described by specifying the optimization behaviour of the individuals plus their forward-looking expectations. He aims to demonstrate that the behaviour of firms is not fixed (or stable) and that it will be changed under environmental modifications, especially following changes in monetary policy. The rationale behind Lucas's critique concentrates on how economic agents usually process information from their environment. Lucas claims that suppliers must respond to a "signal extraction" problem while making decisions based on relative prices in the context of imperfect information. Firms must determine what portion of price changes in their respective markets reflect a general change in nominal prices (inflation) and what portion reflects a change in the real prices for their inputs and outputs.

Lucas hypothesized that suppliers know their businesses better than the prices of other markets. Given the asymmetry in information, a supplier could perceive a general increase in prices due to inflation as an increase in the relative price for its output, reflecting a better, real price for its output and encouraging more production and, thus, more employment. But, Lucas argues, the correlation between inflation and output (on which the Phillips curve is based) is only held as long as individuals cannot accurately anticipate the expected inflation. This "surprise" will lead firms to expand their production throughout the economy. However, once agents recognise inflation for what it is – that what they observe is not caused by a rise in just their prices – they will rapidly revise their attitudes. So, inflation that happens without agents

recognising it as inflation can lead to reductions in unemployment. But when the government uses it as a lever, then agents will recognise it as inflation and respond to it differently. Thus, a proactive monetary policy cannot systematically "buy" some points in GDP, as post-Keynesian economics argues (Phelps, 1970, p. 2).

Now, to turn to the issues of reduction, recall that Nagelian reductions require a derivation of the targeting higher level (macroeconomic) phenomenon from micro facts or/and theoretical principles from the reducing domain. Lucas, in his article, supplies this derivation for us. First, his theoretical assumptions are:

(a) There is only one type of shock in the economy: monetary shock;
(b) The information is imperfect (i.e., the total information about the supply and demand conditions are not given *a priori*);
(c) agents are price takers (i.e., they cannot influence the market price on its own). They must accept the prevailing prices in the market;
(d) agents based their (economic) decisions fundamentally driven by relative (market-specific) prices;
(e) there is no monopoly (which means a scenario of ideal competition);
(f) all individuals' plans are fairly compatible (markets-clear – i.e., they are not characterised by excesses of supply and demand);
(g) agents are rational – i.e., people maximize their utilities or profits under a set of constraints using all information available (*homo oeconomicus*).

Lucas presents his argument through a sequence of equations. As a matter of clarification, here is a list of the symbols:

Table 2: List of symbols

N= quantity of markets in a closed-system economy
i= individual market (e.g., the fish market)
$y_{i,t}$= quantity supplied by each supplier at time t
$y_{i,t}^{p}$ = fixed level of production of a firm (permanent quantity supplied)
$y_{i,t}^{c}$= quantity supplied of a firm sensitive to price changes (cyclical component)
$p_{i,t}$ = specific-market price of a firm i at time t
$p_{i,t}^{e}$ = suppliers' in market i's perception of the common component of the price level in each market
β = behavioural parameter of individuals
P_{RGt} = real general price level at time t
$z_{i,t}$ = random disturbance to each market
θ = parameter determined in part by the variance of $(z_{i,t})$ and the variance from the distribution that market suppliers attribute to (p_t)
p_t = a contribution to $p_{i,t}$ that is common to all markets

$\overline{p_t}$ = anticipated average of the price level (p_t) under the distribution presumed by suppliers I_{t-1} = information based on the past events $\mathbb{E}$ = expectations conditioned to market-specific prices, real general price level, plus past information Y_t = aggregate supply $Y_{c,t}$ = cyclical component of aggregate supply $Y_{p,t}$ = permanent component of aggregate supply

To the problem. Consider a given economy that is composed of *N* markets in different places supplying the same good. The suppliers' production ($y_{i,t}$) in each market *i* at time *t*, ∀*i* = 1...*N* consists in a permanent ($y_{i,t}^{p}$) and cyclical ($y_{i,t}^{c}$) components of supply:

$$y_{i,t} = y_{i,t}^{p} + y_{i,t}^{c} \quad (1)$$

where ($y_{i,t}$) is the quantity supplied by firm *i* at time *t*, ($y_{i,t}^{p}$) is the normal supply unresponsive to price changes (the normal level of production), all in logs, and ($y_{i,t}^{c}$) is the variant or cyclical supply sensitive to oscillation in market-specific prices (it varies from market to market) that is perceived by firms. The cyclical component is supposed to be a linear function of the deviation between the actual market-specific prices (the face value of market-specific prices at time *t*) and the perceived common component of the general price by producers in market *i*:

$$y_{i,t}^{c} = \beta(p_{i,t} - p_{i,t}^{e}) \quad (2)$$

where ($p_{i,t}$) corresponds to the actual price (in logs) in market *i* at time *t*, ($p_{i,t}^{e}$) is the suppliers' in market *i*'s perception at *t* of the common component of the price level, and β denotes the intertemporal substitution possibilities (the agents' decision rules that vary across the time and under different policies) in supply: technological factors and tastes for substituting labour today for labour tomorrow (Lucas, 1983 [1976], p. 275). Lucas assumes that the specific-market at *t* in market *i*, ($p_{i,t}$), which is observed by agents, is determined by:

$$p_{i,t} = p_t + z_{i,t} \quad (3)$$

where (p_t) is a variable common to all markets and ($z_{i,t}$) is a random disturbance – normally distributed, with mean 0 and variance τ^2, independent from the distribution of (p_t) – representing variations in relative prices across time and markets. Based on past information (I_{t-1}), market suppliers suppose that the distribution of (p_t) is normal with mean ($\overline{p_t}$) and a

variance σ^2. The two components of $(p_{i,t})$ cannot be observed separately by market suppliers. The real general level of prices at time t (P_{RGt}) is simply described as the average across the N markets of the i particular prices:

$$P_{RGt} = \frac{1}{N}\sum_{i=1}^{N} p_{i,t} \tag{4}$$

Substituting (4) into (3) yields

$$P_{RGt} = \frac{1}{N}\sum_{i=1}^{N} p_{i,t} = p_t + \frac{1}{N}\sum_{i=1}^{N} z_{i,t} \tag{5}$$

For an N that is large enough, Lucas assumes the second term will be small. So, we have:

$$\frac{1}{N}\sum_{i=1}^{N} z_{i,t} = 0 \tag{6}$$

Thus, general price level in the economy at time t is (P_{RGt}), and is equivalent to (p_t), as expressed in equation (8):

$$P_{RGt} = p_t \tag{7}$$

Suppliers in each market are supposed to determine their supplies according to their expectations $(p_{i,t}^e)$:

$$p_{i,t}^e = \mathbb{E}((p_t \mid p_{i,t}, I_{t-1}) \tag{8}$$

Lucas then claims that suppliers base their supply decisions on their estimate of (p_t) – i.e., $(p_{i,t}^e)$. He then assumes that "this estimate $p_{i,t}^e$ is the mean of the true conditional distribution of p_t" (Lucas, 1983 [1976], p. 274) – based on the current price and in their market prices past information.

Calculating this from the assumption about the distribution of (p_t) and $(z_{i,t})$ yields

$$p_{i,t}^{e} = \mathbb{E}(p_t \mid p_{i,t}, I_{t-1}) = (1 - \theta)p_{i,t} + \theta\bar{p}_t \quad (9)$$

where $\theta = \frac{\tau^2}{\sigma^2 + \tau^2}$ (τ^2, recall, denotes the variance in the distribution of $(z_{i,t})$, the relative price variation for market *i* at *t*, and σ^2 is the variance of the general price level estimates conditional on (I_{t-1})), and $\overline{(p_t)}$ which is the average of (p_t) calculated by the normal distribution that suppliers suppose(p_t) has.

Substituting (9) into (2) shows the cyclical component of the individual suppliers' supply is then

$$y_{i,t}^{c} = \theta\beta(p_{i,t} - \bar{p}_t) \quad (10)$$

Let's assume, again, that the aggregate output can be represented analogously to the output in individual markets. In this case, the aggregate output (Y_t) is composed of the permanent $(Y_{p,t})$ and cyclical $(Y_{c,t})$ components of all individual suppliers. We then have:

$$Y_t = Y_{p,t} + Y_{c,t} \quad (11)$$

Following the equation (10), the cyclical variation of the level of output in a given market *i* directly depends on the spread between the actual price level in that market *i* and the anticipated average common price level $\overline{(p_t)}$. The cyclical component of the macro quantity aggregate supply, which I hereby prefer to label as $(Y_{c,t})$, is identified with the average over all markets *i* – the cyclical component supply in each market, as in:

$$Y_{c,t} = \frac{1}{N}\sum_{i=1}^{N} y_{i,t}^{c} \quad (12)$$

Assuming that *N* is large and invoking the law of large numbers,[50] we may finally get the cyclical component of aggregate supply:

[50] In probability theory, the law of large numbers is a theorem that describes the result of performing the same experiment a large number of times. Accordingly, the average of the results obtained from a large number of trials should be close to the expected value and will tend to become closer to the expected value as more trials are performed.

$$Y_{c,t} = \theta\beta(p_t - \overline{p_t}) \quad (13)$$

By reintroducing the permanent component ($Y_{p,t}$) in the equation (13), we therefore obtain the aggregate output:

$$Y_t = Y_{p,t} + \theta\beta\ (p_t - \overline{p_t}) \quad (14)$$

From this, we have:

$$Y_t = Y_{p,t} + \theta\beta\ (P_{RGt} - \overline{p_t}) \quad (15)$$

The difference ($P_{RGt} - \overline{p_t}$) is the difference between the real general price level (P_{RGt}) and the average expected by suppliers based on past information $\overline{(p_t)}$. Here we see the importance of the micro reduction: it is the last term on the right-hand side of the equation (equation 15) that can generate the Phillips curve. Under the condition of stability among the parameters θ, β, an observable (econometrically estimable) stability between output and inflation follows from this.

3. Model's lesson: to teach about capacities

In his model, Lucas isolates some key economic variables, which may determine the manifestation of the Phillips relation by their causal features. In that idealised picture of the social world, Lucas does not describe any actual events – there is no *actual* agency from the suppliers. Instead, he aims to comprehend in which possible ways the set of target variables might behave in an analogous (plausible) economic system when they are unimpeded. In such model, isolation is a powerful tool employed by the model builder to understand how the causal powers of some target entities might behave when they are harmonised with other elements in the system (Cartwright, 2009).[51]

[51] The type of isolation employed by Lucas is what Uskali Maki (1992) calls "theoretical isolation", i.e., when, in purely theoretical contexts (e.g., economic models) some chosen elements represented in the model are closed from the causal interference of some other elements present in the model.

Capacities or other equivalent terms, like tendencies, propensions, and dispositions, refer to stable economic tendencies. These tendencies are not only technical devices but are also supposed to work in the real world. For instance, in accord with microeconomic theory, firms produce up to the point where their price equals marginal cost. Marginal cost depends on the price of the firm's input to production. In a competitive economy, quantity choices are determined mainly by agents' reactions to market-specific prices.

We may expect a representative firm *i* will produce more if the price of its output rises relative to the expected value in *i* of other prices in the economy, which in Lucas's model are represented by $(p_{i,t}^{e})$ – the expected level in market *i* of the average general price level expected input prices. Consequently, an increase in the price of output up to the cost of its production leads firms to expand production. This microeconomic principle does not describe the actual decision-making of suppliers under uncertain conditions – but instead suggests what firms, in a competitive market economy with uncompleted knowledge, normally do. After all, this is what we might expect from a firm, what motivates the decisions of a typical welfare seeker:

> "Laws" such as "People seek more wealth" are supposed to identify relevant causal factors. These may be counteracted by other factors and prevent from operating, and so, for example, people will not always seek more wealth. But a desire for more wealth remains a significant causal factor. When economists state that people desire more wealth, they are stating this truth. (Hausman, 2001, p. 297)

It is sure, though, that some deviations are more than expectable. For example, we may say, conforming to economic theory, that people tend to elevate or diminish consumption according to their costs. Nonetheless, someone heavily addicted to heroin may easily escape this tendency – her demand for heroin tends to be inelastic (i.e., it does not vary with prices). But these deviations do not invalidate that there is an overall *approximation* about the general motives of suppliers (profit), or about what makes households increase their consumption (rise in real wage) etc.:

> Historically, rational economic man [...] captured a real tendency in human behavior, which dominated but did not completely determine, certain individual and social outcomes. Reasoning based on such a pervasive [economic] motive was, therefore, a good approximation for many purposes and circumstances. (Hoover, 2010, p. 337)

Lucas, in his model, tries to uncover some facts about stable tendencies (Cartwright, 2007, p. 221). He wants to describe relevant tendencies that typical suppliers normally have in dynamic free-market economies and which aspects of this socio-economic environment are significant for their economic decisions. On Lucas's view, typical suppliers *tend* to collect as much information as possible and use it economically to optimise their decisions (rational expectations); they *tend* to optimise their resources and look after profit as much as possible; they *tend* not to systematically alter their technological factors under changes in economic

policy; and households *tend* to have an organised ranking of preferences that is unaltered by policy regime.

By taking for granted a set of axioms, Lucas articulates three economic tendencies associated with fixed (exogenous) parameters, beliefs, and (rational) expectations of economic agents under imperfect information scenario that operating together are pictured as capable of performing the explanatory role for the possible rise of the Phillips curve. The macro phenomenon's generative mechanism is performed by suppliers' properties. The list below offers an overview of each of these properties deputised to suppliers:

(a) *Stable dimension of agent's behaviour – the fixed parameters*: In economic theory, there's a common-ground assumption that two elements of agent's economic dynamic behaviour are relatively stable across the time: preferences and technology. The credo among neoclassical scholars like Lucas is that, on demand side, every consumer has an ordering, a comparative ranking of preferences in their minds whatever options are available. Give me a set of options, say choosing between (1) David Lynch's *Eraser head* or (2) Marvel's *Spider-man,* and I will easily tell you which movie I'd prefer – I go for (1). An economist infers my preferences based on what I have chosen; this is what is possibly (indirectly) observable.[52] The formal mapping of preferences takes for granted that households have perfect information about the objects of their choices.[53] On the supply side, each firm has technological possibilities and constraints to transform raw materials (input goods) into commodity outputs. In theoretical models, these parameters are what Boland (2017, p. 94) labels as the "exogenous givens" parameters,

[52] There is a current vibrant debate on the nature of preferences in economic theory/philosophy of economics. Dan Hausman (2011), for instance, argues in favour of a mentalistic view: preferences are complete subjective comparative evaluations of a set of choices. In Hausman's account, preferences, in the context of economic theory, do not refer to a simple desire but instead to a sophisticated cognitive process. To see a criticism against Hausman's approach, see Angner (2018). Alternatively, the mainstream understanding on preference equates it with actual or hypothetical behaviour (Revealed Preference Theory, RPT). As a representative summary, see, for instance, Harsanyi (1977), Binmore (2008), Gul and Pesendorfer (2008). For a recent philosophical defence of the revealed preference approach, see Thoma (2021). More recently, though, Francesco Guala (2019) suggests a dispositionalist view on preferences. In his view, preferences are neither behavioural nor mental: they are belief-dependent dispositions with a multiply realizable causal basis. In this debate, Lucas seems to assume by "preference" as a primitive notion within the economic theory that does not require explicit definition about its content. As Angner (2018, p. 665) claims, "[…] economics as a science can be independent of any claims about the origins of preferences and their relationships to subjective phenomena". In this debate, I assume a mentalistic perspective about preferences. Following Dietrich and List (2016, p. 268), who believe that the best scientific practice on decision-making is committed to the reality of mental states. Mental states can be taken in functional terms: something is a mental state simply because it plays the relevant role for agents' attitude.

[53] The standard axioms of preferences are *completeness* (given any two options *A* and *B*, consumer can rank them – *A* if preferred to *B*, *B* is preferred to *A*, or they are indifferent), *transitivity* (if *A* is preferred to *B*, and *B* is preferred to *C*, then *A* is preferred to *C*) and *reflexivity* (*A* is at least as good as *A*). This represents the standard starting point of choice theory. Nowadays, the representative paradigm of choice theory is Revealed Preference Theory, historically defended by Paul Samuelson (1938), equates preferences with actual or hypothetical choice behaviour.

the deep parameters that are assumed as constant under changes in economic policy. On the supply-side, rational suppliers do not systematically change their technological conditions to maximise profits, nor do households systematically change their ranking of preferences. As Lucas (1981, p. 220) says, it is "[...] reasonable to hope that neither tastes nor technology vary systematically with variation in countercyclical policies".[54]

(b) *Dynamical dimension - Beliefs*: suppliers observe actual market-specific prices and access past information about inflation, the two sources of information. They know their profitability is undissociated from consumers' preferences – households' preferences underneath the face value of suppliers' outputs. Sadly, suppliers have no information about demand conditions. In this regard, firms interpret their collected information according to their background knowledge (mostly guided by the dominant neoclassical theory) as intelligently as possible.

(c) *Dynamical dimension – Expectations*: These are relevant psychological phenomena that cannot be directly observed but affect the economy. Expectations are nothing but "[...] proper attitudes, dispositions, or states of mind which determine our behaviour, or at least accompany it" (Ozga, 1965, p. 23). According to the Rational Expectations Hypothesis (REH) adopted by Lucas, economic agents use current available and relevant information in forming their expectations and do not purely rely on past experiences. It also assumes that the information gathered is processed and analysed optimally and efficiently to arrive at an intelligent estimate or expectation of the value to be taken by the economic variable under consideration (Shaw, 1984, p. 47).[55] In Lucas's model, suppliers form their expectations about inflation based on past information about their target variable (past inflation).

Each of these elements plays a different role. (a) is the stable dimension of suppliers and households' behaviour, the insensitive dimension in their behaviour that, despite all the dynamic behaviour of individuals, these two features (tastes and technology) are relatively unresponsive to environmental changes. (b) – (c), in turn, pay attention to the objective signals

[54] It's worth noticing that the stability of parameters is assumed as fixed in theoretical models. This is not the case of empirical models when economists do not always make such assumption (Boland, 2017). Lucas (1981, p. 11-12) said that the stability of parameters should be evaluated in empirical, rather than logical grounds.

[55] I just summarise what is normally called the weak version of REH. In the strong version of REH, agents' expectations of economic variables will coincide with the true or objective mathematical conditional expectations of the chosen variables. For an overview about different forms and versions of REH, see Redman (1992).

from the economic setting. They are both heavily intertwined. (b) corresponds to suppliers' beliefs about which elements of the economic system determine their profitability. They know their profitability depends on consumers' preferences and production costs. At *t*, suppliers observe market-specific prices and access information about past inflation to optimise their decisions. They process the set of informational sources and base their decisions on taking into account what they believe is the actual causal structure of the economy.

Similarly, (c) also involves suppliers' general understanding of the causal structure of the economy. In Lucas's model, suppliers form their expectations about the future values of the general price level (their target of interest) based on past information. While forming their expectations, suppliers collect and analyse relevant information as if they were welfare maximisers (Shaw, 1984, p. 50). To the extent that suppliers do not collect all information and every piece of information is subject to a certain degree of uncertainty, errors (imperfect foresight) are more than expected. What is denied by (REH) is that agents will make expectational mistakes indefinitely. (a), (b), and (c) – presents a hierarchy of tendencies: (b)–(c) are grounded in (a), and they are both *subordinated to* (a). In other words, while (a) is relatively fixed, i.e., it is regularly unaffected by the context, (b)–(c) are dynamic – we expect the variables we used to represent (b) – (c) to modify their values due to changes in the environment whereas (a), in contrast to (b) – (c) does not.

3.1 How the Lucasian model represents its target

Contemporary philosophy of science presents a vibrant debate concerning the representational power of scientific models, how these abstract entities that intentionally distort their targets in various ways may represent them, and how they can be informative about the targets they supposedly represent. One dominant approach interprets models as fictional stories analogously to literary novels: "[…] modelers often treat model systems in a 'concrete' way that suggests a strong analogy with ordinary fictions" (Godfrey-Smith, 2006, p. 739). Like literary fiction, scientific models are products of the voluntary cognitive activity of modelers who create a fictional world (i.e., a model of perfect market competition) similar to Thomas Mann's *Doctor Faustus*.

In such an account, the atomistic model introduced by Nils Bohr, subsequently replaced by Erwin Schrödinger quantum mechanics, holds the same fictional status as the character Adrian Leverkühn. Both fictional entities are not regulated by factual conditions; they are by-products of their creators. As stressed by Arnon Levy (2012, p. 740), a world of fiction is a world

that is non-regulated by truth conditions or constrained by empirical facts. That is the reason why Roman Frigg (2010, p. 253) qualifies scientific models as "[…] imagined physical systems, i.e. as hypothetical entities that, as a matter of fact, do not exist spatio-temporally but are nevertheless not purely mathematical or structural in that they would be physical things if they are real". So how can these imagined worlds represent their targets?

The question of representation bifurcates into two more-specific questions: (i) the aboutness condition and (ii) epistemic condition. Condition (i) is interested in knowing which property an object *O* in the model must have to be qualified as a legitimate representation of some real phenomenon *P*, what Margaret Morrison (2008, p. 70) calls "[…] the crux of the problem of representation". Condition (ii), in turn, is interested in knowing which property *O* must hold in order to be an epistemic representation of *P*, i.e., how *O* may teach us about *P*.

Indirect accounts of model-world relationship take the mediation between the model and world as made by the model itself.[56] First, the modeller imposes the conditions of the *model system*, a set of instructions that are not regulated by empirical conditions. Such list of instructions constitutes the *model description*. These descriptions are like props that prescribe imagining about the model system. The model system, in turn, is the representational vehicle used to representing its target that could be many things (maps, equations, graphs etc.). And then we have the *model target*, that part of the world that the modeller believes the model system is informative about it.[57] According to Frigg (2010, p. 126), a model system *X* is an accurate representation of its target *Y* iff (i) *X* denotes *Y* (aboutness condition) and (ii) establishes epistemic relevance by imputing properties of *X* to *Y* (epistemic condition).

In Lucas's example explored in this chapter, there is a list of assumptions that impose the model conditions (instrumental rationality, fixed parameters, incomplete information, etc.) – a set of conditions that are the central parts of the model description. Analogously to a map of the city of London that permits us to infer facts from the territory from the facts about the map, Lucasian model permits to translate facts from the model to facts about its target as long as the model system denotes relevant properties of the target *P* – the properties of the model system and its target must be similar in relevant aspects: "[…] when I say that the population of rabbits in a certain ecosystem behaves very much like the population in the Fibonacci model, what I assert is that these populations possess certain relevant properties which are similar in relevant respects" (Frigg, 2010, p. 119).

[56] Representative scholars of the indirect view are Ronald Giere (1988), Peter Godfrey-Smith (2009), and Roman Frigg (2010).
[57] A modified version of this triad relation is claimed by Ronald Giere (1988).

Contrarily, direct accounts of model-world connection claim that model systems are somehow a description of their targets, i.e., they are direct representations of the phenomena they assume to represent: "[…] models are special descriptions, which portray a target as simpler (or just different) than it actually is" (Levy, 2015, p. 791).[58] Despite the differences between these two approaches, both agree that models, in the usual application of several scientific contexts, serve as vehicles of representation, tools in the service of human understanding about some fragments of the world.

Although acknowledging the intuitive appeal of the fiction view of models, I prefer to construe maps, charts, mathematical equations, and other sorts of technical artefacts used to represent their targets as *concrete epistemic tools* (Knuuttila, 2011; Currie, 2017). Novels are not only fictions; they are also the object of inquiry of scholars. Since Mann's publication in 1947, there has been a myriad of scholarly manuscripts (dissertations, theses, and articles) that discuss different aspects of Mann's book: whether the signed pact between Adrian Lerverkühn and the devil constitutes a metaphor to represent the reasons why Germans embraced Nazism (Reiherdt, 1985), the closing gap between music and history (Pratt, 2000), or how the novel partially represents the social and cultural tensions of the Weimar Republic (1918-1933) that subsequently led to Nazi regime (Berland, 1978).

The concrete epistemic tool approach may also respond to both fiction view criteria. To assess the aboutness condition, historians and sociologists may compare the acquired knowledge about the social and cultural tensions in the inter-war period in Germany with some descriptions – or fictional truths – contained in Mann's story. Herein, the intentions of the author are irrelevant when the goal of the scholars is to challenge the representational power of many statements inserted in the book. The fictional statements may acquire their truthlike representational status when they are contrasted with empirical facts. In so doing, we may end up realising that Mann's story potentially shared many of the similarities with the accrued historical knowledge about the cultural effervescence of the Germany during the first quarter the 20th century.

Analogously, the variables within the explanans of the Phillips relation correspond to approximations of the actual capacities responsible for triggering the Phillips curve, they amount to some genuine features that have their causal features manifested in the actual world. In the Lucas's model, the features named in the model are sufficiently similar to its target in some relevant respects. The model lists several actual features from the real world, such as suppliers are profit seekers, people *usually* do not disregard relevant information for the

[58] Representative scholars in favour of direct view are Levy (2012, 2015) and Toon (2012).

purposes of decision-making, and information in real world is highly asymmetrical. Rather than being constrained by some actual state of affairs, all the chosen theoretical assumptions adopted by Lucas are constrained by some actual economic tendencies manifested within the model system.

Despite its purposively adopted abstractions and idealisations, the Lucasian world constitutes a plausible economic world as long it is grounded in such real economical dispositions. Somehow, the model is ontologically constrained: it is limited by some real dispositions that actually occur in the real world (Maki, 2009, p. 32). By preserving the criteria of the model fiction view, the model-world similarity based on capabilities would respond to the aboutness condition – which stand-in relation an object *O* should have to be qualified as a legitimate representation of a phenomenon *P*.

Regarding the epistemic condition, recall that (economic) models can constitute the object of scientific inquiry and knowledge. Economists may contend the Lucasian conclusions by challenging several assumptions in the model: the plausibility of the Rational expectations hypothesis, the stability of parameters, and the instrumental rationality. Criticisms may voice that the Phillips relations stand only due to the stringent conditions and assumptions adopted by the modeller. Such legitimate criticisms are possible because the models constitute the objects of knowledge in themselves (Knuuttila, 2011, p. 267).

We can construe the possible challenges towards the explanans within the Lucasian model as seeking alternative explanations for the rise of the Phillips curve, such as: (i) would the Phillips curve emerge if people form their expectations contrarily to what the REH predicts or (ii) would the Phillips curve be the case if the parameters are not taken for granted as fixed?[59] Lucas's model is an economical cognitive artefact due to its capacity that other scholars have to evaluate it by manipulating its fundamental premises and levelling various *what-if* questions (Knuuttila, 2011; Ylikoski and Aydinonat, 2014). This counterfactual dependence between the explanatory variables within the model and the explanandum grants that the model itself is sufficiently informative about its target.

In conclusion, the Lucasian model serves as case study where we can have a peaceful integration of instrumentalism and realism. When we look at the details, we realise various instruments at the service of the explanation opted by Lucas: idealisations, abstractions, correspondence rules based on definitions, and the like. On the other hand, the model also mimics several real-world tendencies,[60] such as suppliers' motivations, their incentives, how

[59] I come back to this later when I discuss the contrastive view of explanations (chapter 5).
[60] In Lucas'model, the relata of reduction are made of these economical tendencies and dispositions listed in the model, but it could be different types of things as well.

they form expectations, revise their behaviours in accordance with novel data, a mimic that resembles the real-world in some relevant respects.[61]

In the next two sections, I articulate how these tendencies/dispositions illustrate two distinct versions of reductive explanation for the occurrence of the Phillips curve: Nagel's reduction and the mechanistic approach. I firstly layout a Nagel-like version of the Lucasian model (Section 4), and hence I represent Lucas's model according to the mechanistic framework (Section 5).

4. Applying the Nagel's recipe

In this section, I represent Lucas's model according to Nagel's view of reduction. Although Nagel did not carefully assess the possibility of macroeconomics being reduced to microeconomics through his own recipe, he externalised his pessimism (maybe influenced by the large dominance of Keynesian economics of that period) about whether this possibility could someday be achieved:

> To be sure, no proof is available that [...] macroeconomic assumptions cannot be deduced from microeconomic ones. But there is also no proof that the deduction can be effected, economists do hesitate to employ those macroeconomic postulates in their analyses; for, as one student has put the matter, "one may disagree on particular assumptions, institutional or psychological, regarding the saving patterns of individuals or group of individuals, and yet may find the concept of aggregate saving useful in describing the actual or probable behavior of national income". But if this is so, and macroeconomic assumptions enable economists to account for aggregative phenomena no less adequately than do microeconomic postulates, the reduction of macroeconomic to microeconomic explanations appears to offer no substantial advantages. In short, there are nonformal considerations as well as purely formal ones for questioning the merits of the reductive thesis of methodological individualism. (Nagel, 1961, p. 544)

Nowadays, more than fifty years after the publication of Nagel's *The structure of Science* (1961), the microfoundations of macroeconomics became the mainstream programme in economic science. Lucas's Phillips curve is a classic example of a microfounded model that fits nicely with Nagel's prescriptions. In the following lines, I demonstrate how this is possible.

[61] Recently, P. Kyle Stanford (2021) offers a middle-path approach between realists and instrumentalists. To the extent the historical record puts us in a position of identifying which entities are genuinely responsible for the empirical success of special sciences and which are more like the accessory tools, we can embrace a selective form of realist that welcomes many (anti-realist) instrumentalist criticisms against the traditional theses endorsed by realists.

4.1. Formal conditions of Nagelian epistemological reduction

Let's first refresh our memories. In Nagel's reduction, two formal conditions must be satisfied to get reductions throughout: (i) deducibility and (ii) connectability. According to Klein's (2009) epistemological interpretation of Nagel's reduction, condition (ii) is met when the reducing theory uses its theoretical apparatus to represent relevant missing terms present in the reduced theory vocabulary. In this reading, reductive functions (bridge laws) are invoked to channel distinct theoretical domains that were formerly viewed as completely autonomous from each other; they do not necessarily amount to any structural or ontological identity in the world.

In epistemic Nagelian reductions, the bridge laws do not necessarily express identity statements between reducing and reduced terms. If we can infer the top hierarchical statements from the reduced theory via the original lower level statements from the reducing domain with the auxiliary bridge principles without any semantic commitment to the identity of the items referred to in the bridging laws, we, therefore, could enact a reduction in Nagelian terms. So, as Klein argues, identities or biconditionals are non-necessary conditions to link reducing and reduced terms and therefore unfold inferences among different theoretical frameworks.

Schaffner, in contrast, claims that bridge principles are synthetic identities that relate entities and predicates of reducing and reduced theories extensionally. Accordingly, bridge laws should be empirically corroborated, which means, for this particular context (the Phillips curve model), that when the inter-level bridge principles are required, they should be empirically grounded rather than merely expressing conventions. Using the representative-agent-based modelling, Lucas "sidesteps the aggregation problems inherent in macroeconomic analysis by using a representative agent whose choices are assumed to coincide with the aggregate choices of millions of individuals" (Snowden et.al., 1994, p. 265). For this reason, I shall concentrate only on the epistemic account of Nagel's approach couched by Klein.

Lucas's model is essentially an interlevel approach across micro and macroeconomic levels. He aims to derive the manifestation of a macroeconomic regularity – the correlation between inflation and output – from a microeconomic principle. The major premise of his argument (equation 1* below) is the supply function, a purely microeconomic law. It is a micro behavioural assumption that says that quantity supplied is a function of the market-specific prices in the economy. Quantity supplied by firms is made of two components:

$$y_{i,t} = y_{i,t}^{p} + y_{i,t}^{c} \qquad (1^*)$$

where the first term on the right-hand side ($y_{i,t}^{p}$) is fixed, and the second is variant. This equation does not describe the actual production conditions; this is just a theoretical microeconomic assumption. Next, Lucas casts light upon the elements of the cyclical component ($y_{i,t}^{c}$) of this microeconomic assumption:

$$y_{i,t}^{c} = \beta\left(p_{i,t} - p_{i,t}^{e}\right) \qquad (2^*)$$

Equation (2*) is another microeconomic assumption. The cyclical component is determined by the difference between the actual prices at t plus individuals' perception about the common component present in the price of all market-specific prices. These components correspond to the informational inputs that suppliers regularly extract from the environment: (a) signals of the actual market-specific prices ($p_{i,t}$) at time t, plus individuals' perceptions across separate markets about what might be the value of the common component of the price level ($p_{i,t}^{e}$) in each market.

From the point of view of agents, components ($p_{i,t}$) and ($p_{i,t}^{e}$) are means of information: they convey relevant information for suppliers' decisions. In capitalist economies, prices are mostly construed as *tags*: they communicate to economic agents how scarce, desirable, and valuable some goods are. Prices function as an informational source that may incentive entrepreneurs to discover profit opportunities (Kizner, 1992).

Prices are informational vehicles, a sort of social mind-dependent entities that economic agents take into account to enhance their decisions: when prices go up, the level supplied is increased; when they go down, the demand rises. The reality of this type of entity depends partially on agents' subjective attitudes towards them, though these attitudes need not be manifested towards their particular instances.[62] Lucas avoids any philosophical jargon in his exposition, but I think this reading is fairly consistent with his ambition and method.

Next, Lucas details the very components of the market-specific prices at time t:

$$p_{i,t} = p_t + z_{i,t} \qquad (3^*)$$

Accordingly, the market-specific prices at time t ($p_{i,t}$) are composed of an element common to all prices across the markets (p_t) and a component that varies randomly across

[62] For a philosophical discussion about social kinds, see for instance, Searle (1995), Khalidi (2015).

markets ($z_{i,t}$) but on average is 0. This is just another microeconomic assumption that reveals the internal components of market-specific prices. This is again another microeconomic assumption because it says the common component is an average across all the market-specific prices. But then Lucas makes his first inter-level assumption. He assumes that the real general level of prices, which I label as (P_{RGt}), is the *average* across the *N* markets of the *i* particular prices:

$$P_{RGt} = \frac{1}{N}\sum_{i=1}^{N} p_{i,t} \qquad (4^*)$$

This is the first bridge principle; a constitutive part-whole relation where the left-hand side of the equation is the whole while the right-hand side represents individual components – a collection of market-specific prices. In economics, the tag "[…] general level of prices is the generic name given to an average of prices of consumer goods and services relative to a reference period, and the main statistic constructed as a measure of this is the Consumer Price Index (CPI)" (O'Neill, et. al., 2017, p. 2). This first bridging principle pairing a microeconomic component (the collection of individual market-specific prices) to the macroeconomic component (the general price level) expresses the common ground assumption among economists about how to think about a useful notion of general price level.

The gap between micro and macroeconomics is easily channelled because *the concept of the general price is defined as a summary of microeconomic components*. This bridge is a *device* (Sargent, 1979, p. 379), an idealisation with no descriptive ambition. It's a definition, a rule used to permit the derivation between the macro principle and the set of micro principles. It's just a rule, is neither true nor false, employed to infer one statement from a set of statements (Nagel, 1970, p. 123). The average, like all the stages and equations in the model, does not express an empirical claim. As I earlier said, Lucas bypasses the aggregation problems. When we substitute this first bridge principle (equation 4*) into (3*), we obtain a derivative principle:

$$P_{RGt} = \frac{1}{N}\sum_{i=1}^{N} p_{i,t} = p_t + \frac{1}{N}\sum_{i=1}^{N} z_{i,t} \qquad (5^*)$$

According to the derivative principle introduced in the equation (5*), we are led to conclude that the real general price level (P_{RGt}) is equivalent to the common component of the

market-specific prices across all the markets (p_t). Recall that for an *N* that is large enough, Lucas assumes the second term will be small – again, another purely microeconomic assumption:

$$\frac{1}{N}\sum_{i=1}^{N} z_{i,t} = 0 \qquad (6^*)$$

Assuming (6*) and given (5*), we therefore derive another bridge principle:

$$P_{RGt} = p_t \qquad (7^*)$$

Once again, this is another part-whole relation connecting micro-and-macroeconomics – the real general price level (a macro quantity) to the common component to all individual market prices (micro quantity). Next, Lucas argues that suppliers in each market supposedly determine their supplying decisions according to their expectations ($p^e_{i,t}$):

$$p^e_{i,t} = \mathbb{E}((p_t \mid p_{i,t}, I_{t-1}) \qquad (8^*)$$

Accordingly, suppliers base their supply decisions on their estimate of (p_t) – i.e., ($p^e_{i,t}$) based on current market-specific prices ($p_{i,t}$) and in past information (I_{t-1}). In other words, people form their expectations about the common component of the market-specific prices (p_t) – which we know from equation (7*) to equal the real general price level – is conditioned to ($p_{i,t}$) and past information (I_{t-1}), i.e., $\mathbb{E} = (p_t \mid p_{i,t}, I_{t-1})$. Past information ($I_{t-1}$) is mainly formed by agents' *perception* of previous inflation (e.g., how people perceive past-one year inflation, or past five-to-ten years perception about inflation).

(8*) illustrates that individuals are tied to the macro level as a matter of fact: they need to use aggregate concepts as a means of dealing with the complexity of the environments in which they make decisions (Hoover, 2009, p. 393). This entrenched connection between micro-and-macroeconomics does not undermine the conceptual separability between micro-and-macroeconomics.

Like many other surrounding aspects in the world that captivate our attention (e.g., merchandising), inputs from macroeconomics (perception and information about past trends of inflation) and expectations about their future states may surely enter in the calculus of economic agents as a constraining feature of the microeconomic world, whereas they convey relevant information for useful purposes for individuals. This fact should not surprise us once

people do not make decisions in a vacuum (Lucas, 1980, p. 710-711). So, macroeconomic inputs may enter the micro level as passive constraining features upon individuals, although agents continue to be seen as the dominant causal entities in societies.[63]

The point of reduction is to demonstrate, in Nagelian terms, how the Phillips relation, a macroeconomic regularity, logically follows from an optimisation problem undertaken by individuals under a set of constraints: individuals are the causal gears while other potential features inside the micro are the contextual constraints. For this reason, I qualify the equation (8*) as a *hybrid* microeconomic assumption, as long as the macro enters the micro-level as informational inputs that constrain and support people's decisions. Even though macro input is significant for individuals' calculation, they manifest their influence as a passive constraining element.

Calculating the estimation of (p_t) from the assumption about the distribution of (p_t) and $(z_{i,t})$ yields

$$p_{i,t}^{e} = \mathbb{E}(p_t \mid p_{i,t}, I_{t-1}) = (1 - \theta)p_{i,t} + \theta\bar{p}_t \qquad (9^*)$$

Likewise, (9*) is another hybrid micro-macro relation where people's perceptions and expectations cannot be dissociated from the macro level. Recall that agents' expectations are formed by possible values that the general price level (p_t) may assume conditioned the past information about the market-specific prices. Herein the perception and the collection of previous inflation enter as an objective constraint of economic agents. Their forward-looking expectations are mainly guided by previous (a) macroeconomic events due to their perceptions of what happened, (b) actual specific-market prices, and (c) past information. This micro-macro transition expresses a theoretical assumption rather than a description. The extension of the referents on the right and left-hand sides of the equation are non-significant. If we substitute (9*) into (2*), we can derive shows the cyclical component of the individual suppliers:

$$y_{i,t}^{c} = \theta\beta(p_{i,t} - \bar{p}_t) \qquad (10^*)$$

[63] The difference between the applicability of reduction and global supervenience to economics stands on what is the role those other microeconomic entities may have. While reduction accepts other entities but individuals inside its domain, it denies they may have as much causal power as individuals. Global supervenience, in turn, allows that both individuals and other entities of the micro world may have equivalent causal powers to determine the macro. I address supervenience in the next chapter.

(10*) is the third hybrid assumption insofar as it reveals that the quantity supplied is a function of observed market-specific prices plus the average of their common component – which is now a macro quantity that expresses the average of the general price level. Assuming that the aggregate output can be represented analogously to the output in individual markets, we then represent it as follows:

$$Y_t = Y_{p,t} + Y_{c,t} \qquad (11^*)$$

Next, Lucas introduces his second bridge principle:

$$Y_{c,t} = \frac{1}{N}\sum_{i=1}^{N} y_{i,t}^{c} \qquad (12^*)$$

Assuming that firms are all alike (on average), the cyclical component of the aggregate supply corresponds to the cyclical component of the individual supply. This is another bridge principle. The aggregate cyclical component is a simple sum of representative agent' (uniform) decisions. Syntactically, equation (12*) expresses an identity, but its semantic content doesn't express the same sort of synthetic statements imagined by Schaffner as a necessary condition for satisfactory reductions – it doesn't have the same ontic status as an identity (*a posteriori*) claim such as "Water = H_2O". Lucas, remember, wasn't interested in aggregation issues.

Assuming again that *N* is large and using the law of large numbers, we can derive that the cyclical component aggregate supply ($Y_{c,t}$) is

$$Y_{c,t} = \theta\beta(p_t - \bar{p_t}) \qquad (13^*)$$

By rescuing the permanent component from (1*) assuming that firms are alike, we therefore obtain the aggregate output:

$$Y_t = Y_{p,t} + \theta\beta\,(p_t - \bar{p_t}) \qquad (14^*)$$

We get the output-inflation regularity via the auxiliary of two interlevel (bridge principles) assumptions that turn Lucas's derivation possible. The conclusion logically follows from microeconomic assumptions. What Lucas argues is that any economy, under those background conditions, will present a stable relation between inflation and output:

(1*, 2*, 3*, 6*) – Microeconomic assumptions (Pure microeconomic reducing statements)
(8*, 9*, 10*) - Hybrid assumptions
(4*, 12*) – Theoretical bridge principles
(5*, 7*, 13*) – Derivative bridge principles
(11*) – Macro assumption

Conclusion: (14*) – reduced (hybrid) macroeconomic output-inflation trade-off

Recapitulating: what is reduction in Nagelian terms? It consists in a special type of inter-level explanation in which a discipline, through its axioms and principles, explains a law (or a set of regularities) that belongs to a science that concentrates its investigations on a set of kinds and laws that amounts to higher-level of organisation. When scientists transfer the explanans from one level of inquiry to another one showing that the manifestation of a given regularity on a higher-level science is inferred from the lower-level science, we have reduction according to Nagel's view. From these set of assumptions, it's the hybrid and the bridge principles statements that potentially level scepticism about the viability of reduction. I don't think these statements can unsettle reduction.

Hybrid assumptions (8*, 9*, 10*) amount to the mixed character of micro-and-macroeconomics. However, they demonstrate that, at least in this fictional world, the outer elements – micro and macro informational constraints inside the micro-level – have a passive behaviour. These features are the objects of inquiry of suppliers. It's suppliers who bet, form expectations, and articulate their background knowledge with the available means of information at their disposal to enhance their decisions.

Bridge principles (4*, 12*), in their turn, are in service of inter-level explanation. They claim that the microquantities involved are mere aggregations of macro-quantities. Following these principles, combined with former statements and purely microeconomic principles, Lucas infers the macroeconomic trade-off. He explains how the Phillips curve is possible, according to these compilations of equations, using theoretical bridge-principles to deduce the manifestation of a macroeconomic regularity. To explain and understand the possible rise of the Phillips curve, the method of explanation is deduction: "we know what does happen in one these economies because we know what must happen given our general principles and the characteristics of the economy" (Cartwright, 2007, p. 218).

Besides the formal derivation, recall that in the first chapter, I listed five epistemic virtues that, according to Nagel, should be present in every positive example of reduction. It is apparent that these are mostly satisfied in Lucas's derivation. Start with (1) and (2): (1) the regularities from the reducing science cannot be *ad hoc* assumptions and should have some empirical support; (2) reducing law statements should be attained independently from the reduced level. Notice that the Lucas model uses only standard already widely used assumptions and also uses only already independently well-established micro principles. The empirical support, by the very nature of such model, isn't offered and therefore this informal condition is mismatched.

Now look at (3) and (5): (3) reductions should enhance scientific knowledge of the reduced domain; (5) fruitful reductions should offer better predictions. Both of these are accomplished by Lucas showing the dependence of the Phillips curve on what suppliers expect and in particular on whether they mistake inflation for a real rise in price in their domain. Finally, (4) good reductions unfold new patterns of dependence among the target laws from reducing and reduced domains that unify the explanations. The target laws in the macro domain are all the equations linking supply and price. The derivation shows the dependence on agents' expectations which in turn are affected by the accessibility of knowledge about real facts, like real price rises versus system-wide inflation. Lucas, therefore, reduces the Phillips curve in a suitable way to satisfy Nagel's demands at theoretical level. More work needs to be done to provide an empirical ground of such reduction.

5. The cogs and wheels behind the Phillips curve

In chapter 2, I said the mechanistic-based explanation is applied case-by-case. Despite the specificities of mechanistic explanations across various sciences, standard mechanistic analysis shares the following common ground approach: the target of interest is comprehended according to its internal functions, early stages, and processes that contribute to the manifestation of the most visible function under assessment. In virtue of their own approach that seeks the internal components that contribute to what is observed at the surface, mechanistic-based approaches can effortlessly intertwine different disciplines: "What is found in biology is *mechanisms*, mechanisms built with chemical components and that are often modified by other, later mechanisms added to the earlier ones" (Crick, 1989, p. 138).

The type of mechanism offered by Lucas represents an ideal picture, a caricature of what a social mechanism looks like. By "social mechanism", I follow Hernes (1998, p. 95) who defines

it as "a device for combining actors with a given set of characteristics ("casting") with a particular social structure ("staging") in order to infer what outcomes will result ("plotting")".

As we saw in the previous chapter, a recurrent debate among social scientists and economists is which social units within the social structure – institutions, norms, culture – are causally efficacious and can, or cannot, perform the explanans role in economic modelling. On Lucas's model, individuals are the dominant social units responsible for determining macrosocial phenomena. Lucas tells us a story of a social mechanism in which the most visible phenomenon at the surface, the Phillips relation, stems from fundamental active components at the bottom line (microeconomics): agents with their intentions, beliefs, and expectations, limited by a set of constraints in their environment, determine the macroeconomic result. Analogously to the biological mechanism suggested by Crick, which, by its nature, can bridge biological and chemical properties, Lucas's socio-economic machinery bridges the two economic layers – micro and macroeconomics.

We already saw that agents' misinformation guarantees the rise of the Phillips curve. There are two main items in the architecture of this socio-economic mechanism: (a) agents and (b) social setting. (a) depends on the properties of those agents, and (b) is the economical space where agents manifest their properties. For Lucas, market-specific prices are the most relevant information available in (b), but other economic signals and sources of data are also appropriate (e.g., information about past inflation). Other existing elements within this setting (institutions, culture, norms) are non-relevant; they exist but have been abstracted away due to their non-significance for Lucas's target of interest. But what are the relevant agents' properties? Hernes (1998) plots four questions which, by their answers, we can track agents' properties of this social mechanism:

(1) *What do they want*?
(2) *What do they know*?
(3) *What can they do*?
(4) *What are their attributes*?

Let's revisit Lucas's story to respond to each of these questions. In his closed-economic system, the environment is fundamentally made of N number of suppliers supplying the same good i selling at the market-specific prices ($p_{i,t}$). These social entities, N suppliers and this mind-dependent social entity (market-specific prices), are the two primary elements of this system. In this economic system where the suppliers are all alike – i.e., they rationally respond

to the same sorts of incentives – they base their decisions about quantities (the number of goods that should be supplied according to the context) on market-specific prices. Each supplier has a normal (or permanent) level of production ($y_{i,t}^{p}$), that is, a quantity supplied of goods insensitive to external or novel environmental conditions.

Suppliers also pay attention to their context: to make a profit or alleviate losses, they optimise their decisions according to the variations in the price of their goods ($p_{i,t}$). This is why typical firms who look after profit have a varying component of supply ($y_{i,t}^{c}$): the quantity supplied is sensitive to changes in their environment. Notice that, as rational, forward-looking agents, whichever decision suppliers will take today (to expand or shorten their production) also depends on their expectations about the possible values that their goods might assume. As Carter and Maddock (1984) stress, "since virtually all economic decisions involve taking actions now for uncertain rewards in the future, expectations of the future are crucial for decision making".

It is also worth noting that, from the very start, suppliers are tied to macroeconomics: they use the money issued by the government, they know the face value of market-specific prices is contaminated by the amount of money circulating in the economy, and they pay wages for their workers in money terms. So, it is natural that, as rational agents, suppliers form expectations about the future states of the macro (e.g., the value of general price level) and use macro data to enhance their decisions, reducing thus the risk of misallocations.

In this regard, suppliers are desperate to know how much of the face value of the goods they observe at *t* represents demand conditions (households' preferences) and which portion corresponds to nominal shocks that affect all the prices in the economy (inflation). Suppliers would always maximise their production level if they could perfectly discriminate demand shocks from nominal (money) shocks and anticipate, for instance, the future rate of inflation. But they cannot; information in this world (like ours) is heavily asymmetrical.

For this reason, suppliers call upon other accessible pieces of information from their environment to optimise their decisions and form their expectations about the future status of the general price level: information about (i) past states/values of the general price level, (ii) real general price level at t (P_{RGt}), which together are major components of (I_{t-1}), and (iii) market-specific prices, which, accordingly, is made of a common component to all prices (p_t) and a random component. Among those, (i) – (ii) are macroeconomic sources that enter the microeconomic level as informational constraining items.

Lucas assumes that the real general price level (P_{RGt}) corresponds to the average of all market-specific prices at t ($P_{RGt} = \frac{1}{N}\sum_{i=1}^{N} p_{i,t}$). We saw (equation 7) that the real general level

contains, i.e., equal in value to the common component to all market-specific prices (p_t), the component that suppliers are desperate to know and try to forecast. At this point, I believe we can answer the questions (1) – (4):

(1) *What do suppliers want*? Money. They intend to maximise their profits and diminish losses as much as possible. This orientation of suppliers towards profit is one of the background axioms – the item (g) of the model (see it in section 2).

(2) *What do suppliers know*? They know their profitability depends on demand conditions, which aren't a priori given due to incomplete information – a presupposed assumption of the model represented in item (b), and that the price of their goods is contaminated by a nominal component – there's only one type of shock in the economy, which is monetary, as stated in item (a) among the mains assumptions assumed by the model. Thus, as vehicles of information, relative prices aren't the exact mirrors of demand conditions.

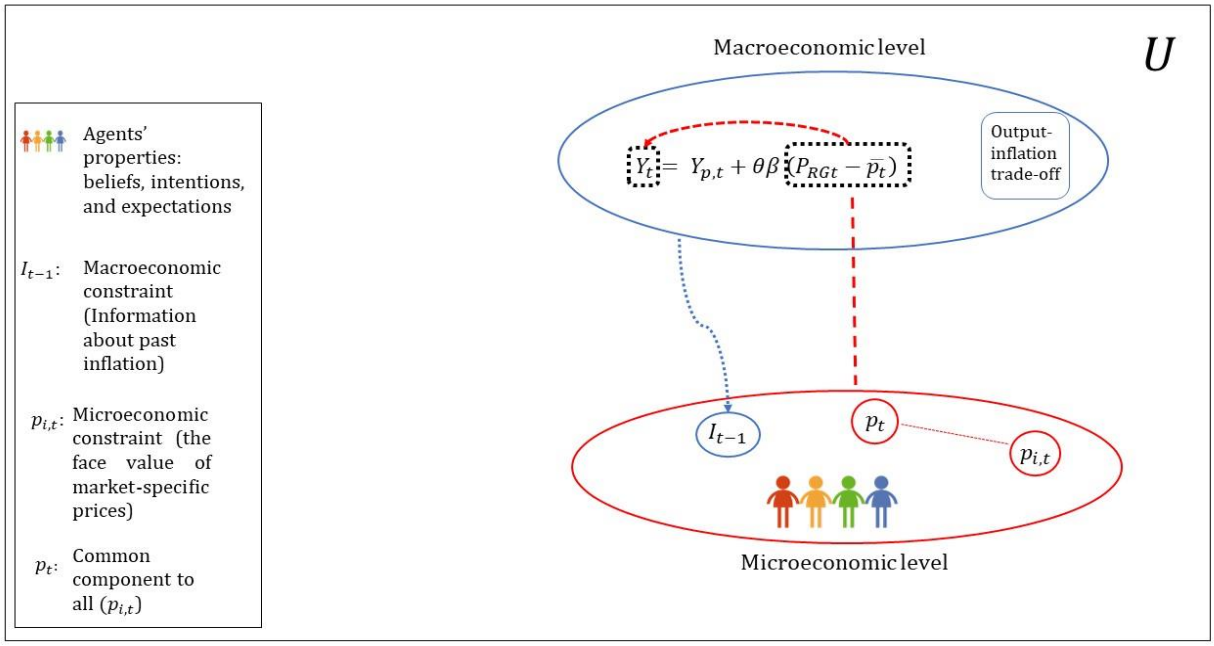
(3) *What can suppliers do*? Typical suppliers have a normal (permanent) and variant level of production (equation 1), where the variant component ($y_{i,t}^{c}$) is supposed to be a linear function of the deviation between the actual market-specific prices and the perceived common component of the general price by producers in market *i* (equation 2). It is assumed that the actual market-specific prices ($p_{i,t}$) are composed of two components, (p_t) and ($z_{i,t}$), where the former is a variable common to all markets and the latter is a random disturbance – normally distributed, with mean 0 and variance τ^2, independent from the distribution of (p_t) – representing variations in relative prices across time and markets. Based on past information (I_{t-1}), market suppliers suppose that the distribution of (p_t) is normal with mean ($\overline{p}_t$) and a variance σ^2. The two components of ($p_{i,t}$) cannot be observed separately. So, to accomplish their goals (maximise their profits herein represented in answer 1), suppliers can rationally expand production when they observe the actual face value of their goods at *t*, in comparison with t-1, goes up and revise the level of production in the immediate future in case of the adding points in production does not convert in sales. Moreover, suppliers can anticipate the value of the general price level, forming their expectations according to the variables represented in equation 8. When they anticipate the rate of inflation, they will not produce more than is demanded by their consumers.

(4) *What are their attributes*? Suppliers are rational agents who try to maximise goals under every set of constraints using all possible information and act according to

economic incentives only. Their maximising attitudes are constrained by the given rules of the model already explored. But only some of the agents' attributes will, in the end, constitute the macro-phenomenon to be explained – the relation of inflation and output that gives rise to the Phillips curve. These attributes are the price that each charges and the amount of output that each supply, which under the bridge principles described aggregate into overall output and overall price.

In their continuous pursuit of wealth, suppliers are always trying to optimise their decisions according to the observable signals of the environment. That is the dynamic component of this social system, although there are two features that remain unaltered: preferences and technology. The trade-off between inflation and output emerges from suppliers' misperceptions about the possible values of the general price level. When suppliers, trapped by the signal extraction problem, base their quantities' decisions on past information about inflation, the actual rate of inflation, and market-specific prices at *t*, they inadvertently produce more or less than is desired by their customers. In doing so, they generate an asymmetrical relation between the general price level (P_{RGt}) and output (Y_t) as represented below:

Figure 2: Representing the constitution of the Phillips curve



Mechanisms can be tackled in a piecemeal fashion. When we look only to the top- level, we perceive an asymmetrical relation between [$\theta\beta\ (P_{RGt} - \bar{p}_t)$] and ($Y_t$). Both are posited at

the same level (excepting the parameters grounded on micro). Under the interventional account of causation, an ideal intervention I on $[\theta\beta\ (P_{RGt} - \bar{p_t})]$ should modify the aggregate output (Y_t) in order to be qualified as causal. However, some of the components of the explanans are not causally autonomous. The cyclical component of aggregate supply is a derivative instead of a lower-level entity – it results from suppliers' intentions, beliefs, and expectations under a set of constraints.

The actual explanans of the Phillips relation is located at the bottom line. It is composed of individual's properties, which, by their features, determine the relations between the real general price level and output. The trade-off between inflation-output is the explanandum, while the properties of individuals correspond to the explanans. All the material (or non-individual) elements inside the micro structure, micro-macro data and the face value of market price tags are passive pieces of information that limit the range of options for suppliers' decisions, beliefs, and expectations.

The tags (constraints) across the environment have a passive role in determining the macro. They are the objects of inquiry of economic agents. The properties of individuals prevail – these are the actual causal gears capable of generating a disequilibrium on the macro level between the general price level and output. The suppliers respond to incentives, perceive prices, have beliefs over the economy's causal structure, and hold expectations – they are the *dominant* causal agents in this system, although not the exclusive ones (Manicas, 2006, p. 75). Thus, in the composition of causes, in the combination of individuals under a set of constraints, it's the agents' properties that hold a causal relevant effect for triggering the Phillips curve. Therefore, we have reduced the Phillips curve in mechanistic terms so long as the properties of individuals prevail over other properties from the microeconomic level.

Contrary to this interpretation, Hoover alleges that the irreducibility of the general price level (P_{RGt}) turns reduction impossible. A better interpretation should accept a non-reductive reading where the output-inflation trade-off supervenes on the micro-set – the combination of individuals' properties plus the set of micro-macro constraints within the microeconomic level that generated the outcome. Hoover believes that once the macro-level is ontologically mixed with micro-level, it turns a reductive explanation into an unbearable theoretical enterprise. In summary, the manifestation of the trade-off doesn't exclusively stem from people's properties.[64]

I don't think these remarks are challenging for reduction. The dispute between reduction v. supervenience in economics should not be argued in terms of "take it or leave it". It should be assessed in a case-by-case analysis according to the explanatory purposes of interest. Fixing the

[64] I address Hoover's defence in favour of supervenience in the next chapter.

ontological problem prior to the process of model construction is an option instead of a necessary condition for economists provide their explanations (Ruiz, 2021). Moreover, in the Lucas's model, all the elements plotted by Hoover as the potential ontological barriers for reduction – the irreducibility of the general price level and the contamination of macroeconomics in the micro domain – do not display any significant causal influence on the explanandum. None of these entities figure in the explanans. As far as they are non-relevant for Lucasian purposes, they can be ignored:

> In the cases we are concerned with, several familiar properties are deliberately ignored in the theorizing that takes at the social level. Some of these properties may pertain to individuals, and some may well be materialistic. But this does not mean that they are all on a par: although some of them play a key role in the constitution of *S*-level causal relations, others are likely to be completely irrelevant. (Guala, 2022, p. 7)

Look back at the bridge principles from the model for the quantities that appear in the output-inflation equation. These quantities are mere aggregations of features of agents at the micro-level, namely the amount they supply and the price they observe. So the macro relation is entirely constituted by relations of micro features. In Lucas's world, only the individual's properties are the genuine realizers of the Phillips relation, even though they are not the unique entities inside the micro level.

As I see it, the difference between reduction and supervenience is better accounted for in the following terms: sometimes, when other properties within the micro-level (e.g., macroeconomic inputs) have the same causal power as individuals to determine a given macro phenomenon, then we may have supervenience.[65] But this is not the case here, where the explanandum is fundamentally explained by the properties of individuals. None of the outer elements from the macro-level and acknowledging the very existence of macroeconomic entities makes reduction impossible. It's just a matter of explanatory interests and which features are causally important for the rise of the explanandum, which are topics I explore in more detail in the last chapter.

5.1. The mechanistic reduction of the Phillips curve: a counterfactual approach

In the former chapter, I defended a monist criterion for assessing causal and constitutive explanations. To be successful, both explanations should be capable of adequately responding

[65] In the next chapter, I offer an argument in favour of supervenience where the outer elements within the microeconomic level have the same causal relevance as the properties of individuals. In my response to Epstein (2014) criticisms against supervenience, I therein argue that institutional properties, one of possible outer elements inside the microeconomic domain, may have active constraining effects to determine a given macroeconomic outcome.

to our *what if*? queries. A constitutive explanation must answer what may possibly happen if the target phenomenon has some relevant structural components (parts) inactive or (ideally, i.e., in theoretical contexts) modified. Therein, I argue that explanandum *E*, placed at the higher level of organisation, obtains its properties by virtue of the properties of the selecting structural items placed at a lower level or organisation (explanans). By adopting a (counterfactual) monist approach, explanandum *E* should be counterfactually dependent on the explanans (regardless of the relata – i.e., whether they are events or causal powers/dispositions): different values in the explanans features must change the features of the explanandum.

As an example, I illustrated a positive case where an economic system, to hold the property *S being capitalist*, it must contain (at least) two lower-level (individual) components of any capitalist system: (1) *being a labour force* and (2) *being a capitalist* (i.e., means of production primarily concentrated at the hands of the private sector). If none of the units are present in *S*, the individual system will not bear the property *being capitalist*. Mechanistic-constitutive reduction is obtained by showing the counterfactual dependence between the relevant parts (explanans) and the whole (explanandum) – the system has the property of *being capitalist* due to the selecting structural features. So, in (fictional) theoretical contexts, constitutive explanations can exhibit the invariance-under interventions criterion adopted for causal explanations: if the causal features of *Y* are due to the causal powers of its components *X*'s, then if we could change *X* in some ways, *Y* would also be modified.

Now to the counterfactual. Look again at equation 15:

$$Y_t = Y_{p,t} + \theta\beta\ (P_{RGt} - \overline{p_t}) \qquad (15)$$

According to this equation, production will be greater than the ongoing level ($Y_{p,t}$) if, *but only if*, (P_{RGt}) > $\overline{(p_t)}$. Recall that $\overline{(p_t)}$ = anticipated average of the price level (p_t) under the distribution presumed by suppliers. So, if the average that is anticipated is the actual general price level, there will be no increase in supply and hence no Phillips curve. Changing $\overline{(p_t)}$ – i.e., what agents at the microlevel expect – and you affect whether or not the Phillips curve obtains. This is just what Lucas was at pains to show: if the government tries to use inflation as a lever to affect output, it will be hard to conceal that, so suppliers' expectations will be different from when inflation occurs 'naturally' and is thus for a short while at least not recognised as such. The expectations are different in the two cases, and that results in a difference between the

Phillips curve obtaining and not obtaining. This is just the kind of counterfactual dependence we are looking for.[66]

Lucas's model is an example of reductive explanation as far as we articulate an objective counterfactual dependence of the explanandum, placed at a higher-level of economic organisation, from the explanans located at the microeconomic level where the relata *is only performed by the properties of individuals*. In this model, only the properties of economic agents are explanatorily relevant to the target of interest. Models *a la* Lucas can be seen as devices for theoretical experimentation where modellers abstract away some parts of the world and articulate others they believe are crucial for their interest targets. Once they state what the assumptions are and fix the relevant dependencies inside the model, modellers can manipulate the variables within the model and ask various *what-if*-questions. Our capacity to explain, and understand, a given economic phenomenon, depends on our ability to make several what-if inferences:

> The fundamental criterion of understanding is the ability to make inferences about counterfactual situations, e.g., the ability to answer contrastive *what-if* questions by relating possible values of the *explanans* variables to possible values of the *explanandum* variable. Explanations aim to track objective relations of dependence. These dependencies are modal: explanation is not about subsumption about regularities, but about counterfactual dependence. (Ylikoski and Aydinonat, 2014, p. 29)

The Lucasian model represents a positive case of reduction according to Woodward's paradigm, exhibiting a counterfactual dependence between explanandum (the Phillips curve) and explanans, the individual's properties – stable parameters, beliefs, and expectations. In this case, the objective relation of dependence represents a mechanistic reduction.

It's worth noticing that this mechanistic reduction does not entail the *elimination* of macroeconomics. This case holds similar trends to the Weberian hypothesis for the rise of capitalism explored in the former chapter. Herein, the monetary authorities can certainly affect the general price level *if and only if* agents do not anticipate the amount of money issued by the central bank. However, the money supply affects the general price level indirectly; inflation is not a putative entity that authorities can exploit it directly (Blanchard et. al., 2017, p. 485). It's people under a set of constraints that may determine the Phillips curve; the money supply causally determines the Phillips relation only indirectly. Thus, macro entities are causally efficacious, although they manifest their effects indirectly.

[66] These counterfactual differences reappear in chapter 5, when I tackle the topic of explanation using the contrastive approach.

6. Concluding remarks

In this chapter, I presented one (theoretical) economic model that permits a reductive interpretation. It illustrates a theoretical inter-level explanation involving micro-and-macroeconomics, where the explanans, placed at the bottom level of organisation (micro-set), explains the possible rise of the macro phenomenon at the top (explanandum). The model has a reductive character as far as the explanans, located at the microeconomic level, does not claim any outer elements within the micro-set to explain the explanandum. The constraining elements inside the micro-set are the objects of inquiry of economic agents; they are passive tags, pieces of information available in the surroundings of individuals. These elements, on their own, do not play any direct effect on the macroeconomic outcome. It's individuals who value these informational inputs on their behalf, combine them with their background knowledge, and use them to form their particular expectations and from that set the microquantities that will aggregate to yield the macro phenomenon.

Philosophical debates over reduction versus non-reductionistic approaches such as global supervenience in economics are better accounted for in terms of the relevance of the entities that perform the explanans. Questions about reduction are questions about what we can explain, as opposed to other metaphysical notions that do not necessarily provide explanations (Klein, 2009, p. 53).[67] In some contexts, reduction performs a partial explanation of a given phenomenon. This message is explicit in Weber's thesis, widely explored in the previous chapter: as a historical event, the rise of capitalism counterfactually depended on material or non-individualistic events for its rising. With respect to the social property of being capitalist, that new economic system counterfactually depends on its individual properties (its components). But in the Lucas case, the reduction of the Phillips curve offers a complete explanation for the explanandum. Even though the explanation is not well-confirmed, the model is firmly grounded in real economic capabilities, offering a plausible approximate description about the possible causes of the correlation between output and inflation. When we look this way, it's clear that a case-by-case basis of the analysis is a much better approach to judging whether there's a reduction or non-reduction example of an economic phenomenon before us. It depends on our goals, explanatory interests, and which elements play the explanatory role for the target of interest.

[67] In the next chapter I present a positive example of supervenience applied to economics that does not lead to explanation. For this case in particular, supervenience is just a modal notion without entailing explanation.

CHAPTER 4: SUPERVENIENCE, LOCAL AND GLOBAL

1. Introduction

In this chapter, I discuss an alternative relation of dependence that has promised to represent how the micro-and-macroeconomics might be channelled: *supervenience*. The exact origins of the concept aren't entirely clear, although the intellectual motivation behind its birth is quite understandable. In the early 20th century, philosophers raised queries about the nature of the relationship linking certain types of properties, such as those attributed in moral and aesthetic judgements, to descriptive properties. The peculiar status of the connection between the aesthetical properties of a painting to its corresponding physical properties would require an extraordinary form of dependence to represent it. To fill that gap, the concept of supervenience was born.

Many versions of supervenience have been formulated since its inception.[68] They vary in modal force (weak, intermediate, or strong), scope (individuals or entire worlds), and types of relata under supervenience's coverage.[69] Despite the plurality of views, all of them normally share what characterises supervenience mostly: a dependence relation that says once a set of micro-level *B*-properties have been fixed, the macro-level *A*-properties are also set.

Harold Kincaid (1998, p. 487-488) and Kevin Hoover (2001a, p. 120; 2001b, p. 73) have alleged that micro and macroeconomics are tangled in a supervenient way. Apparently, there's no viable pathway to eliminate or reduce macroeconomics. First, macroeconomics is essentially entrenched in the microeconomic domain. Further, macro entities may also respond positively under interventions, a response that depends on individuals' behaviour to influence other macro entities. This indicates a difficulty in construing every macro entity as a mere epiphenomenon. Alternatively, supervenience may offer refuge to capture an ontic bridge between micro-and-macroeconomics.

In economics, supervenience has a non-reductive interpretation, being employed to describe the dependence between micro and macroeconomics as if they were (conceptually) two distinct levels. Nothing special or mysterious characterises microeconomics and macroeconomics. Hoover defines microeconomics as "the economics of the behaviour of individual economic agents" while macroeconomics is "the economy as a whole, focusing on the

[68] A summary about the topic is McLaughin and Bennett (2018).
[69] Usually, the supervenience' relata are facts or properties. The concept is not normally employed to relate representational items such as theories or their fragments. Butterfield (2011) has applied the concept to relate different scales of physical theories.

behaviour of the economic of aggregates, such as gross domestic product (GDP), employment, unemployment, inflation, and interest rates" (Hoover, 2010, p. 330). Supervenience is claimed as the bridge between this great whole to microeconomic facts:

> Most obviously, supervenience would seem to describe the relation between macro-level and micro-level explanations in economics. For example, if all the microeconomic facts are set, so too are the macroeconomic, or if all the facts about individual behaviour are set, so too those about the corporate behaviour. But such dependence does not entail reduction. (Kincaid, 1998, 487-488)

The version of supervenience that seems best to accommodate the vertical determination in economics is *global* supervenience, which claims that *A*-properties globally supervene on *B*-properties *iff* for any two worlds, if these two worlds have the same *B*-property pattern of distribution, then they also hold the same pattern of *A*-property distribution. If, for example, the universe of microeconomics consists of the economic behaviour of individuals, then two worlds with the same facts about individuals ought to share indiscernible macroeconomic results.

Julian Reiss (2004) and Brian Epstein (2014) mistrust supervenience's capacity to depict the right connection between micro and macroeconomics. Reiss (2004, p. 232) argues that macroeconomic facts causally affect microeconomic ones, undermining the applicability of supervenience in economics. Epstein, in his turn, claims that macroeconomic facts do not globally supervene on microeconomic facts: It is logically possible to have two identical microeconomic worlds that do not share the same macroeconomic results. In fact, "[...] there is no complete microeconomic explanation of all macroeconomic phenomena since the microeconomic properties do not even determine macroeconomic properties" (Epstein, 2014, p. 15).

More recently, Peter Galbács (2020) contests Epstein's conclusions. In his view, Epstein's challenge towards supervenience only stands if we rule out the entire institutional setting from the microeconomic setting where individual behaviour occurs. When we re-insert the socio-economic structure into the microeconomic set, we reset a bottom-top global supervenience determination, showing that Epstein's argument is off the mark (Galbács, 2020, p. 188).

In this fourth chapter, I respond to Reiss and Epstein's criticisms. I first approach Reiss's criticism, arguing that the downward causation does not constitute a fatal blow to global supervenience necessarily. We may have cases where a macro set_1 supervenes on micro set_1 and causally affects another micro set_2. By exploring one example, the application of macroeconomic Taylor's principle, I demonstrate a paradigmatic case where there is no competition between downward causation and supervenience.

Second, in response to Epstein's argument of the autonomy of macroeconomics, I follow Galbács's intuition: macroeconomic properties are fixed by a combination of institutional and people's properties. A sharp separation between microeconomic properties and institutional features misrepresents the socio-economic setting where economic behaviour occurs and consequently engenders the macroeconomic sets in the budget scheme scenarios proposed by Epstein.

The chapter is organised as follows. I first lay out the fundamental features of supervenience (Section 2). Moving forward, I introduce its variation – global supervenience (Sub-section 2.1). Then, I present Hoover's defence of global supervenience to relate facts across micro and macroeconomic sets (Section 3). Next, I present the two selected criticisms (Section 4): the downward causation (Sub-section 4.1) and the autonomy of macroeconomics (Sub-section 4.2). Then I sketch Galbács's criticism towards Epstein (Section 5). Finally, in the final section, I offer my response to this discussion (Section 6). I initiate tackling the downward causation problem, demonstrating its potential consistency with global supervenience (Sub-section 6.1). Next, I argue why the structure (the legal apparatus) dismissed from the micro level in Epstein's argument should be included in the microeconomic set (Sub-section 6.2).

2. What supervenience is

By supervenience, I follow Kim (1993), who lists six fundamental features that characterise supervenience:

(1) It is a synchronic relation;

(2) Covariance: once we fixed the micro-level *B*-properties, the macro-level *A*-properties would also be set;

(3) The two sets involved in the relation must be qualitatively distinct;

(4) It relates sets of properties that cannot be linked through biconditionals (we cannot state identities);

(5) Supervenience is non-symmetrical: sometimes supervenience is asymmetrical; others, symmetrical;

(6) Supervenience admits transitivity: if *A* supervenes on *B*, and *B* supervenes on *C*, then *A* supervenes on *C*.

Consider John Coltrane's *Alabama*. Maybe you, as well as I, adore Jazz, especially Coltrane as a horn performer. Probably you are inclined to ascribe plenty of praising descriptions to that song. Imagine an aesthetic property such as *being a sorrowful piece of music* associated with the theme. Now ask yourself: What is the nature of the relationship linking this aesthetic property to the non-aesthetic (structural) features of Coltrane's song, such as the slow tempo, lines, and scales? For some aestheticians, the non-aesthetic and aesthetic properties of songs overlap themselves in a supervenient way:

> [...] melody and rhythm supervene on acoustic events, there is no difference in melody or rhythm without some difference in a physical acoustic phenomenon; to make your music more beautiful, you must perform physical acoustic operations and thereby alter the acoustic subvenient bases of the musical properties you want to improve. (Ihara, 2020, p. 124)

Like two paintings, such as the Picasso's *Guernica* and Munch's *The scream* –the aesthetic differences of two different musical works like Mozart's *The Magic Flute* and Beethoven's *Moonlight Sonata* are related to their non-aesthetic attributes. The nature of this relationship is (1) synchronic: their non-aesthetic arrangements determine the aesthetic qualities of the song at the same time. The relation is synchronic inasmuch it vindicates a simultaneous determination between the two different qualities of properties. We cannot say that the aesthetic property is an effect produced by the non-aesthetic properties. Unlike causality, which is arguably temporarily asymmetric, non-aesthetic and aesthetic properties co-vary instantaneously. Supervenience also requires (2) covariance (correlation) between the sets of properties at issue. This is satisfied in this case: there cannot be a difference in the aesthetic properties without a difference in the non-aesthetic properties. Notice that the example encompasses two types of properties whose qualities are distinct (3).

Consider what's wrong with the alternative, that aesthetic and non-aesthetic base are identical. First, different media (vinyl, digital media, or CDs) may equally perform the song. The track can therefore be *multiply realisable*: different platforms reproducing the same tune with the same aesthetic property.[70] Second, that aesthetic property could be recognized in other sorrowful songs. Assuming that qualitatively identical things must share their properties, we may discard the relation between non-aesthetic/aesthetic properties as a matter of identity. The aesthetic may be based on distinct non-aesthetic properties and may be equally applied to other songs. Thus, those properties would not be adequately bridged via biconditionals (4). Further, the relation between non-aesthetic and aesthetic properties is also asymmetrical (5); the non-aesthetic fixes the aesthetic but not the reverse.

[70] In the philosophy of mind, the multiple realization thesis claims that one mental kind can be equally performed by distinct physical kinds. For an overview, see Bickle (2020).

For supervenience theorists such as Kivy (1980) and Davies (1994), aesthetic properties such as *sadness* and *melancholic* are not mere subjective projections from the audience towards the musical works: they are present *in* the music. These aesthetic properties are attributes of musical works that normally follow a specific structural pattern: downward melodies, slow tempo, restrained dynamics, minor tonalities, continuous flow in phrases, etc. The aesthetic of *turbulence*, in turn, requires another musical structure: unstable chords and harmonic progressions, dissonances, minor tonalities, unresolved cadences, high-level rhythm activity, staccato, and increasing speeds (Azcarate, 2017, p. 98). Sad songs move downwards and slowly, while happy tunes often proceed by leaps. Thus, the aesthetic properties ascribed to musical pieces are not mere imaginary objects or experiences; they are supposedly real. For realists about the reality of aesthetic properties in musical works such as Kivy (1987), it is the reality of aesthetic properties which allows different persons invariantly associate equivalent aesthetical values to the same song: there is something engrained in the music that displays in the audience's experience the common aesthetical experience.[71]

And finally, supervenience permits transitivity that span different levels of properties (6). In one sense – crudely poor, I would say – music is nothing but sounding waves instantiating various mathematical properties. If musical properties supervene on physical properties, and physical properties supervene on mathematical properties – assuming mathematics as an autonomous formal science and taking numbers in a realistic Platonic (or Pythagorean) sense – then musical properties would supervene on mathematical properties. Although music cannot be reduced to physics and therefore, to mathematics, supervenience may capture the structural features of music anchored in the real – i.e., the mathematical – bedrocks of the world.[72] Thus, we may say that for any

> Two objects (e.g., artworks) that differ aesthetically necessarily differ non-aesthetically. [i.e., there could not be two objects that were aesthetically different yet non – aesthetically identical.] [i.e., fixing the non-aesthetic properties of an object fixes its aesthetical properties.] (Levinson, 1984, p. 93)

Alternatively, imagine how moral properties relate to natural properties. Moral realists take moral judgements at face value. They claim moral properties do exist; they are not mere subjective judgements verbalized by individuals. According to Hare (1952) – when the notion

[71] Peter Kivy (1987) defends a Platonist view of musical works. In his view, musical works are abstract objects (universals, types, or kinds) identical to their sound structures. Their performances are particulars, tokens, or instances. The Platonist view of music is the dominant interpretation about the ontology of music, but there are other interpretations. For some idealists about the nature of music, such as Collingwood (1938), musical works are mental entities. I leave entirely open whether an idealist view about musical works could accommodate supervenience. For an overview, see Kania (2017).

[72] In *The Jazz of Physics*, the physicist and jazz musician Stephon Alexander (2016) argues that several symmetries that plays a significant role in quantum mechanics are musically represented in Coltrane's songs.

of supervenience gained its popularity – to say that *X* is good supervenes on *X*'s non-moral characteristics:

> […] let us take that the characteristic of "good" which has been called its supervenience. Suppose that we say "St. Francis was a good man". It is logically impossible to say this and to maintain at the same time that there might have been another man placed exactly in the same circumstances as St. Francis, and who behaved in exactly in the same way, but who differed from St. Francis in this respect only, that he was not a good man. (Hare, 1952, p. 145)

Which mechanism enables different kinds of properties to be glued across multiple levels (nomic equivalencies, or type-type identities) is irrelevant; supervenience is neutral about how these relations hold (Post, 1995, p. 74).

The examples above represent local cases of supervenience. They take intrinsic properties as the lower-level base properties that fix the supervening (higher-level) ones. The distinction between intrinsic/extrinsic properties that became the standard approach in metaphysics was put forth by David Lewis. As he puts it, an intrinsic property is a property "which things have in virtue of the way themselves are" in opposition to the extrinsic ones, which things have "in virtue of their relations or lack of relations to other things" (Lewis, 1986, p. 61). Our mass is an example of an intrinsic property. If you are (luckily) wealthy enough, like Jeff Bezos, to visit the moon, your mass is one of the things that you shall keep unaltered when you step on the moon's surface. On the other hand, you shall weigh differently: it does so due to the moon's gravity. Our body weight is an example of an extrinsic property: a property that is sensitive to its possessor's relations with the world. Other examples of intrinsic properties are *water's boiling point* – a property that is inherent to the sample – or *having hair of a certain length*. In contrast, *having a long-haired brother* or *being the tallest man* are examples of extrinsic properties.

One might suppose that the supervenient base for aesthetic and moral properties should contain only intrinsic properties, but amongst aestheticians, there are many disputes over the applicability of supervenience. Several aestheticians argue that the aesthetical properties of *Monalisa* also depend on its extrinsic properties: the period in which the painting was made and who painted it increases its value.[73] Similarly, well-educated persons in music or knowing the historical context and motivations behind Coltrane's composition may make people experience Coltrane's *Alabama* differently.[74] Therefore, the supervenient set of properties could not be made of intrinsic properties exclusively; extrinsic properties should also be included. A less

[73] See, for instance, Walton (1970).
[74] The song was inspired in the events occurred at a Baptist church, in Birmingham, Alabama, when four black girls were killed during a bomb attack leaded by the Ku-Klux-Klan in 1963. See, Leonard & Culture (2019).

restrictive weaker version of supervenience was shaped to permit this, the so-called global supervenience, outlined below.

2.1 Supervenience in its global version

Contrary to local supervenience, global supervenience was forged to obtain top-level determinations when dependence on just local intrinsic lower-level features could not be met.[75] A more permissible version of supervenience was then couched to include a broader scope of cases while preserving the essence of supervenience. The standard version of global supervenience is formulated as follows: "*A* globally supervene on *B* iff any two worlds with the same distribution of *B*-properties have the same distribution of *A*-properties as well" (Paull and Sider, 1992, p. 834).

Certain versions of physicalism are examples of global supervenience. These say that for every non-physical property that individuals who walk on the surface of Earth may have, physical conditions determine their non-physical properties.[76] If two sibling worlds are alike physically, then both worlds must share their non-physical conditions. In its global version, the supervenience jargon "there cannot be an *A*-difference without *B*-difference" relates to the set of macro, the supervening properties taking into account the entire possible base world. Base (micro) and supervening (macro) properties, the original relata of supervenience relations, continue to be attributes of individuals or objects. Entire worlds do not hold clusters of properties; instead, the individuals who inhabit these worlds have properties.

Accordingly, given a universe *U*, once *B*-property (a set of properties placed at the base) distribution has been set, the *A*-property distribution (the set of properties located at the top) is also fixed. Global supervenience relaxes the determinacy requirement of local versions in which the macro properties of an individual or object are determined by their micro properties. The global version says that given a universe of individuals *N*, the configuration of *B*-properties that characterise the individuals in *N* shall globally determine the distribution of *A*-properties of those individuals.

Consider the physicalist jargon that asserts that everything that exists is somehow physically determined. If any two worlds are physically identical in their property configurations (biological, chemical, and the like), they will end up being also equal in respect

[75] Like the notion of local supervenience, philosophers have also provided different notions of global supervenience with different modal status. McLaughin (1996), Stalnaker (1996) and Sider (1999) separate strong from weak global supervenience. Later, intermediate notions came out in literature. See, for instance, Shagrir (2002) and Bennett (2004).
[76] See, Stoljar (2021).

of their non-physical properties. That jargon supports *transitivity*: if *A* supervenes on *B*, and *B* supervenes on *C*, then *A* supervenes on *C*. To get determinations from the bottom to the top, we usually assume that the bridge across the extremes is filled by supervenient relations channelling the intermediate levels. In this regard, saying that the mental globally supervenes on the physical is true because biological properties globally supervene on chemical properties, and chemical properties also globally supervene on physical properties.

Let us say that the physical states of my brain determine my mental states. Rather than its local version that would say that the properties of my brain are the unique responsible features in determining my mental properties, global supervenience claims a broader supervenient basis:

> […] what mental properties I have is determined, not only by what physical properties I have, but also by what physical properties other things have, by what physical relations I bear to those objects, and by what physical conditions those objects bear to each other. Objects with the same physical world perspective must have the same mental world perspective. (Sider, 1999, p. 6)

More precisely, it's my neural states plus wider relations with other dispersed properties throughout the environment that are responsible for determining my mental states. Suppose my mental states do supervene on physical states. In that case, it does so by including inside the supervenient base other properties that shall be correlated with my mental states, a pattern of determination that should be replicated across every individual that inhabits the same environment. In a physically possible world where there is a *doppelganger* – an individual that is a perfect copy of my entire physical dispositions – we may welcome that we both share the same mental states – so long as the states of all other physical features in the world stay fixed. Then we may have global determination without local determinations (Kim, 1993, p. 85).

Several criticisms have been raised against global supervenience by philosophers. Nonetheless, for those like Hoover who defend supervenience, global supervenience is the right non-reductive version that can accommodate the connection between micro-and-macroeconomic properties. As I shall detail in the next section, global supervenience applied to economics demands a *global* bottom-top determination from micro towards macroeconomics, taking facts as their primary relata: once all the microeconomic facts are fixed, the macroeconomic facts have been settled.

3. Hoover's defence on global supervenience

Hoover construes supervenience as his anti-reductionist bridge between micro and macroeconomics. Hoover is concerned with *ontology*; he did not claim supervenience for

bridging some macroeconomic theory to some microeconomic theory (Hoover, 2001b, p. 122). One of the primary conditions to get supervenience is that both relata, supervening macro properties, and correspondent base (micro properties) are real. Therefore, the supervenience defender must commit herself to the reality of macroeconomic properties. Macroeconomic properties should be seen as real as the properties and behaviour of economic actors (i.e., microeconomic properties) are. According to Hoover, the evidence shows that macroeconomic properties cannot be eliminated.

Macroeconomy variables are often employed as inputs by economic agents. When, for example, people evaluate what economic decision to make, they form their expectations and take their actions based on many things, including information from macroeconomic variables. For instance, as an economic agent, my decision regarding the best time and place for travelling abroad may consider the actual exchange rate at the time. There are undoubtedly some periods of the year and destinations more affordable than others.

Moreover, evidence has also shown that macroeconomic aggregates are sensitive to manipulations: e.g., some macroeconomic variables under the direct control of central banks can affect the value of other macroeconomic entities. Take, for instance, the interest rate. The potential causal powers that one variable may have in keeping other macroeconomic aggregates under control may offer a reliable source to (indirectly) infer its existence.[77] We infer the existence of the real rate of interest in terms of its capacity to produce real effects upon the yield curve in the same way that physicists justify electron's existence:

> […] Consider two irreducibly macroeconomic aggregate entities: the real rate of interest […] and the yield curve […]. Both the real rate of interest and the yield curve are synthetic aggregates and both are entities with causal powers in some economic theories. Every macroeconomic theory that I know predicts that actions that increase the general price level or the federal funds rate will shift the yield curve upwards in the short run. And, at least if the changes are unanticipated, increases in the general price level will reduce the level of interest rate. The empirical evidence for these effects is overwhelming and indeed is easily confirmed by anyone willing to read the Wall Street Journal regularly for a month. Just like the electron, some macroeconomic aggregates can not only be controlled but can be used to manipulate other macroeconomic aggregates. (Hoover, 2001a, p. 125)

However, causal relations among macroeconomic variables need not be direct between two contiguous events. Supposing supervenience to hold, causal relations at the macroeconomic level are mediated by the microeconomic level: the real rate of interest works as a positive signal to economic agents, inducing their actions. In practice, it is people's attitudes that directly shift the yield curve. Nevertheless, although the causal relation between the

[77] Initially, Hoover incorporated the so-called Hacking's (1983) Entity realism. Accordingly, we are entitled to infer the existence of any putative entity as long as we use it to interfere over other entities.

interest rates and the yield curve is indirect, this does not exclude that the two macro entities are not counterfactually connected: different values of interest rates would entail a different yield curve.

Following Wilson (2021), we may take as the positive mark of the existence of putative entities their *causal powers*; whatever causes an entity can bring about is directly associated with its features. So, we can say that the real rate of interest is real inasmuch it can modify other macroeconomic entities. If the cause is real, then the effect (the yield curve) is also real. Hence, macroeconomic aggregates hold causal powers; they can affect how people drive their economic actions and modify the value of other macroeconomic entities – even though indirectly.

Claiming the reality of macroeconomic properties does not entail dualism. Macroeconomics is not orbiting in a free-floating stratum of reality. Each macroeconomic aggregate has its anchor in elements placed at the microeconomic scale. Hoover classifies macroeconomic aggregates into two groups. There are natural macroeconomic aggregates. These are those that add up their components into one single type of measurement or average. For instance, the total level of employment, a macroeconomic aggregate, is labelled as natural insofar it is measuring the same thing as its microeconomic components, employment or non-employment: aggregate employment measures the number of workers that are employed. There is a close analogy between the aggregate and the individual parts. Consequently, for these sorts of aggregates, we may say they are a summary of their parts.

The other group is the synthetic entities. They receive this label because they do not preserve the same degree of analogy that natural aggregates hold with their microeconomic components. As Hoover (2001a, 113) says, they are "[…] fabricated out of components in a way that alters the structure of the components". Real GDP or the general price level are examples of synthetic aggregates. Such aggregates are built from what is originally heterogenous. They aggregate dissimilar properties (different goods) into a common measurement (money). Still, each macro entity – assumed as real – is anchored in microeconomics. Consequently, a complete dualism – where macroeconomics is a completely divorced domain from microeconomics – is discarded. A middle range view conciliating the causal autonomy of macro entities with lower-level dependency is descriptively better, Hoover argues. For that reason, he claims that macroeconomics, globally speaking, is an *emergent* phenomenon:[78]

[78] Hoover himself did not offer details about emergence. On my view, Hoover bears in mind a weak notion of emergence. Wilson (2021) states two criteria for weak emergence. First, a token higher-level property *H* is weakly metaphysically emergent from token lower-level property *L* on a given occasion when (i) *H* and *L* are synchronically related and (ii) *H* has a non-empty proper subset of the token powers had by *L*. Condition (i) specifies synchronic dependence while (ii) captures a weak sense where an emergent entity *E* is causally

> […] macroeconomics could have an ontological anchor in the individual, while preserving an ontological independence for causally interacting aggregates. […] macroeconomics could be seen as emergent properties of the economy – ones that would not exist without the underlying microeconomic agents – but ones that, like mental properties emerging from physical properties – were ontologically distinct. (Hoover, 2009, p. 390)

On global supervenience, the macro is fixed when all the micro facts – past and present – are fixed, so all is well for global supervenience if the macro properties that have effects on micro properties are always fixed by (possibly earlier) micro properties. So, we may bet that when all microeconomic facts are fixed, the macroeconomic facts are also set:

> Macroeconomic aggregates supervene upon microeconomic reality. What this means is that even though macroeconomics cannot be reduced to microeconomics, if two parallel worlds possessed exactly the same configuration of microeconomic or individual economic elements, they would possess exactly the same configuration of macroeconomic elements. It is not the case, however, that the same configuration of macroeconomic elements implies the same configuration of microeconomic elements. (Hoover, 2001a, p. 120)

Even though specific macroeconomic properties may causally induce individual accountings in their decisions, at the end of the day, the decisive direction of causality is vertical – from individuals to aggregates. The causal efficacy of macroeconomic entities is real, although it is momentary and indirect: people's behaviour count as the primary social unit capable of shaping the reality of the macroeconomy. Therefore, following Hoover, we may bet that when all microeconomic facts are fixed, the macroeconomic facts should also be set. A conclusion that may calm the ontological fears and revulsion of many economists who reject the possibility of getting pure and direct causation at the macroeconomic scale, frequently labelled as "spooky" (Ross, 2014, p. 114).

4. Two challenges of global supervenience

In this section, I lay out the two selected criticisms voiced against global supervenience. I begin by introducing the downward causation problem levelled by Reiss (2004). Then next I reconstruct Epstein's challenge to global supervenience.

autonomous, and ontologically distinctive from its base entity. This view is, according to Wilson, the proper characterisation of non-reductive physicalism, which indicates why Hoover qualifies his approach to economics as non-reductive strategy.

4.1. The downward causation problem

The first appearance of "downward causation" comes from the work of psychobiologist Roger Sperry. Crudely put, the problem appears when the direction of casual influence extends from higher levels of organisation down to lower levels. The idea presupposes an ontological architecture of the world disposed of in layers. The structure of reality may be seen as consisting of a plurality of layers, where each domain consists of different entities with different properties defined over the respective domains.

Reiss's criticism does not depend on any metaphysic resources to stand on its own. To reject supervenience, Reiss drives our attention to the economics' practice and how people – the fundamental unit of microeconomics – normally behave. We may agree with Hoover that macroeconomic properties are real and accept their existence are somewhat anchored in the microeconomic world, but the combination of these two assertions does not logically entail that micro-and-macroeconomics are supervenient related. After all, if macroeconomic events can causally affect micro-events – a fact that Hoover does not deny – then how should we endorse (global) supervenience? As Reiss opposes,

> I do not see how the macro could supervene on the micro in any strong sense [...]. One point is that, clearly, macro entities causally influence micro entities (when, for example, agents react to inflation or recessions or changes in the federal funds rate). This contradicts both the spirit and the letter of supervenience theories. (Reiss, 2004, p. 232)

A direct influence from macro data on individual decision-making contradicts the message of supervenience in economics. Macro facts are codified by people; individuals adjust/alter their decisions according to the circumstances, including the facts from macroeconomics or their expectations about their future status. This fact denounces that the direction of causality does not take a unique bottom-top direction as supervenience necessarily requires. Consequently, it seems that microeconomics, the realm where intentionality reigns, is not purely micro. If macro facts are read by individuals as their informational inputs, then the micro-and-macroeconomic universes are not strictly disjointed, as the supervenience thesis also demands. When micro facts are fixed, the micro-set is entirely contaminated by facts or expectations from macroeconomics. Without disentangling these two economic domains, it is impossible to follow supervenience. Therefore, the claim of supervenience to portray the relation between micro and macroeconomics is wrong: two of its crucial criteria are simply not matched.

4.2. The autonomy of macroeconomics' properties: Epstein's critique

To test the plausibility of supervenience, Epstein (2014) argues that it is important firstly to fix what is permissible within micro and macroeconomic sets. Very often, microeconomic models include entities beyond individuals: goods, governments, and even the environment may belong to microeconomics. A too comprehensive microeconomic base that would enlarge its boundaries far beyond the scope of individuals may risk turning the interlevel dependence under the umbrella of supervenience into a trivial exercise. He thereby defines that nothing but individuals and specific kinds of features should make up the microeconomic set: the microeconomic set is composed exclusively of individuals' intentions plus the things they are in close causal contact with (sensible objects and other people). For example, when you buy a bottle of wine at a supermarket, you materialise a microeconomic fact: a market transaction between a firm and a household. The means of payment you used to pay your bill (money or debit card), the wine (economic good), and the checkout assistant constitute the socio-economic features in close causal contact with you. The legal apparatus that makes that economic fact legitimate under the eyes of the law does not make up the microeconomic set. Weather conditions – an external (exogenous) variable – are also not included.[79]

On the other hand, the macroeconomic set includes all macroeconomic aggregates (GDP, inflation, interest rates). If macroeconomics supervenes on microeconomics, then macroeconomic results would be fixed by this set of microeconomic facts, Epstein insists. He formulates an imaginary scenario to make his argument against supervenience. Imagine a hypothetical country randomly suffering from weather disasters. To diminish their effects on the population, local Tories voted for a public expenditure package. The approved rule follows the criteria below:

(a) If storms hit houses, then the government must pay $ 10,000 per damaged household;

(b) If hurricanes impact houses, then the government must pay $ 50,000 per household.

Consider two possible scenarios:

Case *A*: The country has been affected by a storm, generating damages for 1000 people.

[79] By "close causal contact", Epstein understands a superset of the microeconomic properties whereby each intrinsic property of individuals is included plus the properties of parts of the world that they are under direct contact with. This expression must be read reasonably and pragmatically. The exclusion of legal facts from microeconomic set is the point of dispute between Epstein and Galbács.

Case *B*: The country is impacted by a hurricane, generating damage for 1000 people.

In both scenarios, the same number of people is impacted; i.e., there is no difference in both microeconomic sets:

> In case *A*, the government shall accrue $ 10 million in obligations while in *B* it shall get $ 50 million. Notice that this obligation is simply a legislated fact. Whether or not the government has this obligation depends on a physical fact about the world. This physical fact is independent of the epistemic or other states of the individuals in the population. In case *B*, the presence of a hurricane sets government obligations at a higher level than they are set in case *A*, irrespective of whether there are causal consequences for the individuals. (Epstein, 2014, p. 11)

Thus, the necessary and sufficient conditions for the government's incurring obligations are determined by the weather conditions: if the evidence shows the case in place was a storm or a hurricane, one or another type of obligation will be triggered.

Epstein assumes that a weather event happens, enabling thus one or the other of these obligations. At that time (year 1), the necessary and sufficient conditions to unlock the government liabilities were present. Let's also suppose that the government only gets acquainted with the true nature of the disaster a few years after the disruption. Table 3 shows the economic effects of storm's outbreak in case *A*:

Table 3: Some macroeconomic statistics in case A

	Year							
	0	1	2	3	4	5	6	7
Event	Normal	Storm	Calm	Initial payment	Acquire information	Normal	Normal	Normal
Interest rate (%)	4	4	4	4	4	4	4	4
Government obligations ($ million)	–	10.0	10.4	–	–	–	–	–
Tax levied and collected ($ million)	–	–	10.4	–	–	–	–	–
Government disbursements ($ million)	–	–	–	10.8	–	–	–	–
Current surplus (deficit) ($ million)	–	–	10.4	–	–	–	–	–
Productivity growth (%)	3	2	2	4	3	3	3	3

Source: Epstein (2014, p. 12)

In *A*, the interest rates ratio was fixed across the years (from year 1 onwards, the interest rate ratio is 4% per annum). When it occurs, the storm generates $ 10 million in governmental obligations. In the year after (year 2), the obligation should be corrected according to interest rates, totalising $ 10,4 million and payments were finally made in year 3 (adjusting the final value, once again, by the interest ratio). In year 4, the official weather committee produced their

report stating that it was a summer storm that had broken out in year 1 instead of a hurricane, and then things went back to normal: the productivity growth that was hit since the impact returned to its regular pattern in the years ahead (from 4 afterwards).

Consider now the situation in case *B*. The table below depicts the economic consequences triggered by the hurricane:

Table 4: Some macroeconomic statistics of case B

	Year							
	0	1	2	3	4	5	6	7
Event	Normal	Storm	Calm	Initial payment	Acquire information	Corrective tax	Corrective payment	Normal
Interest rate (%)	4	4	4	4	4	4	4	4
Government obligations ($ million)	–	50.0	52.0	43.3	45.0	46.8	–	–
Tax levied and collected ($ million)	–	–	10.4	–	–	46.8	–	–
Government disbursements ($ million)	–	–	–	10.8	–	–	48.7	–
Current surplus (deficit) ($ million)	–	–	10.4	–	–	46.8	–	–
Productivity growth (%)	3	2	2	4	3	1	4	3

Source: Epstein (2014, p.13)

Like in *A*, the interest rate is also fixed. However, unlike what happened in *A*, a hurricane was responsible for the government liabilities in *B*. Notice that, until year 4, that is, before the authorities got the information and assessed the nature of the phenomenon, taxes and disbursements were the same as in *A*. However, government liabilities have kept growing since year 1. After a careful evaluation from the weather committee, authorities finally agreed that it was a hurricane that had destroyed a thousand houses. Consequently, in years 5 and 6, additional taxes were levied to pay the obligations, provoking a massive impact on total disbursements in year 6. While what has been done by authorities in case *B* is the standard accounting procedure, they wrongly proceed in case *A* – they did not adjust taxes to cover the original value of the obligation dictated by the law. What is the moral of the story?

Between 1-to-4, individuals' conditions – e.g., the number of houses damaged and information about these citizens (microeconomic facts) – was the same in *A* and *B*. In both situations, macroeconomic data only became acquainted about the effects upon individuals three years later. It was only then that knowledge about people's conditions was processed and finally analysed, generating their effects on macroeconomic features (in years 5 and 6). However,

the amount of government obligation did not start in year 4. Macro states differed between *A* and *B*. Each one has also changed step by step from 1-4, each later step caused by earlier macro steps while micro remained unchanged. The amount of government disbursements in years 5 and 6 is a result of the contracting obligation accrued over these years *regardless* of the micro facts about people's conditions until year 4:

> […] the facts about the government year 1 liabilities matter, even if no one knows those facts until year 4. The facts themselves are not exhaustively determined by the epistemic states of individuals, or by their intrinsic properties, or by things they are in close causal contact with. Despite the microeconomic facts being indiscernible for the first few years in the two situations, the facts about the government's respective liabilities differ. (Epstein, 2014, p. 14-15)

Since the initial year, both macroeconomic sets ran autonomously. Their results in year 6 were not impacted by microeconomic facts. Across the years 1 to 6, there was no persistent co-variation between microeconomic facts and macroeconomic facts: macroeconomic effects have only macroeconomic causes. Contrary to supervenience that would require a complete bottom-top determination, what we actually have as the outcome is (partially) independent from micro-level: "[…] different levels of government expenditures induced by different environmental outcomes trigger differ different macro-situations, whilst the micro-level remains the same" (Galbács, 2020, p. 184).

Events *A* and *B* are better explained, Epstein concludes, by the partial autonomy of macroeconomics from microeconomic sets. We should therefore abandon supervenience, recognising its failure to link micro and macroeconomics. Microeconomics casts light upon several economic phenomena (e.g., how relative prices are formed), but it cannot serve as macroeconomics' foundation. Perhaps, as John King suggests, micro and macroeconomics are horizontally related rather than vertically:

> […] microeconomics and macroeconomics are related to each other horizontally, not vertically. The phenomena that they study are rather obviously closed related, and their practitioners most certainly can (and should) cooperate with each other and learn from each other. But neither should be seen as the "foundation" of the other. (King, 2012, p. 10)

5. Global supervenience is still in the game: Galbács's response to Epstein

To Galbács, Epstein's cases fit perfectly well under the microfoundational picture of the economic world advocated by Lucas. In a nutshell, he disputes what should be included in the microeconomic set. According to Galbács, Epstein's argument goes through only if we set aside the socio-economic structure from the microeconomic set – the actual institutional surroundings where individual actions occur. In doing so, we act as if people's beliefs and

intentions take place in a vacuum since their environment is disconsidered. However, individuals' economic actions always take place in some socio-economic structure. The structure, plus its features (social, cultural, and legal constraints), must be inserted into the supervenient base, Galbács argues. The rules that codify the macroeconomic liabilities in *A* and *B* were built into their structures. Thus, Galbács argues, the microeconomic set should be comprised of individuals and the institutional setting where they interact. In reviewing the micro set, we re-orient Epstein's hypotheses, turning it more suitable to Lucas's ambitions:

> At the bottom line, even though along dissimilar arguments, Epstein and Hoover see eye to eye on the failure of microfoundations project as agents and their properties in themselves cannot define the macro level. […] Macroeconomics in itself fails to rest exclusively upon microeconomics for macroeconomics contains a lot of elements that cannot be derived from the micro. In a strict reading it follows that if a model contains "things" other than agents and their properties, then microfoundations are insufficient. But what are these "other things" in Lucas's models? A well-defined setting taken as the environment. A playground that imposes also well-defined restrictions upon the interactions of agents with a government and some institutions that can be implicitly regarded as created and carried by individuals. (Galbács, 2020, 186-187)

In Epstein's scenarios, the macroeconomic results are not directly determined by weather conditions. Hurricanes and tornados can make people homeless instantaneously, but they cannot trigger macroeconomic liabilities on their own. What happens in Epstein's macroeconomic sets could only come about due to the underlying microeconomic structures that create the rules. Weather events were coded by these rules (the economic structure), and hence we have macro liabilities:

> Passive (rule-based) responses to stochastic events like changes in weather fail to lead to changes in the structure in the economy. Weather outcomes do not establish differing macro-level settings, rather possible reactions to changes in the weather are embedded or coded in the structure. (Galbács, 2020, p. 188)

If the macro shows different states, it is because different states as responses to external events were coded into the micro—into the rule-based budget. The same number of individuals across the micro sets *A* and *B* is not *sufficient* to determine that those sets are identical. It is also *necessary* to consider the institutional features that make up the micro sets *A* and *B*, respectively. According to Galbács, the difference between the macro liabilities *A* and *B* is obvious: they result from *different* micro bases. Different institutional settings – micro bases *A* and *B* – fix different macro properties. So, the introduction of an informational void between the outbreak of facts (in year 1) and their appreciation in (year 4) in one of the cases does not create any barrier to global supervenience: Even though macro results in year 6 were accrued horizontally from 1 to 6, underlying structural micro conditions were present along the line – people's interactions within a socio-economic setting. The macro results were only possible due to structural facts that come up from the micro. People's properties combined with institutional properties fix the

macro sets: these are the necessary and sufficient conditions to fix the macro. I shall explore these properties further in section 6.2.

6. Possible answers

In this final section, I provide the potential replies to Reiss and Epstein's criticisms. I begin by tackling the downward causation problem raised by Reiss (Sub-section 6.1) and then I approach Epstein's critique, which endorses Galbács's argument but advances it significantly (Sub-section 6.2).

6.1. Accommodating the downward causation with supervenience

In response to Reiss, I argue that the downward causal input from a putative macroeconomic entity towards microeconomic units does not collide with global supervenience necessarily. Like many other real-world features that daily constrain our economic decisions (e.g., merchandising, culture), macroeconomic data at t_1 is undoubtedly among one of these real-world environmental settings influencing economic choices, and this is consistent with supervenience so long as the macro could not be different given all micro properties at t_1.

For supervenience theorists, a macroeconomic set_1 – a collection of macro data (real GDP, rate of inflation, etc.) at a certain spatio-region of time – supervenes on properties about individuals (e.g., a microeconomic set_1) but not vice versa. At another point in time, say t_2, the microeconomic set_2 could be causally affected by previous macroeconomic set_1 but never by the supervenient macroeconomic set_2. So, the macro effects upon the micro (the top-down causation) and the micro globally fixes the macro (global supervenience) is possible as long as these determinations do not happen simultaneously. To see a scenario where this is possible, consider a famous macroeconomic principle, Taylor's rule.

The rule, baptized with the name of its proponent, John B. Taylor, consists of a monetary-policy prescription that stipulates how much the central bank should change the short-period nominal interest rate in response to changes in inflation and output or other economic conditions. In particular, the rule dictates that the central bank should raise the nominal interest rate by more than one percentage point for each one-percent increase in inflation. According to Taylor's principle, the short-period nominal interest rate is a manageable variable at the hands of central banks to keep inflation (and other macroeconomic variables) under

control: "the policy rule […] has the feature that the federal funds rate rises if inflation increases above the target of 2 percent or if real GDP rises above trend GDP" (Taylor, 1993, p. 202).

Taylor's principle is employed chiefly when monetary authorities adopt the inflation targeting regime. Since the 1990s, various central banks across the globe whose top concern is the stabilization of the level of prices and the output (e.g., New Zealand, Brazil, Chile, etc.) have embraced inflation targeting policies to achieve their goals. Summarising, inflation targeting is a public announcement of official quantitative targets (or ranges) for the inflation rate over one or more time horizons with the explicit acknowledgement that low-stable inflation is the long-term monetary policy. Very briefly, the central bank forecasts the future path of inflation and compares it with the rate of inflation the government believes is appropriate for the economy. The difference between the forecast and the target fixes how much monetary policy must be adjusted.

Rather than focusing on always achieving the target, the approach has emphasized achieving the target over the medium term—typically over a two- to three-year horizon. This approach allows the policy to address other objectives—such as smoothing output—over the short term. Thus, inflation targeting provides a rule-like framework within which the central bank has the discretion to react to shocks. Because of inflation targeting's medium-term focus, policymakers need not feel compelled to do whatever it takes to meet targets on a period-by-period basis.

Adopting a nominal anchor is an assertive monetary policy for several reasons, but most importantly is justified by its effects on uncertainty about the future states of the economy. As we saw in the earlier chapter, as forward-looking agents, several economic decisions depend upon people's expectations about the future. Firm's decisions to expand production depend on its expectations about the future of inflation. A rise in inflation might indicate that the price of goods supplied by a firm might be higher in the next term. Second, firms do not pay nominal wages to their employers (wages paid in terms of money); they pay real wages (wages adjusted by inflation). Production cost and profitability are directly associated with inflation expectations. Given the importance of inflation, setting up a target over a medium-term may calm agents' expectations.

Very often, when a shock disturbs the rate of inflation, central banks adjust the nominal interest rate, ensuring that the real interest rate moves in the right direction to restore price stability. A shock, for example, that shifts inflation upwards makes the central bank lift interest rates. This action cascades its effects on the microeconomic level: the refinancing costs of commercial banks increase, whereas their demand for refinancing operations is likely to

decrease (Herger, 2019). Economic activity is therefore sensibly affected. The momentary slowdown in the economy pushes inflation down towards its point of equilibrium envisioned by a central bank.

To illustrate the mechanism that engenders central bank actions, the downward effects on the microeconomic level, and therefore over inflation, consider the recent breaking news coming from Brazil. In June 2021, Brazil's Monetary Policy Committee (Copom) announced its new inflation targeting for the next two years:

> The Economy Ministry said the 3,00% target for 2024, with a tolerance margin of 1.5 percentage points on either side, will reduce uncertainty and allow households, companies and the government to plan better the future. "The 0.25 percentage point reduction from the 2023 target is consistent with the high credibility of monetary policy", the Economy Ministry said in a statement. (Reuters, 2021)

Since 1999, Brazil has adopted the inflation targeting regime. Copom tries to intervene in the nominal interest rate to render control of inflation (measured by the Brazilian National Consumer Price Index – IPCA). In June 2021, for example, in an attempt to reduce inflation, Copom raised their benchmark Selic rate by 75 basis points for a third time to 4.25%.

Recent news and public announcements from Brazilian's monetary authorities reveal their heavy commitment to keep inflation under control. Last June 2022, Brazil lifted the Selic rate by 50 basis points in line with market expectations.[80] So, we might expect that Copom will keep lifting the nominal interest rate above the inflation to push inflation towards the desired target to calm down agents' expectations. Considering what has been said about this mechanism used to control inflation, let us come back to the issue of downward causation and supervenience.

Following Hoover's account, the nominal interest rate is a macroeconomic variable capable of rendering control upon the general price level. The variable is real insofar as it may keep other macro aggregates under control – even though indirectly. We assume a potential causal relation at the macroeconomic level – where the announcement of the nominal rate of interest, macro set_1, is the variable subject to interventions (the cause), while the rate of inflation, macro set_2, represents the outcome. Herein, the causal powers of interest rates are more linked to the announcement itself rather than the proper means employed by the authorities to manipulate interest rates (selling bonds, open-market operations, etc.). As stressed by one of the architects of the contemporary monetary policy, Ben Bernanke (2015):

> When I was at the Federal Reserve, I occasionally observed that monetary policy is 98 percent talk and only two percent action. The ability to shape market expectations of future policy through public statements is one of the most powerful tools the FED has.

[80] See: https://www.reuters.com/markets/us/brazil-central-bank-raises-rates-by-50-bps-signals-another-hike-2022-06-15.

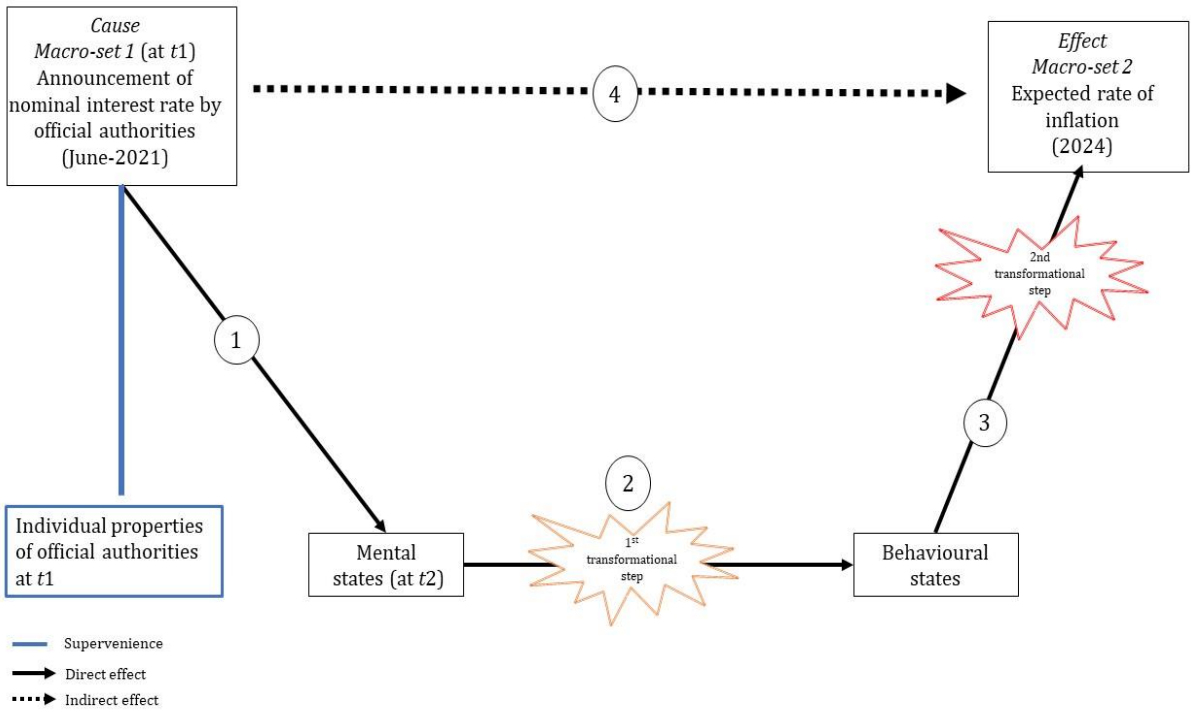
> The downside for policymakers, of course, is that the cost of sending the wrong message can be high.

This is not at all surprising. After all, under the paradigm of the rational expectations hypothesis, insofar rational agents use all the available information economically, then the announcement is *per se* sufficient to provoke economic effects:

> As economic agents assimilate policy intentions as their own personal expectations, they do the work of central bank. Expectations (guided by central bank rhetoric) can thus influence the course of inflationary and deflationary processes independent (or in anticipation) of conventional interventions on interest rates, in a form of self-fulfilling prophecy. The bridge to the ephemera of expectations – expectations that shape economic behaviour prospectively – is constructed with words. (Holmes, 2018, p. 177)

As mentioned above, the elevation of the nominal interest rate dissuades spending among economic actors by lifting the real interest rate. At t_1, when the central bank implements the policy, the effects upon economic actors (the micro base) cascade at another different regions of space, perhaps at t_2. If people behave predictably – assuming that other things are held constant – inflation in 2024 tends to decrease due to a gradual momentary slowdown in the economy – people alter/adjust their economic actions in the face of new information from the macro. The transitory slowdown of the economy broadly pushes the relative prices down, thus affecting the rate of inflation. The diagram below represents the causal machinery that takes us from the first causal push to the ultimate effect:

Figure 3: Brazil's inflation targeting: the structure of the three-year horizon deadline



Like the explanation of the origins of capitalism explored in chapter 2, this example displays similar trends in bridging the initial cause until the outcome. On top, there is a causal relation at the macro scale that can be read counterfactually: different values of interest rates would bring about different results of inflation – there is a horizontal causal determination from the macro-conditions to the macro-outcome (arrow 4). This is mediated by the announcement concerning nominal interest rates imposing an objective constraint upon individuals capable of influencing their decisions. There is a difference-making from the top to the base (downward causation).

We first have the announcement (macro set 1), and *then* people react to that new information. In this regard, the psychological states of individuals – including their expectations about inflation, are explained in terms of their environment (arrow 1). Then, people's behaviour is explained in terms of their mental states. Arrow 2 depicts this transformational step between different properties inside the micro base, from mental to behavioural properties. And then, we have an inter-level transformational step, from behavioural properties to the macroeconomic outcome (arrow 3). This is the circuit of macro-micro-and, once again, macro transitions.

For the purposes of this economic explanation, we need not focus upon the macro-micro supervenience relation at t_1. Even though the bank's announcement depends on the actions of individuals at the micro-level, in many explanatory contexts we may ignore the precise mechanisms upon which the announcement depends. In this example, monetary authorities' individual properties at the bottom line are their beliefs, dispositions, intentions, attitudes, local situations, and relations. On top (macro set$_1$) are the committee's pre-existing roles and positions where a group of individuals occupies these positions.

Copom was formally created in 1996. Originally conceived to set the monetary policy and define the rate of interest, one of the primary concerns of the committee is to keep the general price level at a manageable margin capable of attending to the interests of the economy. Experts who occupy the positions in the committee must seek the best monetary policy capable of enhancing economic welfare, taking into account various issues – inflation, market expectations, government finances, economic activity, and external scenarios.[81] This is what the commission was legally designed for. The committee has a *purpose*, a goal to achieve. So, if intentions are, loosely speaking, states in a planning system that settle matters about what to do, then individuals, robots, bees, and the committee itself can all qualify as *intentional agents*:

> Suppose that and action-focused output of a social procedural rule of a subcommittee says: send medical supplies to C. And suppose this is an institutional output. My claim is that this crystallized, institutional output will be set to function in ways characteristic of intention. We thereby infer that it is, functionally speaking, an intention of Medic

[81] See: https://www.bcb.gov.br/en/monetarypolicy/committee.

> Supply. It is a construction that is an institutional intention because it satisfies the functional specification for such an intention. (Bratman, 2022, p. 135-136)

Ideally, embodied with public spirit, committee members try to materialise, as much as possible, the intentions of the institution.[82] The committee members' intentions are assessed in accordance with their impact over some specific targets, how the members are functioning to accomplish the purposes of the institution. Inevitably, all the members of Copom hold a social position – i.e., the position that members occupy in hierarchy of the institution – that is fundamentally characterised by the role they play for the institution's function.[83]

Taken in isolation, individual members' decisions are irrelevant to economic agents. Only the committee's collective decision can produce economic effects. However, when the announcement is publicised – a collective decision turned public in the institution's name, the macro set$_1$ is not independent of individual's decisions (micro set$_1$). The relation between individual decisions and a collective decision is necessary. The properties on the top co-varies synchronically with individual's decision. So, we may say the group decision supervenes on individual's decisions: only when all individual decisions have been fixed, the group decision is also set, even if the latter does not reduce to the former.[84]

For some authors, this conclusion may suggest an unacceptable thesis that the economic effect – the agent's formation of their expectations based on the announcement by the authorities – has two sufficient causes for the same effect (individuals and the group decision). Agents' expectations would thereby be causally overdetermined. However, on close inspection I do not think this example presents a problematic proliferation of causes.[85]

The two levels – collective decision-making and individual decision-making of committee members – are in harmony to determine the beliefs of economic agents at t_2.

---

[82] Adolph (2013) argues that the formative working years of central bankers (partially) explains central bank policy decisions. In countries whereby there is no stringent rules (or no rules at all) regulating revolving doors (i.e., individuals who switched private for public sector and then return to private sector), the interest rates are higher. Revolving doors policies are fixed to avoid the use of privileged information directly get from central banks.

[83] This approach looks at collective intentionality in terms of their relevant role in some specific explanatory contexts, how they are causally relevant for the manifestation of a given target phenomenon. When we move down to the individual level, we realize that this view is consistent with *functionalism* about the nature of (individual) mental states which understands that what makes something a mental is identified by its role, i.e., its causal relevance for individuals' behaviour. Recently, Franz Dietrich and Christian List (2016, p. 268) provide a defence of mentalism in economics that is tuned with functionalism. This approach is ontologically neutral about the content of individual (and collective) mental states – sometimes it may amount to mental states, others not necessarily as this case of collective intentionality aims to endorse.

[84] For a more detailed defence in favour of supervenience between individual and group decision-making and a variety of supervenience arrangements between individuals and groups, see List and Petit (2006, 2011).

[85] Kim (1988) became famous for levelling what has been named as the causal exclusion argument in the philosophy of mind: events cannot have more than one sufficient cause. Unlike Kim, many authors do not accept his worry, and argue as I do above that proliferation of causes is not necessarily problematic. I do not have the space to discuss this issue further, but for further details about how this dispute plays out, see, for instance, Yablo (1992), Bennett (2003), Stoljar (2008), and Kroedel and Schulz (2016).

Concerning economic agents' beliefs/expectations, individual and institutional/collective properties are not metaphysically independent – there's no competition between committee participants' decisions and the group's decisions (the announcement). Nonetheless, in this specific case, macro level explanations are much more illuminating than appealing micro level explanations. Knowing the nature of divergences inside the group can be interesting for some purposes (e.g., to understand the theoretical and political biases of some committee members), but it is not relevant for economic agents. Only the committee as an institutional agent is epistemically important. This kind of inter-level structure is also present at various levels of science. For instance, if I have a headache and my partner asks, "Why did you awake in the middle of the night?", I shall probably answer, "I was in pain and decided to go down to the toilet to take a paracetamol". My wife doesn't expect me to be answering with a lower-level neurological explanation like, "C-fibers in my brain were firing. So, I decided to intervene on it".

In our case, the downward causation relates the macro set_1 with a different variable, a micro set_2, at different times. The set of free empty spaces with roles and positions in the macro set_1 is occupied and instantiated by the individual properties of the committee's participants. Although the committee can be qualified as an intentional agent with its own purposes, it can only display its causal powers and seek its goals *via* the members who occupy its positions and manifest their properties (decisions, beliefs, and attitudes). First, the macro set_1 *instantiates* the micro set_1 properties – the pre-existing social roles and positions are fulfilled by individual's properties, and then the macro set_1 downwardly causes the micro set_2.

Interestingly, this case reveals that supervenience can appear behind the curtains. The most visible phenomenon is the downward causation from macro set_1 towards micro set_2. Supervenience between micro set_1 and macro set_1 is less transparent, but it exists. However, when our epistemic goal is to understand which feature of the world is causally significant to understand the economic phenomenon, this example shows that sometimes supervenience appears as a mere modal notion without explanatory power.[86]

6.2. Institutional features matter

I follow Galbács's intuition: macroeconomic results are essentially derivative – they are products of the microeconomic set. His interpretation is aligned with the intentions of global supervenience: to permit a global determination between micro and macro sets of properties

[86] In the next chapter, I address the issue of explanation, showing how supervenience, depending on the case, can be explanatory. Epstein's case is an example of supervenience that serves to explain the target of interest.

that could accommodate intrinsic and extrinsic properties. However, Galbács does not state what is his criterion to accept and reject extrinsic properties from the micro set. Knowing this criterion is crucial to inform us about the role of extrinsic properties inside the micro set, how they contribute to the rising of the macro set.

Epstein, in contrary, offers his criterion. The types of extrinsic properties permissible in the micro set are only those that, besides the intrinsic properties of individuals, have direct causal connection to individuals. For instance, in my ranking of breweries, *Durham Brewery* is at the top; *Guinness* comes in second, and *Carlsberg* is granted third place. Assuming that this set of options is all available at the supermarket, I would prefer *Durham's*. But sadly, my favourite beer is currently extremely expensive, so I make the move for *Guinness*. In Epstein's account, the prices of goods function as the real features, the extrinsic properties, which are in immediate contact with people's economic actions – they exert a strong influence upon decision-making. I manifest my preference for *Guinness* over *Carlsberg* when I pick up the pack of the former even though the latter was cheaper than the Irish brew.

On Epstein's criterion, the institutional setting does not have the same kind of causal influence as the relative price when it comes to realising this economic exchange. It seems that, in his view, inasmuch the institutional setting is more like an environmental (pre)condition (or a constraint) that makes possible every microeconomic fact, that condition is not *on par* with other crucial extrinsic microeconomic properties such as relative prices. So, Epstein is at least explicitly committed to the idea that this part of the social world, this (institutional) extrinsic property that is not directly modifying or influencing decision-making, can be placed outside the micro set. This is visible in his argument where only the epistemic states of individuals are taken into account, and nothing else is relevant for the rise of the corresponding macroeconomic sets.

Contrary to Epstein, I contend that even though institutional features do not exert the same type of causal influence as the features directly involved in microeconomic facts – taking causation as a sort of difference-making upon decision-making, the institutional features impose an objective (causal) *constraint* upon individuals. Judicial rules, contracts, economic rules etc. are structural features that make economic attitudes possible, they are the rules of the game – the formal constraints (North, 1990, chapter 6). Legal rules do not have to be directly causally related to individuals to constrain their behaviours: '[…] the subject's mental states depend upon social facts about the conventions and knowledge available within a community, even where those facts make no direct causal or perceptual impact on the subject' (Currie, 1984, p. 354). People take for granted the legal standards required to materialise their economic

exchanges; ordinary persons take the existence of institutions as given – a social fact (Brisset, 2016, p. 9): 'legal rules, constitutions, and other forms of institution "constitute a part of individual's circumstances which together with his aims determine his behaviour"' (Agassi, 1960, p. 247).

Normally, this sort of institutional constraint shapes, guides, and imposes limits upon the behaviour of a given system or upon an individual. Legal rules dictate restrictions, the range of values of some target of interest. Institutional and individual agencies are interdependent on each other. According to what is our target of interest, they can both work together to produce the same social outcome. Joseph Agassi qualifies this view as *Institutional individualism*: we can both see individuals *and* institutions as exogenous variables capable of bringing about social effects. Taking into account what we aim to respond to, they can be equally seen as the primary social units entitled to perform as the explanans in sociological explanations (Agassi, 1975, p. 152).

While institutional facts and facts about individuals can both count as causal, they manifest their causal powers differently: while the institutional factors impose the constraints, human agency is more like the run-of-mill cause in the social realm – the triggering causes of social facts. So, depending on what is our focus and how ambitious we are to elucidate the phenomenon at stake, its complete explanation may require these two dimensions – the institutional aspect and the human agency.

In my view, in Epstein's scenarios, these two dimensions, the institutional and the individual, are both crucial for fixing the corresponding macroeconomic sets. Without the combination of these two factors, we can't make sense of what is going on. On my interpretation of Epstein's cases, the existence of the macroeconomic sets is subordinated to the existence of their corresponding micro-sets; the latter (the micro set) is more fundamental than the former. If there was no rule and no effect upon individuals caused by the storms, the budget would not have existed.

James Woodward's (2003) interventionist account of causation provides a straightforward answer to understand the relevance of the legal rules for Epstein's argument. Following Woodward, to acknowledge there is a causal relation between variables *X* and *Y*, they must exhibit a robust counterfactual dependence: if we change the value of *X* in several ways, *Y* should be also modified.

Let's say the existence of rule *A* = [If hurricanes impact houses, then the government must pay $ 50,000 per household] assumes the value (1), while its counterfactual – the nonexistence of the rule, is (0). For the case under assessment, it is obvious we can only make

sense of how the obligation was possible under the presence of the rule: if there was no legislation, the macro set would not have been brought about.[87] At the same time, the government obligation partially depends on the nature of the external event *and* the impact of the storm on the population (Epstein, 2014, p. 11). The budgets schemes were not *always there*, placed at some spatio-temporal region of reality, patiently waiting to respond to external interventions.

The budget schemes come up from more fundamental units of the social world – the epistemic states of individuals damaged by the storm – the *raison d'être* of the rule – *plus* the legal statute that defines the public expenditure. They emerge in virtue of the effects on people and the rule which enables the expenditure procedure, fixing the limits of how much should be paid for rebuilding houses. Like the Hobbesian idea that the existence of society is derivative from its very components (individuals), the macro sets' existences are also derivative: they emerge from more fundamental features – people's conditions plus institutional properties, the fundamental constraint of the macro sets. Institutional properties should be placed inside the micro set; they constitute a necessary condition for the emergence of the macro sets in question.

In some measure, it seems Epstein is (partly) aligned to this view. In his budget scenarios, legal facts are *presupposed* by microeconomic facts; they are implicit. We should transform what is implicit into explicit: the micro-set must welcome legal facts – institutional properties and people's properties make up the base. When the weather events took place, they were coded by the micro set. Once they have been globally fixed at the micro level, so they are the macro level. The circuit of determination of the macro set flows from the initial exogenous variable (weather conditions) towards the base. It is the base that codifies the external events that fix the macro. It is only when the micro set is settled, that the macro is also fixed.

Epstein might retort, arguing that this does not explain the logic of the budget, fixed step by step along the years without requiring information from individual conditions, demonstrating the autonomy of macroeconomics. It is worth noting that to reject supervenience, Epstein defends the autonomy of macroeconomic facts via *intra*-level determinations, i.e., relations of dependence that happen inside a given set/or level. Comparing the situations of the macro facts before and after the fourth year, he argues, we are naturally led to conclude there is no room for a bottom-top determination from micro towards macro. Each macro fact, an individual spatio-temporal region or event, is exclusively determined by its predecessor macro fact, no matter what went on at the bottom line.

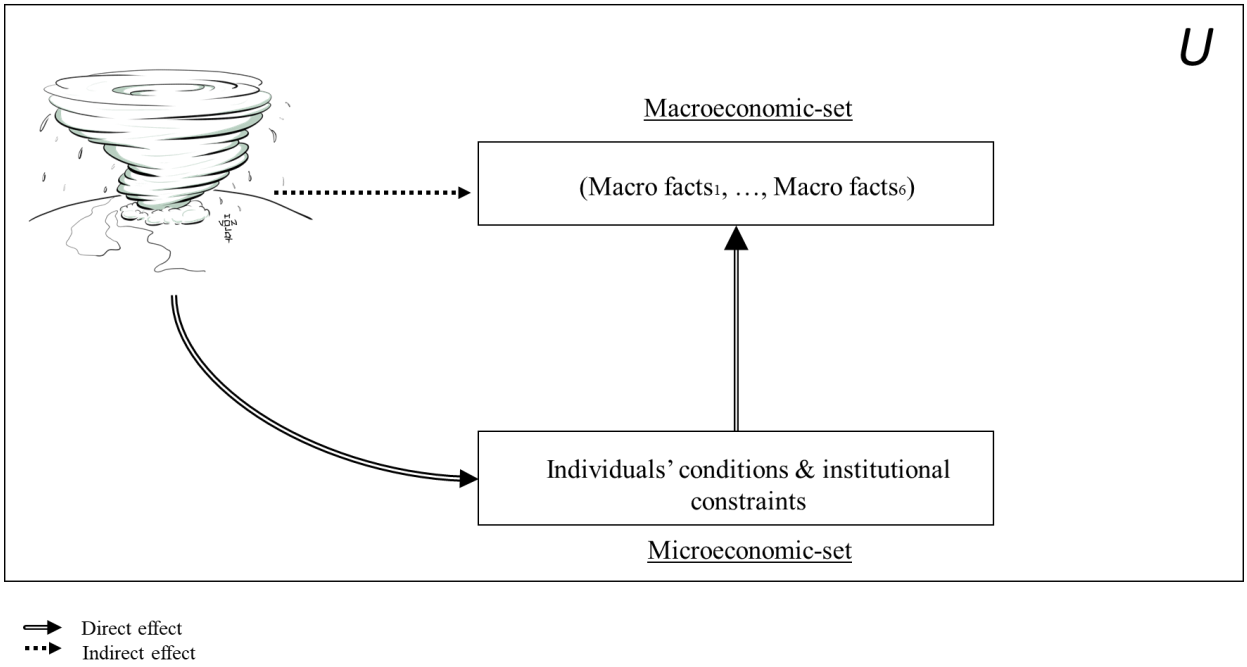
[87] A similar idea was recently put forward by Ross (in Review).

However, in turn, global supervenience aims to fix *inter*-level determinations. It looks after global determinations without local determinations. From years 1 to 6 (i.e., the complete macroeconomic set), there's no single macro event whose determination has been directly fixed by something other than the combination of people's conditions plus the institutional facts in which they are embedded. It is the institutional facts that impose the limits of the budgets, also fixing the right procedures for their adjustments over time. So, the intra-level relations inside the macro set are all predicted by the rules. The supposed autonomy of macroeconomic facts is not at all autonomous in any strong sense of the word; they are granted and *expected* in accordance with the rules. Macro facts, as Hoover says, are anchored in the structure – they *emerge* from the micro.

We surely may agree with Epstein that there is a higher-level cause between the tornado and the macro sets; both events are counterfactually channelled. The external event is an obvious necessary condition for the rise of the macro set. What I disagree with is that the relation between these two events – the weather event and the emergence of the macro sets – is direct: they are *not*. The timing lag issue levelled by Epstein is insufficient to demonstrate the autonomy of macroeconomics, that something outside the micro set is directly responsible to determining the states of the macro. Both institutional facts and the individuals' conditions represent the crucial factors to directly determine the macroeconomic outcomes.

In global supervenience, none of these microeconomic properties – properties of individuals, or institutional properties – have the dominant power to determine the macro. It is through the *combination* of these bottom-level structural features that macro properties are directly fixed. A potential methodological conflict between institutions vs. individuals, regarding which social unit prevails in determining the collection of facts at the macro-scale, simply fades away. The figure below aims to capture my interpretation of what goes on in Epstein's example:

Figure 4: The causal-chain: from weather event towards the budget scheme



The tornado fixes the macro set, but only indirectly. The cause (the tornado) and the effect (the macroeconomic set) are real. If it wasn't a tornado event, the macro set would not exist. Both events are counterfactually linked. However, the causal link connecting that external event to the macroeconomic set is mediated by the microeconomic set, made of individuals' properties plus their formal constraints. The properties of both (micro and macro) sets are (a) real, and (b) qualitatively distinct. Also, they are (c) correlated, (d) synchronic related (micro-set fixes the macro instantaneously), and asymmetric – the base fixes the top but not the other way round. The base functions as a mediator variable, relating the initial cause (the tornado) to the final effect (macro facts). Therefore, we can conclude that, in Epstein's scenarios, given that the micro set contains the kinds of institutional facts which are presupposed by the decisions of individuals and are crucial to fixing the boundaries of the macro sets, then the macro globally supervenes on the micro.

7. Concluding remarks

In this chapter, I evaluate two general claims against the applicability of global supervenience to relate micro-and-macroeconomics: the downward causation problem and the alleged autonomy of macroeconomics. I argue the downward causation raised by Reiss does not necessarily challenge supervenience. I tried to show that we may have supervenience and causation operating at different phases of the same economic phenomenon. By exploring a

paradigmatic example in economics, the Taylor's rule, I argued it is perfectly possible to have a collection of microeconomic properties (individual decision-making) fixing the macro properties (collective decision-making), whilst the latter causally determines another set of microeconomic arrangements. The competition between supervenience and downward causation is only apparent.

In the second case, I have shown why macroeconomic properties are arguably globally fixed by the corresponding microeconomic sets. When we look at the inter-level relation between the two economic sets, it's clear that the base fixes the top. Without the rules and the events cascading upon individuals, the macro set would not see the light of the day. The circuit of determination flows from weather events to the micro base and then we get the macro set.

This chapter reveals that the applicability of supervenience in economics may have different nuances. Sometimes, supervenience relations may function as a constraining operator as the Epstein's case illustrates; others as a mere modal relation with no explanatory power – it exists, if you look at carefully you may see it, but it is not always epistemically relevant for the economic event. It's not fruitful to discuss the applicability of supervenience in economics in terms of all-or-nothing: either we have reduction or supervenience. Like reduction, we may also have different types of supervenience with different roles and explanatory performances.

CHAPTER 5: THE MANY FACES OF EXPLANATION: THE CONTRASTIVE ACCOUNT

1. Introduction

In the former chapter (chapter 4), I argued that the downward causation problem raised by Reiss and the potential autonomy of macroeconomics levelled by Epstein are not fatal blows for the potential applicability of global supervenience in economics. In the first case, I argued that the announcement of an increase in nominal interest rates, the collective decision publicised by monetary authorities at t_1 (a macroeconomic event), supervenes on authorities' individual decisions synchronically – at t_1. The announcement's influence on individual decision-making materialises at different times – say, upon a collection of economic agents at t_2. So, there is no competition between downward causation and supervenience. When people undertake their actions, transforming mental states into behaviours, they shall fix the properties of the macro elsewhere (perhaps, at macroeconomic event$_2$).

In the second case, I argued when the institutional structure is inserted into the microeconomic set, the corresponding macroeconomic sets end up being directly fixed by the micro. There is nothing over and above the micro sets that is responsible for directly fixing the macro: once the micro set has been set – the set formed by institutional and individualistic properties – the macro is also settled. But what does knowing this bottom-up determination teach us?

Supervenience –local or global – states that some patterns of property variation (or facts) hold. It is just a modal notion rather than an explanatory one. It is silent about *why* these patterns are possible. It is not every supervenience case that is explanatorily illuminating. The application of Taylor's recipe in the former chapter exemplifies this question. Although the downward causation and supervenience are not mutually excluding one another, knowing the nature of the dependence between the committee's decision and its participants is irrelevant for economic agents. Thus, more work needs to be done to show there is also an explanatory connection between the supervenient target sets (the base and the macro). Can we sometimes learn something else from this sort of inter-level determination between two different economic layers? Can understanding the supervenience base help provide a previously unavailable explanation for macroeconomic properties?

In this chapter, I shall argue that, in paradigmatic cases that are of serious interest to economists, the supervenience base can explain macroeconomic properties. This remark is about supervenience, but a similar issue arises concerning reduction. Knowing that the macroeconomic properties could not be other than what they are once the microeconomic ones

are fixed and knowing just what those micro properties are in a given case is not enough to show *why* the macro properties are just as they are and not any other way. I raise this because I think the demands for explanation are crucial to understanding the appeal for economists, especially among mainstream scholars, for the call for microeconomic foundations for various macroeconomic phenomena. It is thought that without understanding what is going on at the micro-level we cannot understand these macro phenomena. Without a solid knowledge of microeconomics, the macroeconomic policy would be drastically uninformed, compromising the reliability of economic understanding and economic forecasting.[88]

More strongly for those who want to *displace* macro models and *replace them* with micro models for certain macro phenomena is the idea that microeconomics by itself is both necessary and sufficient for understanding those macro phenomena. So, I am here going to proceed from the assumption that a key reason for searching for microfoundations in a given case is to say that information from the bottom line (microeconomic level) is necessary for a *complete* explanation of the macro phenomena under analysis, and perhaps (depending on the context) even sufficient by themselves. Or even further, they may be necessary (and possibly sufficient) for accurate predictions about these macro phenomena.

In this final chapter, I focus on a special kind of explanation, particularly contrastive explanations. According to contrastive theorists, *X* explains *Y* if we contrast *X* (explanans) and *Y* (explanandum) with a set of alternative situations. When we confront *X* with potential alternative explanatory scenarios for the occurrence of *Y*, *X* must emerge as the unique truth possible answer to *Y*: it's *X* rather than *X**, that explains *Y* rather than *Y**. Currently, a wide range of disciplines employs the contrastive account: artificial intelligence (Miller, 2019), cognitive science (Chin-Parker and Bradner, 2017), historical studies (Northcott, 2008), and economics (Lawson, 2009; Morgan and Patokämi, 2017).

I open the chapter laying out the motivation and the main features of ternary contrastive accounts: *c* explains *e* rather than *e** (Section 2). Moving on, I lay out a one-step further approach, a quaternary account of causation that contends that causation requires contrast in *c*- and *e*-sides: *c* rather than *c** explains *e* rather than *e** (Sub-section 2.1). Following the chronological structure adopted throughout this thesis, I shall begin to revisit the examples explored in the dissertation. This is not, despite what it may seem, mere repetition.

I intend to demonstrate that, when we pose the contrasts in terms of questions and answers, we visualise other salient aspects of causal and noncausal relations capable of

[88] In stressing the explanatory role of the reductions and the supervenience as in paradigmatic economic explanations I stay neutral about whether these are really good or correct explanations.

connecting inter level relations. Unlike, for example, the usual ontological approach to reduction and supervenience in which these relations are fundamentally represented in binary forms, a contrastive view of explanation can exhibit explanatory forms of reduction and supervenience in a quaternary way, shedding light on different nuances involved in reduction and supervenience. As we shall see, the contrastive approach serves multiple roles beyond explanation: it can be used to reject hypothesis and sometimes is capable of being employed for predictions.

For this reason, I turn once again to Weber's explanation for the rise of capitalism explored in the second chapter (Section 3). I then revisit the Lucasian Phillips curve to show how in that case, the microelements cited (properties of economic agents) provide a good contrastive explanation of the macro phenomena in question (Section 4). Similarly, the same pattern of explanation can be extended to supervenience. In my reply to Epstein, I argue the characteristics of global supervenience are manifested in his example. Now, I step forward to show that Epstein's case can be squared into the logic of contrastive explanations (Section 5). The base provides a good contrastive explanation for the rising of the macroeconomic sets, a type of explanation based on constraints (Sub-section 5.1).

2. The pragmatics of explanation: van Fraassen's view

I was walking in Newcastle during the last Christmas time when I suddenly came across a gorgeous brand-new vinyl record store in the city centre. Naturally, as a vinyl collector, I did not resist and decided to go through that store. Like if I were a kid at Disneyland, I was amazed at how many unique items that store had. After 2 hours of browsing the shelves, I finally identified three missing items in my collection available at the store. Situations like these demand choices for ordinary people – we all have budget constraints, right? So, on that day, I had to choose among three wishing-list albums with the same price: (a) *Songs of love and hate* (Leonard Cohen), (b) *Disintegration* (The Cure), and (c) *Rumours* (Fleetwood Mac). I picked up (a) assuming that (b) and (c) must be easier to find across the UK – they were both recently re-impressed, so I guess it is easy to get better bargains for these albums in the future.

After shopping, I was ready to get the train back to Durham. It takes 15-min walk from the vinyl store to the Newcastle central train station. While I was on my way, I visualised a family in front of the *Smith Toys*: a mother, father, grandmother, and a child. I was walking by when I heard the father tell his kid: – "*Mum says*: *No*! *You must obey mommy*". It was Christmas eve, and everyone could rapidly notice what was happening. The boy was asking for every toy in town,

and his mom denied it. The kid, alternatively, asked for another toy from his nanny, who promptly agreed: - "*Of course*, *darling*". However, the father feverously rebuked the boy, arguing he had already subscribed to his mother's decision – getting two action figures for Christmas was more than enough.

These two dissimilar situations exhibit a similar pattern in one aspect. In the first case, my choice for Cohen's record entails (momentarily) saying no to *Disintegration* and *Rumours*: I picked up (a) rather than (b) or (c). In the second case, the father who tells the kid to obey her mom meant the boy should not listen nor follow the rules of anyone else but her mother, even though his nanny agreed to purchase an extra toy. We may see choices and even authority as being contextually dependent: I chose (a) under a given set of options, while the kid should respect his mother's authority and not play according to the nanny's rules. In a different setting and circumstance, the grandmother might dictate the rules, rather than someone else (e.g., in contrast to any other adult who is not a family member).

We can represent multiple daily human experiences and situations (choices, moral judgements, behaviour, and the like) in terms of '*rather than*' *clauses*: I regularly go shopping on Saturdays rather than Mondays (bodily movements); I genuinely believe people should be free to decide their religion rather than being enforced by someone else (family) or outer institutions – e.g., State (beliefs); and John Denver would have preferred to be in Colorado rather than somewhere else in the world – at least this is what he sang in his classical 1971's country hit (preferences). The contrast between what is the actual case and the unactual set of alternatives is the key message of contrastive explanations.

The idea dates from Fred Dretske (1972), but it gained more attention and popularity in the philosophy of science after Bas van Fraassen (1980) accommodated it as an important ingredient in his Constructive Empiricism. Van Fraassen suggests that scientific explanations are neither (a) arguments nor (b) express a singular relation between theory and reality. Instead, explanations, including the scientific ones, are speech acts heavily dependent on contextual factors (Wilkinson, 2014, p. 2). They are answers to why-questions that require relevant descriptions that serve a contextually determined purpose.

For van Fraassen, why-questions are interrogations that imply tacit declaratory sentences. Questions like, for instance, "Why did Russia invade Ukraine?" or "Why do leaves get yellow in autumn?" entail that, from the questioner point of view, Russia had invaded Ukraine, and leaves get regularly yellowed in autumn. These real-world events are what is being called for an explanation, for what reason they occurred rather than any other alternative scenario.

They are what van Fraassen (1980, p. 141) calls the "topic of the question", declaratory propositions equivalent to the interrogations from which they derived.

On van Fraassen's semantics, a why-question Q is identical with a triple $\langle P,X,R\rangle$, where P is a proposition which holds a truth-value (the topic of the question), X is the set of propositions where P is included alongside the relevant non-actual alternatives and the only true member in X (the contrast class of the question Q), and R is a (contextually) determined relation of explanatory relevance between a proposition and the topic/contrast-class pair $\langle P,X\rangle$. For example, let's say we aim to explain this canonical biblical question (van Fraassen, 1980, p. 127):

(1) Why did Adam eat the apple?

Contrastive accounts show in a straightforward manner why this query is ambiguous. For example, the inquirer might have meant any of the following explanatory requests:

(1a) Why did Adam eat the apple, rather than throw it away?
(1b) Why did Adam eat the apple, rather than any of the other fruits?
(1c) Why did Adam, rather than Satan, eat the apple?

A suitable answer to (1) depends on which one of the questions (1a)–(1c) was meant. A statement such as "Because he was hungry" is a possible answer to (1a), but not to (1b). It counts as a satisfactory answer to (1c) just in case we know that Satan was not hungry. Since the explanatory requests (1a)–(1c) should be answered differently, it is reasonable to assume that the inquirer asks us to explain different things. But if that is so, then question (1) underdetermines the explanandum. Van Fraassen, therefore, concludes that inquiries like (1) only appear to posit non-contrastive explanandum; they are actually ambiguous, and once disambiguated, it becomes clear that the actual explanandum is a fact embedded in a contrast class, a set of alternative events that did not come true. Contrasts are crucial to disambiguate the question. Here is another example, suggested by Alan Garfinkel (1981).

Willy Sutton is jailed in virtue of having robbed a bank. When, during a visit to the prison, Sutton was asked by a priest why he robbed the bank, he replied: – "Because that is where the money is". Garfinkel indicates that while Sutton explains why he robbed banks rather than other possible venues (jewel stores, gas stations), the priest was seeking to understand why Willy robbed banks rather than leading a morally good life. From Sutton's point of view, to say he robbed a bank because this is where most of the money is located constitutes a satisfactory

answer. He externalises the intentions for his action at stake. This is so since Sutton *presupposes* this is what the priest is seeking to know. From the priest's eye, Sutton's answer did not explain what he was inquiring about; he wants to know why rob anything at all. The original question, "Why did you rob the bank?" turns out to be ambiguous. To disambiguate it, we need to add contrasts. Without the contrasts, we cannot know what is being inquired about and thus appropriately respond to it.

Explanation, for van Fraassen (1980, p. 129), is a three-place relation between theory, a fact, and a *context*. A why-question does not arise independently of a context. The context fixes the contrast class and therefore a why-question is ambiguous until the appropriate contrast is inserted. Once the contrasts are brought to light, we can formulate the explanation in the form of *c* explains *e* rather than *e**, where the contrast stands on the e-side.

Suppose that X= {P, P_1, …, P_k, (…)}, and that P is not one of the P_ks. Where A is any proposition, van Fraassen (1980, 144) defines A as a direct answer to Q relative to X to be any proposition having the following truth conditions:

(2) P and, for all $k \geq 1$, $\sim P_k$, and A.

The standard word of a direct answer to a question Q uses the word 'because' in place of the second logical conjunction "and" in (3):

(3) P, in contrast to the rest of X, *because* of A.

On van Fraassen's view, the word "because" is simply Boolean conjunction (i.e., the truth value of the final conjunction typically depends on the truth value of the other operators involved) as indicated in (2): P must be true and all the potential alternatives to P (P_1, P_2, …) must be false (therefore, $\sim P_k$). In (3), the word "because" plays the pragmatic function of indicating that (3) is being employed for an explanatory ambition, not to give a non-truth-functional dimension to the truth conditions of (3). Proposition A is said to be relevant to Q *if and only if* A bears the relevance relation R to $\langle P, X\rangle$. To ask why P is to ask for a reason, and R varies according to the kind of reason that is being requested by the inquirer according to the context. The contrast class (which can be finite or infinite) facilitates us to distinguish different why-questions that have the same topic.

Consider, for instance, the scandal during the last edition of Oscar's Academia Movie Award when Will Smith left his seat and walked toward Chris Rock to slap him. What everybody

witnessed on that night is that Will Smith became furious after the presenter joked at his wife, Jada Smith. Let's portray our why-question:

(4) Why did Will Smith slap Chris Rock during the Oscar?

We already saw that, in the contrastive account, questions like (4) are non-sufficiently specific. When we add appropriate contrasts according to the context, we can break the question (4) into (at least) three more accurate why-questions:

(4a) Why did Will Smith, rather than another actor, slap Chris Rock during the Oscar?
(4b) Why did Will Smith slap Chris Rock on the Oscar set, rather than on the Red Carpet?
(4c) Why did Will Smith slap Chris Rock, rather than dialogue with him?

All these questions have a common general topic –Will Smith's aggression against Chris Rock. But when we add contrasts, it becomes clear we have different questions demanding different explanatory requests about the same event. Each question in place entails a different explanandum. For example, question (4a) entails the statement M = "Will Smith, rather than another actor, slapped Chris Rock during the Oscar". So, M is the topic of question (4a). Possible contrasts to the explanandum M may be M_1 = "Javier Barden slapped Chris Rock during the Oscar", or M_2 = "An individual S from Oscar's security staff slapped Chris Rock during the Academy Award". For question (4a), propositions M, M_1, and M_2 are all members of the contrast class X in which only M holds the truth value. We are in search of the relevant explanans of M. According to what has been widely reported, a direct answer to (4a) is

A = Because Smith's wife was offended by Chris Rock.

The target of the comedian's joke was Jada Smith rather than, for example, Penelope Cruz (the actual Javier Barden's spouse that was also present in the Award). A holds a relevant relation with M but not for M_1 or M_2. So, we may say A (the explanans) explains why M was the case, rather than M_1 or M_2 – A has the relevant explanatory relation R to $\langle M, X\rangle$.

In the second question, the interrogation (4b) entails the statement N = "Will Smith slaps Chris Rock during the Oscar, rather than somewhere else". According to the context, plausible candidates of unactual events may be N_1 = "Will Smith slapped Chris Rock on the parking lot", or N_2 = Will Smith slapped Chris Rock on the street". Like the question (4a) above, these three

statements ⟨N, N_1, N_2⟩ are all members of the contrast class Y where only N holds the truth value. A possible answer to (4b) is

B = Because Jada Smith was offended in the Oscar ceremony.

Again, B is a relevant answer to N, but not to non-actual events N_1 or N_2. B holds the relevant relation R to ⟨N, Y⟩. So, B, the explanans, explains N.

The third question (4c) entails the statement O = Will Smith slapped Chris Rock, rather than dialogue with him". Other alternative scenarios may be O_1 = "Will Smith whispered something to Chris Rock", or O_2 = "Will Smith showed dissatisfaction to Chris Rock". A reasonable answer to (4c) is

C = Because Will Smith has explosive behaviour.

These three statements are members of a contrast class Z. C holds a relevant relation R to O, and not to any other unactual events (O_1 and O_2). C has the relevant relation R to ⟨O, Z⟩. Thus, we may consider C as the explanans of O: C explains O rather than O_1 or O_2.

2.1. Evaluating the answers

According to the contrastive view, a why-question presupposes (i) that its topic is true, (ii) that only its topic is true in the contrast class, and (iii) that at least one proposition bearing the explanatory relevance relation R to the topic/contrast-class pair is true. All these three clauses are respected in the abovementioned example.

A, B, and C – the three explanans of the questions (4a)–(4c) – respond to different aspects of the same topic. In the case of A, it answers which fact in the world has triggered the fury of Will Smith – the joke of Oscar's host at Jada smith. A is a satisfactory answer to (4a) but not to (4b) insofar as it does not provide relevant information about the location of the aggression – which is the topic of the question (4b).

B, in turn, responds to another aspect of the same topic: the location issue. It puts light on Smith's explosive behaviour. While B is a good answer for N, it is not a good answer for alternative contrasts like N_1 so long it may suggest that the actor may have planned what may be the best place for the aggression, calculating likely harmful implications for his image and career.

Finally, *C* highlights the psychological aspects of Smith's behaviour. Apparently, from Smith's perspective, only an aggressive behaviour could sufficiently respond to that level of offense that his wife had undergone. Other degrees of behaviour, as represented by O_1 or O_2, was not enough. On that night, Will Smith had time to reflect on what he was about to do: he left his seat and walked a couple of meters until he slapped Chris Rock. He could have stopped along the way, but he did not.

Interestingly, the relata of explanans *A* is a contingent, distinct, and contiguous event temporally connected to the explanandum while explanans *C* is a *condition*, a psychological *disposition* of Will Smith's behaviour. While *A* is a clear causal explanation of *M*, *C* (the explanans "Because Will Smith has an explosive behaviour") complements our understanding of why *M* has happened by supplementing structural information about the psychology of Will Smith, an answer whose relata is not an event – the canonical standard relata of causal explanations. *C* is not a contributing cause either – a cause that is not sufficient although it is necessary to bring about the final effect; it sounds more like a structural condition that, under certain stimulus conditions, activated the aggressive behaviour of Smith, favouring the rise of *A*.[89]

This example illustrates two important things about the contrastive view. First, depending on the context, contrastive explanations are not restricted to the canonical causal explanations relating two actual and distinct events. Other types of relata are perfectly welcome by contrastive accounts. In paradigmatic cases, the contrastive account can be employed to represent examples of noncausal explanations. Second, the inclusion of contrasts reveals that the simplest questions can be broken into more specific interrogations, showing that, as Hempel (1965, p. 421-423) says, scientific explanations only answer to certain *aspects* of events. The contrastive account highlights this feature of explanations in a very appropriate manner.

As we saw, the Will Smith's scandal met all the three conditions claimed by van Fraassen: (i) the topic of the question is true, (ii) only the topic, rather than any of the propositions from the contrast set, is true, and (iii) there is at least one relevant proposition *A* that is true. When (i) or (ii) fails in virtue of the contextually determined background knowledge in play does not imply both (i) and (ii), the why-question does not arise. When the (iii) fails, then the why-question has no proper answer even if it arises.

For instance, suppose that paresis hits someone who has untreated syphilis. Given a universe of ten people who have untreated syphilis, and exactly one of them, Carl, got paresis,

[89] For some philosophers such as Armstrong (1968), Block (1990), and Jackson (1995), dispositions are not causal or causally explanatory for their manifestations. In contrary, Bird (2020) argues the triggers of dispositions are the causes: *C* causes *E* when *C* is the stimulus of some disposition of which *E* is the manifestation. Such account suggests that causes are conditionally sufficient (rather than necessary) for their effects. I leave this question entirely open. My point is to demonstrate how contrastive accounts accommodate multiple relata.

there may be no answer to the question "Why did Carl, rather than any of the other nine, contract paresis?" Since paresis develops indeterministically from syphilis, nothing favours Carl – in contrast to the other nine syphilis patients – as being more likely to develop paresis.

Contrastively, if Julie and Hilary had syphilis, the question "Why did John, rather than Julie and Hilary, develop paresis?" does have an answer: "John developed paresis, rather than Julie and Hilary, because John had syphilis, but Julie and Hilary did not". As in the first example, the why-question in this case requests causal factors that led to John's getting paresis while the others figured in the contrast class did not. Thus, in both cases, the same relevance relation *R* is in charge because they request of the same kind of information, namely the causal factors leading to the truth of the topic rather than the other members of the contrast class.

If there are no such causal factors, like in the first case of paresis, the question is to be dismissed. If, as in the second scenario of the paresis case, there are such causal factors, so that at least one proposition holds the relevance relation to the topic/contrast-class pair, then a potential answer *A* is judged according to three criteria: (a) how acceptable or likely *A* is, (b) the degree to which *A* favours the explanandum over other members of *X*, and (c) whether *A* is made irrelevant by other rival answers (i.e., alternative explanans).

2.2. Ternary relation isn't enough: the quaternary approach

Some scholars like Schaffer (2005) and Northcott (2008a), argue that causation is a quaternary relation rather than ternary only because causes need contrasts too. This approach is defended on the grounds that the quaternary approach adequately responds to many paradoxes around causation, such as the causation by absences.[90] Plainly put, causation is a four-placed relation in the form *c*-rather-than-*C** determines *e*-rather-than-*E**. Causation includes, in this account, 4 slots: two for events that occur – the explanans cause and the explanandum effect – and another two dedicated to their contrasts. Sentences in the form "*c* causes *e*" semantically express different propositions under different contexts. What shifts from one context to another are their contrasts, where contrasts represent salient alternatives both to the actual cause *c* and the effect *e*. In fact, behind the curtains, the binary form of causation implicitly communicates a double contrast on both sides: {*C**} and {*E**} are pairs of non-empty sets, that must be filled by unactualized events added to these sets. What contrasts an utterance of a causal sentence expresses depends on the questions structuring the causal inquiry and the

[90] For more detailed discussion over the advantages of quaternary approach of causation, see Shaffer (2005), Gallow (2022, Sub-section 1.2.2).

alternatives explicitly (or implicitly) evoked by the conversational participants. When one says, "*c* causes *e*", one speaks elliptically and relies on context to supply the two implicit relata *C** and *E**.

As indicated above, one of the advantages of the quaternary approach is to respond satisfactorily to examples of causal explanations where the explanans is a non-actual event. Realists and physicalists about causation reject causation by absence. If absences are not actual events, then they are not physically connected to the events that we often speak of as their effects. So, the supposed causal factor, a non-actual event in the world, is not contiguously linked by a causal process to what supposedly triggered it. David Armstrong (1999, p. 177), for example, says: "Omissions and so forth are not part of the real driving force of nature. Every causal situation develops as it does as a result part of the presence of positive factors alone". A realist about causation like Armstrong would question anyone who trusts in causation by absence about how such an absence can be causally efficacious given that absences do not initiate a causal process leading to the effect. Consider the daily situation below.

At my home, Friday nights are called "Pizza's night". My wife and I are passionate about pizza and every Friday we have pizza for dinner. Last Friday, we welcomed ten friends for our Pizza night. Denise wondered if two bottles of wine would be sufficient for the entire night. So, she put the first pizza into the oven and said: - "Alex, keep your eyes on the oven, please. I'll go to the supermarket and be right back in 15 minutes". It was my duty to be aware of the timing. In the meantime, I received a call from my aunt. My phone was charging in the living room and someone next to it alerted me. So, I left the kitchen to answer it. To not disturb our guests, I decided to answer the call in the master room. My aunt and I have started chatting and (sadly) I simply forgot there was something being cooked in the oven. I know you can imagine what has happened to the pizza: it got burnt. Every common-sensical person who heard this story would blame me for letting the pizza get burnt. It makes perfect sense to say that "Alex not removing the pizza has caused its burning". If I were aware of and paid attention to the oven, the pizza would not get burnt; the result was more than expectable, was predictable.

On a contrastive account, we can represent this Friday night fatality by the statement "Alex not removing the pizza from the oven right on time caused its burning". This worldbound occurrence brings implicit contrastive unactualized events. For example, we can say "Alex chatting with his aunt, rather than looking after the pizza, caused the pizza's burning, rather than perfectly baked". My long-lasting phone call in my room rather than paying attention to the oven is causally significant for the effect, whereas being on a phone call rather than playing

videogame did not. For the quaternary view, an accurate structure of causation is given by the form "*c*-rather-than-C_0 causes *e*-rather-than-E_0 *iff* C_0 counterfactually entails E_0":

> We have reached the following picture: metaphysically, causation is defined on 4-tons (c, e, C_0, E_0). For any given such 4-ton, whether c causes e depends on whether C_0 and E_0 are members of $\{C^*\}$ and $\{E^*\}$ respectively, and also on whether C_0 counterfactually entails E_0. (Northcott, 2008a, p. 121)

If an explanation is an answer to a why-question, a description that serves a contextually determined purpose capable of satisfying our explanatory interests, then we can extend the quaternary view of causation to explanation. My disastrous Friday night can be represented by the following why-question:

(5) Why did the pizza get burnt?

This question, according to contrastive view, is non-specific enough. We should therefore add the contrast:
(5a) Why did the pizza get burnt, rather than perfectly baked?

When we turn the contrast explicit, we crystalise the topic of concern entailed by (5a) in the form of the proposition *P* = "The pizza was burnt, rather than perfectly baked". As I already recognized, I should be blamed for that result, insofar as my distraction has generated the effect. From a quaternary view, the structure of the explanation would follow this pattern:

(6) *P*, in contrast to the rest of *X*, because of *A*, rather than *B*.

The quaternary view adds contrasts on both sides: *P* is the explanandum and *X* is the contrast slot on the e-side, while *A* is the explanans and *B* is the contrastive slot on the c-side. In my example, the direct answer to question (5a) is *A* = "The pizza got burnt rather than perfectly baked because Alex was on the phone in his room rather than in the kitchen". This is a direct answer for (5a) that, following the quaternary approach represented by (6), is equivalent to the bracket:
C = Alex answered the phone in the master room.
E = Pizza got burnt.
C_0 = Alex was standing in front of the oven.
E_0 = Pizza was perfectly baked.

This scheme has two slots for the events and another two for non-actual events. Contrast sets permit counterfactual pairs like {C_0, E_0} that denote individual contrast events. While C and E are not actually connected by a causal chain, the contrasts *would be*. There is a would-be connection (Shaffer, 2005, p. 332) relating C_0 and E_0: if C_0 was the actual event, the desirable result E_0 would be brought about. E_0 is counterfactually dependent on C_0 – the latter entails the former. The non-actual event C_0 exerts a positive difference-making for the rise of the alternative explanandum E_0. So, we can say that C is the explanans that answers why explanandum E occurred.

The relevant explanatory relation between the explanans and the explanandum is given by the counterfactual entailment between the two. So, the explanation of the event E is given by the explanans C = "The pizza got burnt rather than perfectly baked because Alex was chatting on the phone in the master room rather than in front of the oven". My absence from the front of the oven denotes an *actual* event (my conversation with my aunt on the phone in the master room) with positive effects for the occurrence of the outcome. Similarly, an analogous example is that

> […] given that the gardener napped and my flowers wilted, 'The gardener's not watering my flowers caused my flowers not to blossom', is to be interpreted as: 'the gardener's napping rather than watering my flowers caused my flowers to wilt rather than blossom'. (Schaffer, 2005, p. 331)

Under God's eye, there are, metaphysically, loads of causes for a given effect. In this example, I have been blamed for the result, but someone could reasonably argue it is fair to share my responsibility with the other ten individuals who were hanging out in the living room. While I was on the phone, Toni, Jonathon, and the others were constantly passing by the kitchen to get drinks and snacks. So, they could also prevent the event right on time, right? This is the metaphysical point: we could equally distribute the responsibility among all the eleven individuals (ten guests plus Alex) who were perfectly capable to avoid the occurrence of the explanandum (this problem is what we named the proliferation of causes).

The other side of the coin is that, epistemically, a why-question comes up from the context and depends on our epistemic *interests*. From an epistemological point of view, some of the causes will be relevant but not *all* the others. That night, there was only one person who was (socially) obliged to keep their eyes on the oven: Alex. I am the householder while the others were only guests. For my wife, no one in the living room was causally responsible for letting the pizza get burnt except me, who should be aware of and knew from the very start something was being cooked in the oven. If she would be asked why the pizza get burnt, she would not hesitate to respond that "Because Alex was not paying attention to the oven", an answer that I would

highly subscribe to. So, the contrastive account provides a pragmatic reason why we should welcome causation by absences.

Here is a toy example of causation by absence from economics. We can see that it can be framed using the quaternary contrastive structure. In the mid-2010s, influential economists and politicians alerted that the trajectory of public deficit in Brazil was far beyond control. Their diagnosis was that, since the early 2000s, public debt in Brazil has gradually escalated and without an institutional intervention, nearby in the future, the government should inevitably raise taxes to alleviate the pressure of public deficit, provoking thus an enormous impact on economic growth, compromising the lives of future generations. Therefore, sooner or later, the hard work would have to be done.

To respond to this issue, in December 2016, the Brazilian parliament approved Amendment 95. From December 2016 up to 2036, public expenses (e.g., investments in health care, social security, education, and social care) shall be frozen – the government expenditures cannot grow above the rate of inflation, not even if the economy is performing well over these years (Roznai and Kreuz, 2018, p. 38). For those who subscribed to the inclusion of this Amendment in the Constitution, the absence of a rule that limits the governmental disbursements was favouring the escalation of public deficit. Following the quaternary view of explanation and the discussion about this issue, we may represent the situation before 2016, i.e., before the Amendment was brought in, as follows:

C = Actual legislation about public expenditures permit governmental investments over and above the rate of inflation, which has historically allowed a gradual and intolerable elevation in public deficit.

E = Future economic growth shall be compromised.

C_0 = Public deficit is tolerable.

E_0 = Future economic growth is not compromised necessarily.

So, explanandum E is about what Brazilians may probably experience in the future in the absence of a rule that formally limits public expenditure based on the diagnosis about the actual situation represented by C. This is the mainstream interpretation of the fiscal problem in Brazil that justified the introduction of Amendment 95. Gradual and persistent elevation of the public deficit is a problem and only an institutional response could render control over the public expenses over time. C entails an absence event C_1 = the absence of a rule inhibiting the public expenses facilitates the persistent increment of public deficit.

We earlier saw that my absence from the front of the oven was entailed by a positive actual event – my conversation with my aunt on the phone in the master room. Similarly here, before 2016 the absence of Amendment 95 was entailed by the actual legislation which permitted government investments over the rate of inflation. Thus, this example has a similar structure to the analysis of absence causation that I gave in the pizza example. In both cases, reference to purely negative events can be avoided.

Interestingly, in this case, the contrastive scenario represented by $\{C_0, E_0\}$ is the usual heterodox interpretation of the problem. Mainstream scholars often fear that a gradual, consistent, stable, and permanent elevation of the deficit may lead society to default in the future. Heterodox scholars retort, arguing that countries that issue their currency (USA, UK, Japan, Australia, etc.) do not experience the usual constraints argued by mainstream economists. It is the federal government, not the taxpayers, that finances all the governmental expenditures. So, the government does not need to get the taxes first to then expend.[91] Despite the details involved in each of these positions, the contrast between these two explanations shows that the quaternary approach can serve as an effective tool to compare alternative explanatory hypothesis in economics even in the context of absent causes.

3. Challenging the origins of capitalism: a contrastive test

In chapter 2, I argue that mechanistic-based explanations can provide solid causal and constitutive information about some target phenomenon. In a case in which our phenomenon of interest belongs to a higher level of organisation, the mechanistic approach can inform why different items placed at different levels within the mechanism are causally efficacious and must be (constitutively) active for its entire performance. When we adopt Woodward's (2003) counterfactual view to understanding causal and noncausal relations inside a mechanism, we visualise why we can accommodate the explanatory power of higher-level events – i.e., cases where these events function as the explanans – with reduction. It's a matter of looking at different aspects that contribute to the rising of the same phenomenon: sometimes, the inquirer wishes to identify and understand which previous events and processes contribute to its rising; in other contexts, she aims to understand which parts and components must be active to produce the entire manifestation of the phenomenon under scrutiny.

By embracing the mechanistic framework, we can convincingly attend to Weber's intellectual intentions and prescriptions. First, we can demonstrate that the birth of capitalism

[91] Stephanie Kelton's *The deficit myth* (2021) has already become a best-seller book in these discussions.

is counterfactually dependent on the large influence of a previous actual and distinct large-scale event – the heavy influence of Protestant doctrine upon individuals: if the latter were absent, the former would not have happened. Similarly, the capitalist system counterfactually depends on the crucial units/parts to be generated – capitalists and labourers. If the parts of the economic system were different – i.e., if a major part of the means of production was not under control of private agents and there were no individuals who sell their labour – then the very economic system would be different, gaining thus a different qualification. In doing so, we indicate how the general recipe of methodological individualism is not incompatible with the explanatory power of higher-level events.

For assessing examples of historical explanations like Weber's explanatory hypothesis for the rise of modern capitalist societies, we should, in case we select Woodward's counterfactual view, run some ideal tests. On a counterfactual account, Weber's explanation runs like this: if Protestantism had not been highly influential in the British social setting of that time, modern capitalism would not have emerged during the 16th century in England. Once we modify the explanans in various ways and the explanandum is also sensibly affected, we are entitled to recognise the explanans as the triggering causal factor for the explanandum. So, the protocol for the analysis of a counterfactual-based explanation prescribes we should first manipulate the *context,* and then we similarly run some counterfactual tests as required by the contrastive scholars:

> Woodward's (2003) manipulationist theory of causation […] implies a contrastive view of token causation. […] Denoting the post-intervention value of the cause variable as c and its pre-intervention value as C_0, and the post- and pre-intervention values of the effect variable likewise as respectively e and E_0, the resultant definition of causation then appeals (from the standpoint of after the intervention) to exactly the same counterfactual C_0-E_0 as does our own definition. (Northcott, 2008a, p. 116)

Thus, although a mechanistic-based approach and contrastive view have been born as independent accounts of scientific explanations, a mechanistic explanation that adopts a (monist) counterfactual approach to assess the relation across the nodes within the mechanism in counterfactual terms turns out to be aligned with the purposes of the contrastive view of explanation.

To the problem. In his classical *Protestant Ethic and the spirit of capitalism*, Weber tries to respond to the following inquiry:

(7) Why did modern capitalism surge in the UK, the US, and the Netherlands in the 16th century?

This why-question entails that the modern capitalist system (a mode of production with a disciplined labour force and a regularised investment of capital) emerged in the UK, US, and the Netherlands during the 16th century – this is the topic of the question assumed as true. Historians and social scientists dispute which social aspects constituted the triggering cause(s) for the birth of modern capitalism in Britain at that Epoque, whether modern capitalist societies date even farther than 16th century – undermining Weber's hypothesis, or if other countries of that era also presented identical capitalistic aspects.

According to Weber,

A = The presence of Ascetic Protestantism exerted a massive contribution to the birth of modern capitalism

where the presence of Protestantism (explanans) and the rise of modern capitalism (explanandum) are (a) two actual and distinct macrosocial events and (b) temporally contiguous. Weber did not cast his research problem in contrast terms. By using the quaternary approach and adding contrasts to question (7), we break Weber's topic of interest into a set of more specific questions such as:

(7a) Why did modern capitalism emerge in the 16th century rather than earlier?
(7b) Why did modern capitalism emerge in the UK rather than in Italy?
(7c) Why did modern capitalism, rather than traditional capitalism, emerge in the UK in the 16th century?

Question (7a) entails the proposition P = Capitalism emerged in the UK in the 16th century – the topic of the question. We can picture a direct answer to (7a) in the following way:

C = Heavy influence of the Ascetic Protestant religion.
E = Modern capitalism was born in the 16th century.
C_0 = Heavy influence of Catholicism.
E_0 = Modern capitalism was in charge in the 14th century.

Hereby, the pair $\{C, E\}$ represents Weber's explanatory hypothesis. These two actual and distinct events are heavily correlated. In this question, the quaternary approach serves to confront Weber's hypothesis. A social scientist who shows evidence suggesting that ($C_0 \rightarrow E_0$) compromises Weber's hypothesis.

Question (7b) entails the proposition P = Capitalism emerged in the UK rather than in Italy. We can picture a direct answer to (7b) in the following way:

C = Heavy influence of the Ascetic Protestant religion.
E = Modern capitalism emerged in the UK in the 16th century.
C_0 = Heavy influence of Catholicism.
E_0 = Modern capitalism was present in Italy in the 16th century.

Like the first scheme, {C, E} represents the Weberian hypothesis while the hypothetical sets {C_0, E_0} are the contrast classes. In (7b), a social scientist shall contend that Italy's renaissance of the 1500s was a typical capitalistic society. Finally, (7c) is portraited as follows:

C = Heavy influence of the Ascetic Protestant religion.
E = Rise of modern capitalism.
C_0 = Heavy presence of Catholicism.
E_0 = Presence of traditional capitalism.

Interestingly, (7a) - (7b) serve to challenge Weber's hypotheses via historical facts by showing that modern capitalism was already existent before 16th century (*time*) and possibly co-existed during that century in different countries where Protestantism was only insipient (*region*). (7c), in turn, puts under discussion Weber's categories of "modern capitalism" vs. "traditional capitalism". A social scientist who offers a different definition of such categories may contend (7c) directly and (7a) – (7b) indirectly.

Traditional capitalism is, according to Weber, the general characteristic of southern European socio-economic systems where most individuals worked to maintain their standard of living rather than striving to accumulate surplus wealth. For instance, in the agricultural sector, at a time when the labour force was the main cog of production, the profit margin was massively dependent on the velocity of harvest. It was the interest of landowners to speed-up the harvest in the face of the uncertainty of the weather. In this system, workers (peasants) got paid per unit: the more you gather, the more you earn (in a system known as piece-rates). Many times, landlords increased the value of piece-rates (the amount of money to be paid per unit) in attempting to stimulate a faster harvest. But it simply did not work. It seems that

> […] the opportunity of earning more was less attractive than that of working less. He did not ask: how much can I earn in a day if I do as much work as possible? but: how much must I work in order to earn the wage, $2^{1/2}$ marks, which I earned before and which takes care of my traditional needs? This is an example of what is here meant by

> traditionalism. A man does not "by nature" wish to earn more and more money, but simply to live as he is accustomed to live and to earn as much as is necessary for that purpose. (Weber, 1992, p. 60)

In modern capitalism, in contrast, individuals look after profit rationally in a systematic and organised matter. Labourers developed a cool self-control and sense of frugality, which increased their productivity enormously. Entrepreneurs, in contrast, reinvest the profit accumulated in the enterprise as a means of future gain. Monetary gain was not to be put to primarily satisfy the urges of gluttony, sloth, lust, ostentation, or luxury; money is at the service of future returns (Poggi, 1983, p. 41). The different patterns of behaviour among individuals in these two different types of capitalism are associated, among other things, with people's beliefs. While in modern capitalist societies the doctrines of Ascetic Protestantism were dominant, in traditional capitalist societies they were not. Protestant doctrines, with their cluster of principles and values, induced methodical ways of living in believers.

It is worth noting that Weber did not characterise the presence of Ascetic Protestantism as *the* cause; this was a necessary but not a *sufficient* element of the causal chain (Weber,1992, p. 27). As highlighted by Anthony Giddens in the introduction of *Protestant Ethic and the Spirit of Capitalism*, other important macro social aspects for the birth of capitalism were:

> 1. The separation of the productive enterprise from the households [...] 2. The development of the Western city. [...] 3. The existence, in Europe, of an inherited tradition of Roman law, providing a more integrated and developed rationalisation of juridical practice. [...] 4. This in turn was one factor making possible the development of nation-state. [...] 5. The development of double-entry book-keeping in Europe. 6. That series of changes which, as Marx emphasised, prepared the way for the formation of a 'free' mass of wage-labourers, whose livelihood depends upon the sale of labour-power in the market. [...] Taken together, these represent a mixture of necessary and precipitating conditions which, in conjunction with the moral energy of Puritans, brought about the rise of modern Western capitalism. (Giddens, 1992, p. xvii-xviii)

In the mosaic of causes that contributed to the birth of capitalism in Britain, the Protestant culture, in Weber's view, *weighted more* than other things. The task of historians and social scientists is to detail which, among this cluster of contributing causes, were more relevant, a task that the contrastive account can once again do service for, but it is far beyond the scope of this dissertation.[92]

All the schemes above contend Weber's explanans on the macro level to test the idea of whether Protestant Ascetic is among the main causes of modern capitalism. At this stage, historians would be testing the explanatory power of that macrosocial event and debate whether the concept adopted by Weber is appropriate enough. So far, the debate circles around

[92] The application of contrastive account to evaluate the weight of explanations in history is provided by Northcott (2008b).

the main event that was potentially responsible for the rise of modern capitalism. But historians are also interested in the internal processes that decisively engendered the birth of the capitalist system in the UK. In so doing, they are not looking after *why* responses but *how* this entire complex social phenomenon called modern capitalism was possible. While the usual relata of why-answers are *events*, how-answers is more about types of *processes*:

> In answering how we usually take as our explanandum not an event but a process (in a broad sense of "process"), e.g., the coming about of an event, the working of a machine, the evolution of a social institution, etc., and rather than characterize the external causal *surroundings* of the process (as we do for events) we characterize its internal *makeup* – we say how the process works. (Cross, 1991, p. 246)

Rather than seek for reasons, how-answers are looking at in which ways the target of interest, the explanandum, was possible:

> (1) *By what road* (How did you get there?)
> (2) *In what manner* (How did you behave at the party?)
> (3) *By what argument* (How will you justify this?)
> (4) *By what method* (How do you perform an appendectomy?)
> (5) *By what means* (How did you get that money?)
> (6) *In what respect* (How do these differ?)
> and last but not least:
> (7) *By what process* (How do DNA molecules duplicate?) (Cross, 1991, p. 248-249)

By answering how a given phenomenon such as the rise of capitalism was possible, we shed light on the internal components that contribute to its rising. So, the how-question is:

(7d) How did capitalism emerge in the UK during the 16th century?

A = By the influence of Protestantism, many individuals across Britain's Elizabethan Era developed capitalistic behaviours.

According to the quaternary view and following what has been argued in the chapter 2, we have something like this:

C = Heavy presence of Ascetic Protestantism
E = People embodied modern capitalistic behaviours
C_0 = Absence of Ascetic Protestantism
E_0 = People did not behave like typical modern capitalists

Both pairs {C, E} and {C_0, E_0} are counterfactually channelled. While in chapter 2 the mechanistic representation of the rise of capitalism breaks the bottom-line (the level of individuals) into more relations of determination – the transformational steps from (i) people's beliefs into behaviours and thus (ii) to people's actions to (iii) macrosocial states – the

contrastive account abstracts away these aspects. It only shows that the bottom line is a necessary mediator variable that channels the macro causal state to the macro effect. It contributes to reinforcing the explanatory power of *C* as one of the main causes (the explanans) for the rise of capitalism. Sometimes, the contrastive account, in comparison with mechanistic explanation, doesn't boost our understanding. This latest explanation (7d) omits one of the nodes from the mechanism (transformational step from people's intentions to behaviours) or, at best, it turns the information over the transformational step implicit. In other cases, the mechanistic and contrastive explanations can attain similar results, as the Lucas's example demonstrates in the lines below.

4. Lucasian aggregate supply: it's the rational expectations, rather than adaptive expectations

In the third chapter, I argued that Lucas's Phillips curve is both consistent with Nagel's and constitutive-mechanistic accounts. The model does not represent a historical chain of causes nor reconstruct a sequence of economic events based on individual agency. Rather, it aims to illuminate how relevant components within an economic system, via their causal features, may possibly determine the manifestation of a given macroeconomic phenomenon.

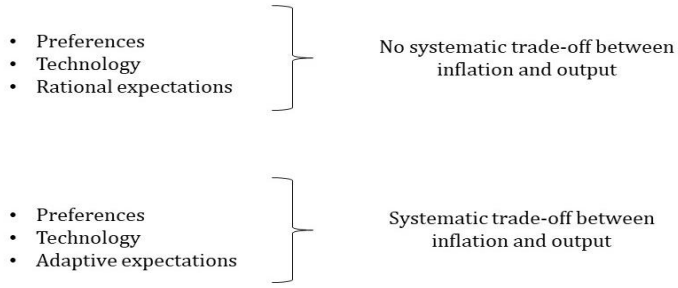
Between the two selected frameworks to unlock reduction in economics, it is the second one, the mechanisms, which is immediately suitable for contrastive questions. In my reconstruction of Lucas's model in mechanistic terms, I illustrated why the Phillips curve is counterfactually dependent on people's (rational) expectations: if people, while forming their expectations, do not consider the economic policy as a decisive informational input, they would commit systematic mistakes. But people are not stupid. They do not discard relevant information capable of diminishing their chances of making mistakes.

Therein, I argue that the sensible contrasts are between two distinct theoretical paradigms about how economic agents form expectations: either (a) they collect all available information and use it economically (rational expectations hypothesis, REH) or, for instance, they (b) base their expectations about some relevant macroeconomic variable on past information only (according to the adaptive expectations hypothesis, AEH). Both paradigms, REH and AEH, accept the stability of behavioural parameters: both tastes and technology are relatively fixed across the time.

According to Lucas, the Phillips curve emerges from a combination of causes: (a) the parameters of tastes and technology are relatively insensitive to changes in the environment (they are the exogenous "givens") plus expectations which, in REH, is an endogenous variable.

In short, we have (a) a (exogenous) parameter and (b) the (endogenous) expectations figuring in the explanans, and the outcome (no systematic trade-off between inflation and output) is the explanandum. By contrast, AEH assumes that parameters and expectations are equally exogenous, which generates a different consequence (systematic trade-off between inflation and output):

Figure 5: The contrasts between two theoretical paradigms



Parameters common to both paradigms fix the necessary causal background where we contrast two alternative explanations (Marchionni, 2006). Then we subjugate both paradigms to the same background conditions and compare the results. REH and its foil, AEH, manifest their causal powers conditioned to the same background conditions. While the parameters are *conditions*, REH and AEH are the *difference-makers*: by modifying the difference-makers, i.e., by switching the difference-making variables in both explanans, we change the corresponding outcomes. Contrastively, the structure of this explanation for this model is

(8) *E* rather than *E**, because of *C*, rather than *C** conditioned to *B*, where:

B = Fixed parameters
C = Rational expectations
E = No systematic trade-off between inflation-output

B = Fixed parameters
*C**= Adaptive expectations
*E** = Systematic trade-off between inflation-output

The scheme above is equivalent to explanation *A* below:

A = There's no systematic trade-off between inflation and output, rather than a systematic trade-off, because people form their expectations rationally, rather than in adaptive terms, conditioned to the stability of tastes and technology.

Economic models *a la* Lucas function as a theoretical experiment where the modeller may test different hypotheses subject to the same background conditions. The method of verification is the manipulation of the relevant variables supposing the invariance of parameters. In the contrast between REH – AEH, we learn that it's only the former, not the latter, that is consistent with economic theory and the maximisation principle; consistent expectations should be consistent with economic theory (Walters, 1971). What the model does is to give us reasons to believe that REH, rather than its rival, is true, at least in the abstract. It sheds light on how the causal cogs of an imaginary social mechanism may produce a macroeconomic phenomenon. As suggested by Cross (1991), this is a typical case where the modeller investigates how-questions: they explore *by which mechanism* the Phillips curve, rather than no Phillips curve, is eventually possible.[93] Consider we formulate this question:

(9) How is the Phillips curve possible?

According to the story that Lucas tells us, an answer to (9) is

A = When firms, trapped by the signal extraction problem can't anticipate inflation, they produce more or less than it is desired by households, generating a disequilibrium between general price level and output.

So, following the quaternary view, contrasts can be added to the given structure, assuming that the cause and the foil are both conditioned to the same background condition (*B*):

B = Fixed parameters

C = Firms do not anticipate inflation

E = Manifestation of the Phillips curve

[93] As an analogous economy, Lucas's Phillips curve presents some features of what is currently called an example of how-possibly explanation, how certain parts and features are somewhat articulated in order to (potentially) generate an explanandum phenomenon (Craver, 2006). These explanations can provide modal information about how a target event might possibly occur (Forber, 2010). They offer modal information about how certain events might be brought in, rather than describing how a given event was actually the case (how-actually explanations). Even though, by their nature, these explanations do not describe actual events of the world, they can be seen as legitimate types of explanation capable of generating understanding (Verreault-Julien, 2019). For more details about how-possibly explanations in economics, see Till Grune-Yanoff and Phillipe Verreault-Julien (2021).

B = Fixed parameters

C_0 = Firms anticipate inflation

E_0 = There's no Phillips curve

Like the other examples earlier discussed, pairs $\{C, E\}$ – $\{C_0, E_0\}$ are counterfactually channelled. Both explanans are conditioned to parameters (B): their explanatory power is subjected to the stability of parameters. By fixing the parameters, modellers can simulate the effects of alternative policies and consequently provide more reliable forecasts. Consider, for instance, someone from Federal Reserve phoning Lucas in order to ask for some advice:

(10) Please, Bob, what may possibly happen if we raise the money supply?

Maybe Lucas could say something like this: "Well, I envision two possible scenarios. In the first one, you should consider a situation where firms correctly anticipate the elevation in the money supply. In this case, the money supply will not affect the real GDP and in turn does not affect unemployment. Or people somehow couldn't anticipate the new policy. Then, in this case, an inflationary policy can have an impact on real output. But beware: this is very short-lived because people are not ludicrous; they rapidly adjust their behaviour according to new pattern of policy".

Lucas has in mind this output-inflation trade-off equation below:

$$Y_t = Y_{p,t} + \theta\beta\,(P_{RGT} - \overline{p_t}) + \mu \qquad (1)$$

where (Y_t) is the output (aggregate supply), ($Y_{p,t}$) is the normal level of aggregate production, ($\theta\beta$) are the dynamic microeconomic parameters, (μ) represents the invariant "deep" parameters of tastes and technology, (P_{RGT}) is the actual rate of inflation, and $\overline{(p_t)}$ is the average of people's expectations about the common component of all market-specific prices in the economy (i.e., inflation expectations). Here is how we can represent this story:

B = Fixed parameters

C = Firms anticipate inflation

E = No impact on real GDP

B = Fixed parameters

C_0 = Firms do not anticipate inflation

E_0 = Momentaneous effect on real GDP

As this case reveals, the contrast account can equally serve to accommodate forecasting questions. As long as we understand in which conditions a given phenomenon may happen, this type of knowledge provides the cognitive path to predictions, which then serve to test and refine the explanations (Douglas, 2009, p. 454).

In this section, I argue that models *a la* Lucas teach us how some key economic variables can possibly generate a macroeconomic phenomenon. Economists can surgically manipulate the variables of their interest and raise various *what-if* questions. Even though the information cited in the explanans does not amount to the actual psychological/neural states of individuals, the model counts as explanatory because it enables us to answer what may possibly happen when the individuals deviate from the theoretical axioms predicted by the theory (e.g., the very assumptions of the rational choice theory – RCT). Although the information listed in the explanans is not actually confirmed, i.e., it does not mention some actual factors that produced the outcome (how-actually explanations), it delivers instead valuable knowledge about an accurate list of potential causes (how-possibly explanations)[94] that may count as legitimate explanations:

> These thin RCT applications provide informative how-possibly explanations of real-world agents' choices because they enable RCTs to accurately answer many what-if-things-had-been-different questions about these choices by pointing to interrelated variations in thin RCT axiomatic requirements and thin RCT implications. (Fumagalli, 2016, p. 68)

The model describes how a given set of agents, namely the suppliers, normally behave in the realistic setting of imperfect information. To the extent that the information listed in the explanans is not discarded by the known facts, we can still count that explanans as valuable (how-possibly) explanation (Salmon, 1989, p. 137). A how-possibly explanation, unlike a how-actually one, does not need to match the external conditions of actual explanations in order to count as explanatory (Verreault-Julien, 2019, p. 28) – the counterfactual responses we can provide with such models license us to consider such models as explanatorily relevant inasmuch they provide relevant counterfactual knowledge and understanding.

[94] Philosophers of economics who express scepticism about the explanatory capacity of economic models have in mind that external validity is a necessary condition to qualify economic models as genuine explanations. Some of these philosophers include, for instance, Reiss (2012), Alexandrova and Northcott (2013).

While REH bears internal validity, AEH does not – this response can count as a legitimate possible answer as long as it does not run in contrary to the known facts. As such, the counterfactual knowledge derived from the model permits that we test its predictive power, which licenses economists to consider the Lucasian model as an approximate (although imperfect) representation of its target, a representation that is highly plausible, conceivable (Mäki, 2009), credible (Sugden, 2000) and (why not) probable. More work needs to be done if the modellers' ambition is to move from how-possibly down to how-actually explanations, providing extra external validity, and validating the how-possibly explanation of the Lucasian model in probabilistic terms, but these extra steps are outside the scope of this dissertation.[95]

5. Epstein's hurricane case: the base explains the top

Taking the quaternary contrastive account as the paradigm, let's have a look at how Epstein's challenge against global supervenience fits it. But let's first refresh our memories. In his gedankenexperiment, Epstein argues that all macroeconomic facts are triggered by an external event (e.g., a hurricane) without collecting information from people's epistemic conditions. Accordingly, the budget scheme is annually accrued, step by step, where each earlier step entails the following macroeconomic result. The direction of causality, Epstein believes, flows from the environmental event to the macro set with no relevant mediation coming out from the micro set.

All the necessary and sufficient information to explain the macro set is contained outside the microeconomic set: the inner logic of the balance sheet plus the exogenous cause (weather outbreaks). So, if microeconomic facts are irrelevant to fixing the macroeconomic facts, then the micro set fails to explain the macro. Therefore, macroeconomics is autonomous from microeconomics in this case. Following Epstein's steps, we can visualise a why-question for the rise of the macro set in the following terms:

(11) Why, at year 6, (in case of *B*) did government disbursements sum to $ 48.7 million?

[95] I agree with Alisa Bokulich (2014, p. 335) who interprets the difference between how-possibly vs. how-actually explanations as matter of degree rather than two rival explanatory categories – there is a continuum from actuality to possibility whose crucial difference stands on the degrees of confirmation. Fumagalli (2020, p. 70) lists some contributions that reveal that the axioms of RCT have received some empirical support and the applications of thin RCT's have also relinquished some of its main assumptions in order to make the requirements compatible with behavioural findings.

This why-question entails that it is true the government did accrue $ 48.7 million in disbursements. So, the event E = *Government disbursements summed 48.7 million at year 6* is the explanandum, which we aim to explain. Epstein argues this real-world event E was triggered by an actual environmental event C (a hurricane). Here is how we can portray the contrasts:

C = There was a hurricane.
E = Government accrued $48.7 in total disbursements at year 6.
C_0 = There was a storm.
E_0 = Government has had no further obligations.

Both pairs, $\{C, E\}$ and $\{C_0, E_0\}$ are counterfactually connected. If there was no hurricane, the government would not have accumulated that amount in disbursements: E is counterfactually dependent on C. If there had been a summer storm, the government would have had no additional obligations – in symbols: $\sim C \,\Box\!\!\rightarrow \sim E$.

A direct answer to (11) is given by the proposition F:

F = The presence of a hurricane, rather than a storm, has caused the government $ 48.7 million in total disbursements at year 6, rather than another value.

Although we may agree with Epstein that the presence of a hurricane is a necessary condition for the rise of governmental disbursements, F says nothing about in which manners a hurricane can affect macroeconomic results directly. We should be more cautioned in accepting F as the necessary *and* sufficient cause. I shall rephrase (11), specifying the appropriate contrasts:

(11a) Why did the government accrue 48.7 million in disbursements, rather than nothing?
(11b) Why did the government, rather than a private agent, accumulate $ 48.7 million in disbursements?
(11c) Why did government disbursements sum to $ 48.7 million, rather than any other value (e.g., $ 60 million)?

Following the quaternary view, we can picture a direct answer to (11a) in the following way:

C = presence of the special legal enforcement rule
E = government disbursements

C_0 = Absence of the special legal enforcement rule

E_0 = Absence of specific government disbursements

Once the appropriate contrasts have been specified, a proper answer to (9a) is

G = Because there is a rule, rather than no rule at all, dictating the specific amount of public expenditures in case of hurricanes, rather than in case of storms.

The counterfactual test guarantees the truth value of G: both pairs representing the actual {C, E} and the non-actual events {C_0, E_0} are counterfactually channelled. Relevant explanatory relation R to the actual events {C, E} is granted by G. For the second question (11b), we have the scheme below:

C = existence of a legal rule strictly designed for the government

E = government disbursements

C_0 = legal rules for the business sector

E_0 = absence of public disbursements

A satisfactory answer to (11b) is

H = Because only the government, rather than the private sector, is legally obligated, rather than having no obligation to expend an exceptional amount of resources to rebuild houses destroyed by hurricanes.

As the answer to the query (11b), this one also presents the counterfactual connection between the explanans (C) and the explanandum (E); the latter is counterfactually dependent in the former. The private sector is more than welcome to cooperate in this process, but it is not legally obliged. Then, for the final question (11c), we have the following two slots for the actual events and another two for the non-actual events:

C =in an annual (fixed) percentage interest rate of 4%, $46.8 million in disbursements were accrued until year 5

E = $48.7 million in total disbursements in year 6

C_0 = in an annual (fixed) percentage of interest rate of 8%, $54.9 million in disbursements are accrued until year 5

E_0 = $64 million in total disbursements are accrued in year 6

Like all the previous brackets, this one also shows how both pairs, the actual and the non-actual events, are counterfactually linked.[96] A direct answer to (11c) is

I = Because the governmental disbursements are annually accrued with a fixed annual percentage of interest rate of 4%, rather than 8%, government accrued $48.7 million in total disbursements in year 6, rather than $64 million in total disbursements.

The explanandum *E* is the ultimate result of a budget exercise initiated much earlier (5 years before). The explanans *C* is the immediate cause of *E*; if *C* were different, say C_0, *E* would also be different (E_0). But notice that *C* is not only the direct cause of *E*, but it is also an effect for the previous amount of governmental disbursements one year ago (at year 4). A causal chain links the first macro fact (year 1) to the final result. In this regard, we expect the entire causal chain in the budget scheme would have been different under different values in year 1 as long as the sequence of macro results (years 2, 3, 4, 5, and 6) are accrued over time. Therefore, we may say *I* has a relevant explanatory relation for this whole sequence of events. Also, this question depicts how the quaternary contrastive account responds well to potential paradoxes involving transitivity:

> Contrastivity resolves the paradox of transitivity. […] [It] treat[s] causation as differentially transitive: if *c* rather than *C** causes *d* rather than *D**, and *d* rather than *D** causes *e* rather than *E**, then *c* rather than *C** causes *E* rather than *E**. (Schaffer, 2005, p. 340)

So far, none of the abovementioned answers to questions (11a) – (11c) mention external events directly, showing that, as argued in the previous chapter, the weather event cannot be viewed as the direct cause for the rise of the macro set. The chain of causality from weather events (the first cause) to the macroeconomic results (the effect) necessitates the micro set as a mediator. This is patently clear in answer to the question (11a):

G = Because there is a rule, rather than no rule at all, dictating the specific amount of public expenditures in case of hurricanes, rather than under the presence of other weather events.

In this response, the piece of information about hurricanes is *secondary*. The economic effects of hurricanes are subordinated to the existence of a legal standard that dictates how much the public sector should expend on rebuilding houses that were damaged by hurricanes. Hurricanes do not depend on anything to produce an immediate impact on households. On the

[96] The values of the foils {C_0, E_0} were corrected according to the new value of annual interest rate (8%). I preserved the original values and the adjustment mechanisms of the accountability exercise provided by Epstein. For more detail, see once again chapter 4.

other hand, in Epstein's argument, the external event can only produce its effects on the macro set by a legal standard that predicts what should be done in case of hurricanes. In other words, the causal powers of hurricanes are ineffective to macroeconomics unless there is a rule predicting what must be done when hurricanes have broken.

As argued in the former chapter, the existence of the legal standard (the institutional apparatus that belongs to the micro level) is not an innocent presupposition with no relevance for the rise of the macro set. When we make explicit what is only implicit in Epstein's argumentation, we shed light on the dependence between the legal rule and the macro set in the form of *C*, rather than *C** in the micro set, explains *E* rather than *E** in the macro set, as demonstrated in the answers above.

To represent Epstein's ambitions via the contrastive account (i.e., to forge a direct causal relation between the external event and the macro set), consider the next question below:

(11d) *Given* there is a rule dictating the specific amount of public expenditures in case of hurricanes, why did government accrue 48.7 million in total disbursements?

In this question, Epstein's presupposition is now explicit. It is represented by a "given clause" *P* = "*Given* there is a rule dictating the specific amount of public expenditures in case of hurricanes" while *E* = "Government accrued 48.7 million in total disbursements" is the explanandum. Herein, every possible contrast answer to *E* must consider only those alternatives that satisfy the presupposition *P* (Garfinkel, 1981, p. 29). When we forge the question in the form of (11d), we can finally capture Epstein's intention. Now, the explanans is the weather event, performing a relevant answer *J* which, in this case, is

J = *E* was *caused by a hurricane 6 years ago*, rather than any other potential environmental events.

So, contrary to what is argued by Epstein, the micro set explains the macro by dictating the conditions in which the external event can also be counted as explanatory. While the outer event manifests its causal powers in terms of the triggering cause, the base behaves as a constraining operator, limiting and imposing the range of possible values of the macro set, rather than triggering it. The presence of the rule is vital for the occurrence of the explanandum because it is through the rule that the hurricane could produce its economic effects. The rule constitutes a necessary condition for the macro set, but it is a not *sufficient one* for its rising: without the presence of a hurricane, the macro would not have existed. The rule itself is not the difference-making of the macro – this role is deputised to the external event. The base

constraints the triggering cause which, as I argue in the section below, constitutes a causal element with supervenient features.

Supervenience is a modal relation that it is not, per se, explanatory. It is a binary relation between two supervenient sets that requires merely that the macro cannot be different without the micro being different. Knowing that a macro state of interest supervenes on a specific base does not ensure that that base helps us understand why that particular macro state occurs. But, as this case shows, instead of mere determination, the application of supervenience in economics can often provide explanation. Moreover, it illustrates that an explanation of why given supervenient (i.e., macro) state holds does need not to be represented in a binary manner. More importantly, this example shows the contrast account explains *why* micro explains the macro but does not inform in which way(s).

On my view, the base functions as constraining operator – creating, shaping and guiding the trigger event (the hurricane) responsible for determining the macro. If the rule was absent or had been different, the macro would not have been existed or it would get another result. The base is a condition, rather than an event, the usual relata of causal explanations. This unique bottom-top explanation played by the micro set should be qualified, in my view, as an example of a constraint-based explanation.[97]

5.1. Evaluating the answers: how the micro constrains the macro

The idea of explanation in terms of constraints has gradually gained attention in the literature. Debates about the nature of constraint-based explanations appear in contemporary philosophy of science (Lange, 2018), social sciences (Haslanger, 2015), and metaphysics (Bertrand, 2019). Recently, several scholars have defended the idea that many scientific explanations across various disciplines do not describe causal processes. Imagine, for instance, we want to explain why a mother cannot evenly distribute a box containing 23 strawberries among her three children. The explanandum for this event is "A mother at time *t* could not evenly distribute 23 strawberries among her three children", while the explanans is "Because 23 cannot be evenly divided by three". The mother did not proportionately distribute the fruits among the children because she *could not* (Lange, 2018). Contrary to causal relationships that are arguably (i) contingent and where (ii) the cause and the effect must be distinct (Craver, 2007,

[97] Ross (in Review) argues in favour of a special sort of causal constrain-based explanations which are pretty much dissimilar from the standard accounts of causation.

p. 179), the modal force responsible for channelling that explanans to the explanandum in this case exhibits relational patterns that do not allow us to qualify that explanation in causal terms.

Like some mathematical truths, Lange argues some laws of nature (e.g., conservation laws) are modally stronger than ordinary laws: the former would be the "transcending" forces of nature. According to Lange, conservation laws actually *limit* the kinds of interactions there could have been, making a non-conservative interaction impossible. These limits are the actual *constraints*, conditions that manifest their existence by imposing boundaries and restrictions for the occurrence of some phenomena rather than *triggering* them. So, constraints are conditions that "both limit and afford a certain scope of possible structures and functions that can be instantiated in a system of a particular type" (Green and Jones, 2016, p. 345-346). Consider, for example, that *Y* is a given phenomenon with its specific causal powers. If *X* somewhat constrains *Y*, then *X* must reduce the degree of freedom of *Y* in multiple ways. Think, for instance, in how the social structure may effectively restrain individuals in the social world.

Sally Haslanger (2015) argues that, according to our explanatory target, social structure (i.e., information about institutions, culture, and other pieces of social world that do not apply to the bodily movements of individuals) operates as a constraining factor of individuals in society. According to the case under assessment, the structure of the social world can perfectly occupy the explanans slot in sociological explanations. Consider this common situation:

> Bus schedule: Jason has a job at a factory in the suburbs. His shift is at 6 am. He is poor and relies on the bus to get to work. He takes the first from his neighbourhood in the morning and after a 45 min commute arrives at his job on time. Due to cut-backs, however, the city has decided to reduce the bus service and there is no bus leaving the city in the morning that will get him to work on time. He asks for a shift change, but it is not eligible. He loses his job. (Haslanger, 2015, p. 7)

Our goal is to explain why Jason has lost his job – that is our explanandum. According to Haslanger, to ascribe to Jason's boss the reason why he's been fired is unfair insofar as the regulations about shift changes constrain the boss himself – this is what the legal standard recommends in cases like this. In the same line, to blame the county may be equally unfair, since the tax base may constrain the county itself.

In conclusion, Haslanger argues that the social structure operates as an objective constraint: no single social agent is directly responsible for triggering the explanandum. On the contrary, the social structure is the explanans – the chain of constraints across the social agents and institutions (legislation of shift changes, city tax base, buses) *moulded* the conditions for the rise of the event. No social (individual) actor related to the event (directly or indirectly) had a choice to behave differently. Nobody has purposely triggered the explanandum. While the run-of-mill causes are often seen as the triggering causes, we also have another class of explanations

that are more structural: conditions that structure, guide, or shape the explanandum outcome rather than firing it. For example, imagine this situation suggested by Dretske:

> A terrorist plants a bomb in the general's car. The bomb sits there for days until the general gets in his car and turns the key to start the engine. The bomb is detonated (triggered by turning the key in the ignition) and the general is killed. Who killed him? The terrorist, of course. How? By planting a bomb in his car. Although the general's own action (turning on the engine) was the triggering cause, the terrorist's action is the structuring cause, and it will be his (the terrorist's) action, something he did a week ago, that will certainly be singled out, in both legal and moral inquiries, as the cause of the explosion that resulted in death. (Dretske, 1991, p. 122-123)

In Dretske's example, the triggering cause manifests its effects *via* the bomb. When the general started the engine, he detonated the bomb: the car's ignition is the trigger while the bomb is the structuring cause. The triggering cause controls the result; it explains why the general was killed on a Monday morning rather than any other day. The armed bomb, in turn, explains how the assassination was possible – an architecture of circuits, wires, and explosives arranged somewhat to produce the expected outcome when someone pulls the trigger (in this case, when the general started the engine). Taken in isolation, triggering and structuring causes cannot produce the desired outcome. Both are necessary and sufficient conditions for the occurrence of the event, although they manifest their contributions in different ways: the former fire the result, and the latter shape/guide/create the conditions of possibility for the triggering cause to work its effects.

Dretske's structuring cause functions as a constraining operator: a structural condition that creates the conditions of possibility of a given target without triggering it. If the terrorist had unarmed the bomb, then the bomb would not have exploded. The terrorist could surgically impede the event by unarming the bomb. Although the bomb itself cannot be seen as the triggering cause, its activation turns the outcome possible. On my view, Epstein's scenarios follow an analogous pattern.

In the previous section, I demonstrated that all the answers for the rise of the macro set offered by explanations *G*, *H*, and *I* illustrate different modes of constraints by the micro upon the macro. In Epstein's argument, "The government accrued 48.7 million in total disbursements" is our explanatory target (explanandum), while the answers *G*, *H*, and *I* describe different constraining aspects exerted by the rule on the macro that explains why the government had to pay this amount of money.

For instance, the explanation *G*

G = Because there is a rule, rather than no rule at all dictating the amount of public expenditures in case of hurricanes, rather than under presence of other weather events

informs which condition in the world – a legal standard – is responsible to dictate the amount of public resources disposable to rebuild houses affected by hurricanes. That macro set was only possible because, as argued exhaustively across the last two chapters, the micro set mediates the external event: the rule is directly responsible for fixing the limits, the range of possible values that the macro might have possibly assumed, while the hurricane represents the indirect cause. An alternative value in total disbursements would be possible *if and only if* the rule would dictate something different. Once the rule determines how much should be spent in reconstructing the houses that have been hit by the hurricane, the total disbursements *necessarily* follow from what is been dictated by the rule, as the answer *G* states very clearly.

In its turn, answer *H*

H = Because only the government, rather than the private sector, is legally obligated, rather than having no obligation, to expend an exceptional amount of resources to rebuild houses destroyed by hurricanes

tells us which social agent (the government) is responsible for reconstructing the houses. The legal standard dictates the actions of the government: if the government disrespects the law, it should be fined for that. The rule therefore constrains a specific social agent (the government) who is legally compelled to follow the law.

Finally, by saying that

I = Because the governmental disbursements are annually accrued with a fixed annual percentage of interest rate of 4%, rather than 8%, government accrued $48.7 million in total disbursements in year 6, rather than $64 million in total disbursements

answer *I* mentions the procedure (a balance sheet) that is often applicable for this sort of case – a debt which is (annually) accrued over time designed to preserve the correct amount of the final payment according to the rule. It informs why the government accrued that final amount in disbursements rather than any other value. All the further corrections made by the authorities to restore the original value caused by the time-lag were dictated by the rule.

The rule, a structural feature of the micro set, (i) creates the conditions for the possibility of the macro set, (ii) imposes its limits, and (iii) fixes its procedures for corrections of the macro values. The triggering cause, the hurricane event, could only have triggered the macro set under the presence of this micro rule which dictates how much the public sector should spend in face of weather calamities.

The triggering cause exerts its effects on the macro through the micro set: it affects individuals and once the authorities identify it was a hurricane rather than a storm, it generates its effects on the macro instantaneously. It does not matter if the authorities have learned it was a hurricane only after the event and must latterly correct the budget scheme. The corrections are all predicted and dictated by the rule: the base constraints all the possible states of affairs of the macro.

In conclusion, supervenience *can* explain not simply by identifying and describing supervenient patterns in economics. In Epstein's case in particular, supervenience exhibits its explanatory power by detailing different modes of constraints, how the structural items from the micro set can actively limit, delineate, and fix the range of possibilities of the macro set. Alongside other examples of constraining-based explanations, supervenience is another one of these that interestingly can be accommodated in the contrastive quaternary account of explanation. The option for a quaternary account, as argued before, is more than justified. It is patently clear that the binary form of explanation omits the crucial factor played by the micro set for the rise of the macro set.

6. Concluding remarks

In this chapter, I explored a slightly modified notion of explanation, namely, the quaternary account of contrastive explanation. This syntax, consistent with Woodward's interventional counterfactual view explored in chapters 2 and 3, is very flexible. The discussions across this chapter reveal that contrastive syntax can be employed to forge explanations far beyond its original scope, designed to cover causal explanations mostly. It can explain the actual causes of historical events, its internal processes and may serve to compare and subject alternative explanatory hypotheses to the available data (Weber's case). It also contrasts contending hypotheses designed to explain the same phenomenon and discards explanations inconsistent with economic theory, serving as a valuable heuristic tool to simulate alternative forecasting scenarios (Lucas's case). And the contrastive account can be used even to track noncausal relations of dependence like supervenience which, metaphysically, is represented in binary terms (Epstein's case).

The contrastive explanatory account is neutral about whether the explanations here in place are good or correct – these epistemic virtues are subject to other epistemological criteria in dispute among the respective epistemic community of historians (in Weber's case) and economists (Lucas's case). Nevertheless, it illustrates objectively why some patterns of relations

hold, explaining why they had been possible rather than some other alternative result. As we see it, the neutrality of contrastive syntax permits us to cover a wide variety of explanations that may allow, unlike what Verrault-Julien argues (2019, p. 26), the same syntax to evaluate how-possibly and how-actually explanations in economics.

In the recent debates around the explanatory power of economic models in philosophy of economics, the application of contrastive view is underexplored, even among scholars who claim that how-possibly explanations are genuine kinds of explanations capable of generate (modal) understanding. Recent literature on the epistemic virtue of economic modelling how-possibly explanations do not pay attention to the proper syntax of how-possibly explanations (if these can be qualified as legitimate explanations). In this chapter, I defend the contrastive view as a general framework in which the defenders of the explanatory power of economic modelling can exploited to articulate their imaginary economic worlds and level various what-if scenarios according to the contrastive account.

CONCLUSION

I begin this final section with a brief account of what contributions I have made in this dissertation and then provide a more discussive discussion of them.

Contributions in brief

In this thesis, I provided reasons that indicate a peaceful co-existence between reductive and non-reductive strategies in the intersection of micro-and-macroeconomics. Following Ernest Nagel's footprints, I treat reduction as a special type of inter-level explanation between the regularities that correspond to a specific (lower-level) theoretical domain and the regularities from the target (higher-level) domain. As an explanatory strategy, reduction serves a variety of scientific purposes that may assume multiple forms. In economics, the purpose of reduction is to accommodate intentionality: to explain the manifestation of every macroeconomic phenomenon in terms of optimised and constrained decisions of individuals.

I qualified this purpose of economists as explanatory parsimony: to explain the rise of macroeconomic events and regularities in terms of individuals' properties (their beliefs, intentions, and expectations). In positive examples of reductions, the explanans consists of individuals' properties as the causal realizers of the macroeconomic outcome (explanandum) – all the extra contextual elements (macro data and market-specific prices) that figure in the explanans are causally inert for the rise of the macro phenomenon; they are the object of inquiry of individuals. To achieve that goal, the relation between explanans and explanandum does not necessarily demand identities, nor identities between the elements that make possible the derivation of the explanandum (the Nagelian bridge principles). As I argued in chapter 3, we may have reductive explanations even if those allegedly ontological issues aren't previously resolved. The requirement to first solve all the ontological issues before employing reductive methods is a matter of preference.

The application of different reductive recipes to relate micro-to-macroeconomics (Nagelian and constitutive-mechanistic approaches) acknowledges the reality of other entities in economics apart from individuals. What is denied is their efficacious causal roles in the context of economical explanations: while individuals are the active causal cogs, these extra non-individual features are the constraining elements. The variables representing the individuals' properties and variables that amount to extra non-individual elements have dissimilar roles in the explanans. As I claimed in chapter 3, the rise of the Phillips curve is only possible due to individuals' features, the actual realizers of the Phillips relation, and their

contextual surrounding relevant sources of economic information (macro data and market-specific prices) as their constraining features. Clearly, the elements figuring in the explanans for the rise of the Phillips curve are not on par: the individuals' properties have a dominant role in the manifestation of the macro phenomenon – it is the dominant role that individuals' properties may have in the face of their constraints that guarantees reduction.

Alternatively, a non-reductive approach has also room to relate micro-to-macroeconomics properties. By targeting some macroeconomic phenomenon (explanandum) and assuming that its explanatory variable (explanans) is necessarily placed on the microeconomic level, in cases where the non-individual entities in the micro-level (market-specific prices, or other potential elements such as institutional features) hold an active profile in determining the rise of the macroeconomic phenomenon, then we have a non-reductive explanation for the manifestation of the macroeconomic target at stake.

In such cases, individuals' properties and their constraints are on par: none of them have a dominant role. In inter-level explanations that integrate micro and macroeconomics, global supervenience is applicable to cases where the selected elements that figure in the explanans are both decisive for the rise of the macro phenomenon, even though they have distinctive roles. According to what the explanandum located at the macro level is, micro and non-micro properties should be allowed to be taken as equally relevant for determining the macroeconomic outcome – both may equally figure in the explanans of macro features in economic explanations.

Epstein's cases in the chapter 4 illustrates this issue in a transparent way. Therein, the rise of the macroeconomic liabilities were only possible due to the institutional features. The counterfactual analysis makes visible the active role that the legal rule possesses for the activation of the macro events: no rule, no liabilities.

What I think of as a central feature of my contribution is that the major role of reduction in economics is to provide explanations: in paradigmatic cases that I looked at, it's supposed that the macro can't be adequately understood or predicted without a clear understanding of the properties of individuals, their properties and behaviours (the microfoundations requirement). This kind of explanation doesn't require an ontological reduction – the microfoundations thesis is an explanatory thesis that may assume different forms which may be consistent with the purposes of reduction.

A second particular contribution I made on this front is to illustrate that micro reductions of this kind often fit contrastive explanation, which matters for predictions and the assessment of failures, as I showed for Lucas's model in the case of the Phillips curve.

A third contribution I made is an elucidation of Galbács's claim. I enriched his argument, showing how the institutional features inside the micro level are necessary for the very manifestation of the macroeconomic properties. I advanced Galbács's argument, arguing that the way the institutional features manifest their causal relevance is by imposing structural constraints, shaping the conditions in which the outer event in Epstein's thought-experiment (hurricane) may manifest its causal powers. I argued here that it makes no sense to picture agents acting and interacting in no setting at all. Structural assumptions are necessary in the micro base to describe the setting and constraints in which they act.

Fourth, there's another contribution: showing that two potential reductive frameworks – Nagel's reduction and constitutive-mechanistic account – are both compatible to Lucas's model.

Fifth, I believe I have enriched the available general understanding of reduction and some related concepts by providing new explanations and illuminating their applications:

- According to the view I have developed, reduction is an explanatory strategy. It serves various goals in accordance with the exploratory interests within a determined scientific community. In economics, every inter-level explanation between micro and macroeconomics where the properties of individuals prevail over their constraints can be qualified as reduction.
- I have argued that supervenience is a suitable concept to bridge micro-and-macroeconomic properties according to the context of interest. According to the case study, we may have supervenience or reduction – it is a matter of determining whether the properties of individuals and their constraints have equivalent powers to determine the outcome (even though they hold dissimilar roles).
- I contend it's not every positive application of supervenience that provides explanation. Whether we have reduction or supervenience is judged by which features are figuring in the explanans.
- Contrary to what many authors have assumed, there's no all-or-nothing dispute between these two approaches, reduction v. non-reduction. The dispute between reductive v. non-reductive approaches should be assessed by the explanatory capacity of each of the frameworks according to the context.

Further Discussion

In this dissertation, I analysed the merit of two reductive (Nagelian and constitutive-mechanistic reduction) and a non-reductive (supervenience) approaches to connecting micro-to-macroeconomic properties/events or assumptions (representative items contained in economic models). Plainly put, the overall message of this dissertation is that whichever virtues or failures are attributed to such frameworks should be evaluated according to their explanatory capacities and which goals the inquirer aims to achieve.

Scholars who reject reduction in economics (such as Kevin Hoover and Harold Kincaid) have in mind an ontological view of reductions. In this interpretation, the goal of reduction seems to be *ontological parsimony*: reduction must decrease the ontology of the universe. That's why Kenneth Schaffner believes that reduction functions should express identity statements – where there appear to be two entities, there's only one. In such an account, the impossibility of reducing some crucial macroeconomic entities to microeconomic components makes reduction a difficult enterprise.

Other concrete issues reinforce the challenge to serious reduction in economics: economical decision-making is not dissociated from the macro – e.g., people use money issued by the government for their transactions and use macroeconomic data as their informational data to enhance their decisions. The overall message seems to be that as long as macroeconomics may influence the microeconomic level, then there's no reasonable pathway to subscribe to an ontological reductive agenda in economics.

These general reservations against reduction in economics concentrate on ontological barriers. In Hoover's view, the fact that crucial macro entities can't be decomposed or eliminated imposes realistic obstacles for reductionists as a matter of fact. Post-Keynesians, in turn, also believe that the large influence of macroeconomics on microeconomics undermines an attempt to understand microeconomics as more fundamental than macroeconomics. According to them, the direction of causality between the two layers indicates that the macro often precedes the micro.

By acknowledging these ontological barriers, neoclassical scholars should discard the microfoundational program and happily embrace the idea that other social-economic entities are as much causal and efficacious as individuals. In so doing, they would acknowledge that other features besides the microeconomic level may equally figure in the explanans of their economic explanations, accepting other methods as equally valid for understanding economic phenomena. The adequate epistemology and methodology are better scientifically oriented if

they are grounded in the correct ontology, which suggests, in economics, an anti-reductionist approach – as Hoover contends.

As I have argued (chapter 3, Section 5), this is a pertinent prescription but not a strict requirement for economists when they forge their explanations and select their methods. An economist may adopt the most suitable methodology according to her target of interest, disregarding other elements that she considers uninteresting for the analysis. For instance, the economic modelling offered by Lucas that I discussed (chapter 3, Section 2) abstracts several elements away from analysis, such as the demand conditions, other possible exogenous shocks in the economy that aren't monetary only, the role of culture for economic behaviour and so on. These are factors that Lucas viewed as non-significant for what he was seeking to understand and explain.

Like other types of explanation, reduction, a special type of explanation, is subordinated to the goals and explanatory ambitions of the inquirer (Steel, 2004), what she aims to achieve and understand with reductive research. In economics, the goal of reduction is often to recover intentionality, preserving the popular definition of economics coined by Lionel Robbins (1935, p. 16): "economics is the science which studies human behaviour as a relationship between ends and scarce means which have alternative uses". In providing microfoundational accounts, economists seek *explanatory parsimony* in attempting to restrain the explanans of economic explanations to the properties of individuals.

Across the dissertation, I treat reduction as a sort of inter-level explanation. Among the various theories of reduction, I evaluate the merits of two reductive approaches applied to a case study in economics: Nagel's reduction and constitutive-mechanism reduction. In both paradigms, reduction is attained when the inquirer forges an explanation of a phenomenon that manifests its features in a higher-level of organisation, based on the principles, assumptions, laws, or properties grounded in a lower-level of organisation. In other words, the reductionist wants to transfer the explanans for the manifestation of the higher-level event to the next level down.

Although necessary, this condition is not sufficient. In the context of economics, there's a reduction of a macroeconomic phenomenon when the explanandum is exhaustively explained by the properties and behaviours of individuals – any other social entity within the microeconomic level should be causally inert. When the surrounding elements inside the microeconomic level actively contribute to the manifestation of the given macroeconomic event, then there's no chance of reduction – as simple as it is.

The first two chapters were dedicated to introducing the two explanatory potentials of reductive frameworks: Nagel's account and the constitutive-mechanistic approach. In the first chapter, I laid out the first contender, Nagel's account of reduction: a type of inter-level explanation that is attained when the laws, or a set of laws, of a higher-level or reduced science are inferred from a set of general principles or experimental laws of a primary or reducing science with the auxiliary of reduction functions when needed. We have Nagelian reductions when these set of conditions are positively met – when they are not, there's no reduction according to Nagel's terms. Reduction, in this account, is the inter-theoretical subsumption of laws with the aid of auxiliary assumptions to bridge the two layers.

In the second chapter, I set forth the second candidate to enact reduction in economics: constitutive mechanisms. Therein, I argued that, very often, a constitutive mechanism is a stage of inquiry when the mechanist tries to understand how the parts of the system (the explanans) interact and how they possibly contribute to the entire manifestation of the system as a whole (explanandum).

Constitutive mechanisms are epistemically illuminating when the mechanist forges constitutive explanations, how the explanandum, placed at a higher-level of organisation, is determined by some of its relevant components. In that chapter, I employed Woodward's counterfactual account of explanation to cover causal and noncausal explanations to illustrate how these two types of explanations are integrated into a mechanistic explanation. In that chapter, I defended a monist approach to explanation, arguing that the explanatory dependence between the explanandum at the top level and the explanans at the lower-level is evaluated according to the counterfactual dependence between the two. If the explanans modifies its value in relevant ways, the explanandum also changes.

In chapter 3, I employed the selected reductive paradigms to assess a case study in economics: the Lucasian Phillips curve. In that chapter, I demonstrated the consistency of these two reductive approaches to Lucas's model. The representative-agent model designed to derive a macroeconomic outcome directly from microeconomic assumptions and principles is a positive example of an inter-level explanation where only the properties of individuals are *causally* efficacious for the rise of the Phillips curve.

In the model, there are other existing economic features inside the micro set: relative prices and macroeconomic information. These extra features in the social setting in which suppliers are embedded have a passive role: they are tags that constrain the set of options of suppliers – these economic entities within the microstructure are the objects of inquiry of suppliers. By their own causal capacity, these elements can't influence the Phillips curve directly

– it's people who make projections, hold expectations, and make decisions taking into account these extra elements as their informational sources. Considering the microeconomic elements that explain the Phillips curve, all extra elements are non-causally significant: only people's properties count as causally relevant for the manifestation of the Phillips curve.

The first reductive framework, Nagel's reduction, fits into Lucas's model in epistemic terms. The macroeconomic principle is inferred from theoretical microeconomic principles, which fits the first formal Nagelian condition: derivability. The second condition is met because the series of reduction functions or bridge principles conceived by Lucas to guarantee the derivation from micro principles to the macro principle can be viewed as *stipulations/conventions* – there's no ontological ambition to connect the very extensions of micro and macro quantities in the bridge principles. The first bridge principle claimed by Lucas, the notion of general price level that he adopted in the model, is a *definition* that is anchored in economic theory. Lucas abstracts away the nature of the relation between the general price level and its microeconomic components. This definition is an instrument at the service of what Lucas seeks to explain – how a macroeconomic principle is possibly determined by a set of microeconomic principles. As Klein (2009) suggests, the capacity of the reducing science in representing some target laws from the reduced domain is sufficient to get Nagel's reduction, and this is what we have in Lucas's model.

Conversely, the model can also be represented in mechanistic-constitutive terms, where the Phillips curve is a type of social macro principle that holds in a kind of structure or "mechanism" and that comes into being as a result of the structural items where suppliers, with their causal powers surrounded by a set of constraints, account for the correlation between the output and general price level. As argued in chapter 2, the relata of constitutive explanations are the causal powers, tendencies, or dispositions that the components of a given whole (the explanandum) hold in virtue of their structural elements located at the corresponding lower levels (the explanans).

The theory of explanation adopted to assess the functioning of this social machinery was the Woodwardian counterfactual view, which suits perfectly to this example. The Phillips curve manifests at the macroeconomic level due to the suppliers' expectations: if they correctly anticipate the general price level, there's no Phillips curve; if they don't, there's a Phillips curve. Thus, the macrophenomenon is counterfactually dependent on individuals' expectations.

The model, an example of an inter-level type of explanation, represents the capacities, tendencies, or dispositions that the chosen economic entities hold and the stable tendencies of these economic entities that may trigger some relevant macroeconomic outcomes such as the

Phillips curve. The reductive frameworks provide different forms of dependences between explanans and explanandum: Nagel's account demands subsumption, whereas the constitutive-mechanistic interpretation under the counterfactual account requires a counterfactual dependence between the macroeconomic result (explanandum) and the individuals' properties (explanans).

As I have presented it, we can see that by investigating economic capabilities, economic models like Lucas's Phillips curve model permit a counterfactual interpretation. But so too is it possible to turn Nagel's view suitable to some kind of counterfactual reading: "all economic models […] pose[s] a question of the form 'What would happen if such and such were the case?' in such a way that it can be answered deductively" (Gibbard and Varian, 1978, p. 668).[98]

In chapter 4, I flipped the coin and evaluated the merit of a non-reductive approach to bridge micro-to-macroeconomics. In this account, the extra economic elements or features (e.g., relative prices, macroeconomic data, institutional elements such as rules, and the like) that don't stand on individuals' properties (e.g., on intentions, expectations, beliefs or, as in the Lucas's model, on other features like the quantity each supplier provides) are as active as the properties of individuals are for determining a given macroeconomic outcome. Accordingly, the explanans of the macroeconomic phenomenon is composed of individuals and their features plus those other extra elements equally essential to determine the macro-outcome. In these non-reductive cases, individuals' properties and material (or non-individual) properties have equal powers in determining the macro.

With this idea in mind, I analysed the relevance of global supervenience as a possible candidate to fill the gap between micro and macroeconomics. I detailed Hoover's argument in favour of supervenience and assessed two potential challenges levelled by Julian Reiss (2004) and Brian Epstein (2014) against supervenience. Plainly expressed, in economics, the supervenience thesis claims that once a set of microeconomic properties has been fixed, the macroeconomic properties are also set instantaneously. Reiss raises doubt about it, arguing that downward causation (macro causing the micro) undermines the condition of supervenience that demands a bottom-top direction. Epstein, in turn, advocates the autonomy of macroeconomics: it's possible to have two worlds indiscernible in their microeconomic aspects with different macroeconomic properties.

In chapter 4, I defended the non-reductive supervenience approach: neither criticism, the downward causation nor the autonomy of macroeconomics, are fatal blows to supervenience. In the first case, I claimed that supervenience might appear behind the curtains:

[98] About a counterfactual reading of economic models, see Morgan (2002).

although the visible and epistemic interesting economic relation is the downward determination from the macroeconomic properties (the announcement of nominal interest rates by the monetary committee) towards the micro (individual's expectations/beliefs and other features they possess), there's a bottom-top supervenient relationship between that macroeconomic input (the announcement) and its corresponding lower-level components (the decisions of the members of the committee). This possibility is identified in a canonical example, Taylor's principle, a guiding tool used by central banks to keep inflation under control.

In the second case, I answered Epstein's worries. The supposed autonomy of macroeconomics from microeconomics is not all autonomy in any rigorous sense of the word "autonomy". His argument fails in demonstrating the autonomy of the macro. As I construed his thought experiment, the direction of causation flows from the macro-to-micro, and hence there's macro once again. His idea that some environmental disaster affects the macroeconomic properties directly stands only if we set aside the institutional properties from the microeconomic set. Part of this discussion is conceptual: how to define microeconomics and which elements it embraces. Epstein contends that the micro set only accepts those surrounding elements in the micro set that are causally related to individuals – in his view, this is clearly not the case of legal rules.

On the contrary, I argued that the rules are constraining features of the macro set: the hurricane could only trigger macroeconomic results due to a legal standard that fixes it as a necessary condition for the government to expend a specific amount in disbursements. On its own, a hurricane can't produce macroeconomic effects – it needs the rule that codifies that event in the world and turns it economically efficacious. This chapter concludes that we may possibly have different types of supervenient relations across micro-and-macroeconomic properties.

In the first example, the application of Taylor's rule, supervenience exists between two targeting elements placed at the two layers, although this relation is non-economically relevant: knowing that the collective decision of the monetary committee supervenes on the decisions of their members isn't epistemically important. In the second case, on the contrary, the existence of a rule constitutes a decisive feature in understanding the chain of causality in Epstein's cases. Without it, there's no macro set. The final chapter (chapter 5) made this issue visible.

In chapter 5, I concentrated on explanation, more specifically on a particular account of explanation, namely, the contrastive view. In that chapter, I revisited the relevant examples explored across the dissertation to argue that, sometimes, reduction offers a complete explanation for the target phenomenon while in others, it doesn't. Lucas's case elucidates this issue, showing that reduction may provide a complete explanation and the epistemic guide

towards better economic predictions. In supervenience cases, I argued that it's not every positive case of supervenience that is explanatorily illuminating, such as Taylor's principle, showing that supervenience may be seen – according to the given case – as a mere modal relationship between two sets of properties. On the other hand, in that chapter, it's shown how vital the acceptance of institutional properties as a member of the microeconomic set in Epstein's case is. In the contrastive form, the possibility of breaking the research problem into a more specific group of questions about the same topic elucidates the crucial condition offered by the legal standard to fix the macro set.

Study limitations and avenues for future research

Although I contended that both reductive frameworks fit the Lucasian model correctly, my argument was silent about the quality of such explanations. For some philosophers of economics (Reiss, 2012; Alexandrova and Northcott, 2013), economic models aren't explanatory on their own. In such interpretations, models *a la* Lucas don't provide explanations inasmuch as they don't illuminate any events of the real world. Moreover, such models, based on idealisations, misrepresent the very targets they intend to represent. In this regard, the Lucasian model couldn't be qualified as a genuine explanation. Alternatively, others (Verreault-Julien, 2019; Till Grune-Yanoff and Phillipe Verreault-Julien, 2021) retort that economic models provide *how-possibly explanations*, how some entities figuring in the explanans may potentially generate the explanandum, rather than revealing how a given event was actually the case. For these authors, these explanations are legitimate sorts of understanding capable of providing modal information about possible states of affairs.

Reduction, as advocated in this dissertation, is a special type of inter-level explanation capable of illuminating objective relations of dependence between the explanandum at the top hierarchical level and explanans on a lower line. The quality of such knowledge, especially when it's applied in the context of economic models, is open to discussion. It's clear that Lucas's model represents a how-possibly case rather than a how-actually example of an explanation. The fact that the reduction analysis is assessed in purely theoretical terms does not make the reduction invalid. As an explanatory strategy, reduction also admits degrees of explanatory power (Marchionni, 2018, p. 610). Lucas's model, the positive example of reduction investigated in this thesis, describes the possible causes for the rise of the Phillips relation based on real-world features. The model shares relevant similarities with its target, allowing us to infer an approximate description from the facts of the model and the facts of the target.

Secondly, it's interesting to investigate whether reductions can conciliate with other non-reductive approaches in economics, such as cultural economics, the branch of economic science that investigates the role of culture in economic outcomes. Recall that, as I said in chapter 2, the Weberian explanation for the rise of capitalism is grounded on the heavy influence of religion to influence people's minds. The mechanistic interpretation of Weber's view has a deterministic interpretation: there's a chain of causes from the initial cause (the dominance of Ascetic Protestantism) to the final event (the birth of capitalism).

In this regard, religion or other cultural elements sound like deterministic factors of behaviour. A more balanced position that conciliates the influence of culture on behaviour (the downward causation) with free-will may contest the downward cause, replacing it by constraining situational mechanism relating the environment to the individual. Once again, the mechanistic schema may be forged using Coleman's boat (chapter 2, Sub-section 5.2) where the cultural property/event constrains the minds of people, and their attitudes determine the outcome: the cultural event indirectly determines the economic outcome and people's behaviour determine it directly.

Finally, it's also important to evaluate the applicability of such reductive frameworks in other inter-field research programs, such as behavioural economics and neuroeconomics.[99] As an example, a behavioural approach to the Lucasian model would suggest a different strategy to explain the rise of the Phillips relation. Instead of taking for granted the theoretical assumptions adopted by Lucas – people use all available information economically (Rational expectations hypothesis), are minimally rational, and have stable preferences – a behavioural economist may suggest a de-idealisation of the explanans, making it more complex and informed by the evidence acquired from psychology and cognitive sciences (Heidl, 2016, p. 2).

Neuroeconomics, on the other hand, wants to go further. It claims that the economic models, in order to be compelling, should be well-informed by the knowledge produced in neuroscience. Craver and Alexandrova (2008) argue that neuroeconomics should be seen as a mechanistic science. In this view, neuroeconomics would follow the pattern of mechanistic explanation, which permits us to articulate how neural mechanisms contribute to individual decision-making. To the extent that the mechanistic account may allow multiple relations of dependencies across different layers subject to their counterfactual connection, there may be a

[99] Behavioural economics is the inter-field theoretical domain that applies psychological findings in economics. Summarising, behavioural economists try "to incorporate more realistic notions of human nature into economics" (Rabin, 2002, p. 657). Neuroeconomics, in turn, incorporates the evidence from neuroscience to develop economic models.

possible pathway to unify these multiple levels (macro, micro, and neuroeconomics) according to the target of interest.

Reduction, as I originally claimed, is an explanatory strategy. The microfoundations requirement of macroeconomics claims in favour of explanatory parsimony: there is a special economic entity (individuals) that prevails in the explanans of economical explanations. When the properties of individuals have a dominant role within the explanans, then there is reduction. Following the same pattern, we could conceive (or construe) economic models that respect this requirement but which, unlike Lucas, base the behaviour of individuals on different assumptions – informed by psychology. In this regard, reducing macroeconomics to behavioural economics would have a different goal: it aims to *correct* the idealised microeconomic theory. Similarly, neuroeconomists want to produce a well-informed economic science grounded in neuroscience – it seeks to correct the wrongdoings of fundamental neoclassical assumptions. There is room for unification based on mechanisms that may permit an integrated knowledge capable of generating more reliable macroeconomic predictions.[100]

[100] Fumagalli (2013, 2016) lists some empirical and conceptual difficulties faced by neuroeconomists in their journey for replacing the traditional utility theory in favour of experienced utility or neural utility.

REFERENCES

Adolph, C. (2013) *Bankers*, *bureaucrats*, *and central bank politics*: the myth of neutrality. Cambridge: Cambridge University Press.

Agassi, J. (1960) Methodological individualism. *The British Journal of Sociology*, 3, p. 244-270.

_____. (1975) Institutional individualism. *The British Journal of Sociology*, 26, p. 144-155.

Allen, R. (1967) *Macroeconomic theory*. London: Macmillan.

Alexander, S. (2016) *The jazz of physics*: the secret link between music and the structure of the universe. New York: Basic Books.

Alexandrova, A. and C, Craver. (2008) No revolution necessary: neural mechanisms for economics. *Economics and Philosophy*, 24, p. 381-406.

Alexandrova, A. and R, Northcott (2013) It's just a feeling: why economic models do not explain. *Journal of Economic Methodology*, 20, p. 262-267.

Angner, E. (2018) What preferences really are. *Philosophy of Science*, 85, p. 660-681.

Armstrong, D. (1968) *A materialist theory of mind*. London: Routledge.

_____. (1999) The open door. in Sankey, H. (ed.) *Causation in Science and the Methods of Scientific Inquiry*. Dordrecht, Kluwer, p. 175-185.

Asay, J. (2019) Going local: a defense of methodological localism about scientific realism. *Synthese*, 196, p. 587-609.

Ayres, M. (2021) Brazil central bank raises interest rates, signals more to come. URL = https://www.reuters.com/markets/us/brazil-central-bank-raises-rates-by-50-bps-signals-another-hike-2022-06-15/.

Azcarate, A. (2017). *Emotions as aesthetic properties of absolute music*. Universidad del País Vasco.

Banco Central do Brasil (2022) Monetary Policy Committee (Copom). URL = https://www.bcb.gov.br/en/monetarypolicy/committee.

Batterman, R. (2016) Intertheory relations in physics. *The Stanford Encyclopedia of Philosophy*. URL = https://plato.stanford.edu/entries/physics-interrelate/.

Bechtel, W. (2017) Explicating top-down causation using networks and dynamics. *Philosophy of Science*, 84, p. 253-274.

Bechtel, W., Abrahamsen, A. (2005) Explanation: a mechanistic alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, p. 421-441.

Becker, G. (1968) Crime and punishment: an economic approach. *Journal of Political Economy*, 76, p. 169-217.

Bennett, K. (2003). Why the exclusion problem seems intractable, and how, just maybe, to tract it. *Noûs*, 37, 471-497.

_____. (2004) Global supervenience and dependence. *Philosophical and Phenomenological Research*, 68, p. 510-529.

Berland, J. (1978) *Historical experience and the fictional imagination in the era of fascism: a study of Thomas Mann's Dr. Faustus*. M.A. Dissertation, Simon Fraser University.

Bernanke, B. (2015) "Inaugurating a New Blog". 30 March. URL = http://www.brookings.edu/blog/ben-bernanke/2015/03/30inaugurating-a-new-blog/.

Bernard, C. (1961 [1865]) *An introduction to the study of experimental medicine*. New York: Collier Books.

Bertrand, M. (2019) Metaphysical explanation by constraint. *Erkenntnis*, 84, p. 1325-1340.

Bickle, J. (1998) *Psychoneural reduction*: the new wave. Cambridge: MIT Press.

_____. (2020) Multiple realizability. *Stanford Encyclopedia of Philosophy*. URL = https://plato.stanford.edu/entries/multiple-realizability/.

Binmore, K. (2008) *Rational decisions*. New Jersey: Princeton University Press.

Bird, A. (2020) A dispositional account of causationwith some remarks on the ontology of disposition. in. Meincke, A. (ed.) *Dispositionalism*: perspectives from metaphysics and the philosophy of science. Exeter: Springer, p. 151-170.

Black, J. et.al. (2017) *Oxford dictionary of economics*. 5th Edition. Oxford: Oxford University Press.

Blanchard, O. et.al. (2017) *Macroeconomics*: *a European perspective*. 3rd Edition. Harlow: Pearson.

Block, N. (1990) Can the mind change the world? in. Boolos, G. (ed.) *Meaning and method*: essays in honor of Hilary Putnam. New York: Cambridge University Press, p. 137-170.

Bogen, J. (2005) Regularities and causality; generalisations and causal explanations. *Studies in History and Philosophy of Science Part C*: Studies in the History and Philosophy of Biological and Biomedical Sciences, p. 397-420.

Bokulich, A. (2014) How the tiger bush got its stripes: 'how-possibly' vs. 'how-actually' model explanations. *The Monist*, 97, p. 321-338.

Boland, L. (2017) *Equilibrium models in economics*. Oxford: Oxford University Press.

Boogerd, F. et.al. (2007) "Afterthoughts as Foundations for Systems Biology", in F. Boogerd, F. Bruggeman, J. Hofmeyr and H. Westerhoff (eds.), *Systems Biology*: Philosophical Foundations. New York: Elsevier, p. 321-336.

Bratman, M. (2022) *Shared and institutional agency*: toward a planning of human practical organization. Oxford: Oxford University Press.

Brisset, N. (2016) Institutions as emergent phenomena: redefining downward causation. *Halshs-01425669*.

Butterfield, J. (2011) Emergence, reduction and supervenience: a varied landscape. *Foundations of Physics*, 41, p. 920-959.

Carnap, R. (1937) *The logical syntax of language*. London: Harcourt.

Carter, M. and R. Maddock (1984) *Rational expectations*: macroeconomics for the 1980s? London: Macmillan.

Cartwright, N. (2007) *Hunting causes and using them*: approaches in philosophy and economics. Cambridge: Cambridge University Press.

_____. (2009) If no capacities then no credible worlds. But can models reveal capacities? *Erkenntnis*, 70, p. 45-58.

Causey, R. (1972) Attribute-identities in microreductions. *The Journal of Philosophy*, 69, p. 407-422.

Chick, V. (2016) On microfoundations and Keynes' economics. *Review of Political Economy*, 28, p. 99-112.

Chin-Parker, S. and A., Bradner (2017) A contrastive account of explanation generation. *Psychonomic Bulletin & Review*, 24, p. 1387-1397.

Churchland, Patricia (1986) *Neurophilosophy*. Cambridge: MIT Press.

Churchland, Paul (1979) *Scientific realism and the plasticity of mind*. Cambridge: Cambridge University Press.

_____. (1985) Reduction, qualia, and the direct introspection of brain states. *The Journal of Philosophy*, 78, p. 8-28.

Coleman, J. (1987) Microfoundations and macrosocial behavior. In. J. Alexander et. al. (eds.) *The micro-macro link*. Berkeley: University of California Press.

_____. (1990) *Foundations of social theory*. Cambridge: Belknap Press.

Collingwood, R. (1938) *The principles of art*. Oxford: Oxford University Press.

Cook, T. D. and Donald T. C. (1979) *Quasi-Experimentation:* design and analysis issues for field settings. Boston: Houghton Miflin Company.

Craver, C. (2006) When mechanistic models explain. *Synthese*, 153, p. 355-376.

_____. (2007) *Explaining the brain*: Mechanisms and the mosaic unity of neuroscience. Oxford: Oxford University Press.

Craver, C. and Bechtel, W. (2007) Top-down causation without top-down causes. *Biology and Philosophy*, 22, p. 547-563.

Crick, F. (1989) *What mad pursuit*: a personal view of scientific discovery. London: Penguin Books.

Cross, C. (1991) Explanation and the theory of questions. *Erkenntnis*, 34, p. 237-260.

Cudd, A. and Eftekhari, S. (2017) Contractarianism. *The Stanford Encyclopedia of Philosophy*.

Currie, A. (2017) From models-as-fictions to models-as-tools. *Ergo*, 4, p. 759-781.

Currie, G. (1984) Individualism and global supervenience. *The British Society for the Philosophy of Science*, 35, p. 345-358.

Davies, S. (1994). *Musical Meaning and Expression*. Cornell University Press.

Douglas, H. (2009) Reintroducing prediction to explanation. *Philosophy of Science*, 76, p. 444-463.

De Vroey, M. (2016) *A history of macroeconomics*: from Keynes to Lucas and beyond. Cambridge: Cambridge University Press.

Dietrich, F. and List, C. (2016) Mentalism versus behaviourism in economics: a philosophy-of-science perspective. *Economics and Philosophy*, 32, p249-281.

Dretske, F. (1972) Contrastive statements. *Philosophical Review*, 81, p. 411-437.

_____. (1991) Mental events as structuring causes of behaviour. in Heil, J. and A. Mele (eds.) *Mental causation*. Oxford: Oxford University Press.

Durkheim, E. (1938 [1895]) *The rules of sociological method*. New York: The Free Press.

Dizadji-Bahmani, F. et.al. (2010) Who's afraid of Nagelian reduction? *Erkenntnis*, 73, p. 393-412.

Epstein, B. (2009) Ontological individualism reconsidered. *Synthese*, 166, p. 187-213.

_____. (2014) Why macroeconomics does not supervene on microeconomics. *Journal of Economic Methodology*, 21, p. 3-18.

Fazekas, P. (2009) Reconsidering the role of bridge laws in inter-theoretical reductions. *Erkenntnis*, 71, p. 303-322.

Feyerabend, P. (1962) "Explanation, reduction, and empiricism". In G. Maxwell and H. Feigl (eds.), *Scientific explanation*, *space*, *and time*, Minnesota Studies in the Philosophy of Science: Volume III. Minneapolis: University of Minnesota Press, p. 28-97.

Fodor, J. (1974) Special sciences (or: the disunity of science as a working hypothesis). *Synthese*, 28, p. 97-115.

Forber, P. (2010) Confirmation and explaining how possible. *Studies in History and Philosophy of Science Part C*: *Studies in History and Philosophy of Biological and Biomedical Sciences*, 41, p. 32-40.

Frigg, R. (2010) Models and fiction. *Synthese*, 172, p. 251-268.

Fumagalli, R. (2013) The futile search for true utility. *Economics & Philosophy*, 29, p. 325-347.

_____. (2016) Five theses on neuroeconomics. *Journal of Economic Methodology*, 23, p. 77-96.

_____. (2020) How thin rational choice explains choices. *Studies in the History and Philosophy of Science*, 83, p. 63-74.

Galbács, P. (2020) *The Friedman-Lucas transition in macroeconomics*: a structuralist approach. London: Elsevier.

Gallow, J. (2022) The metaphysics of causation. *The Stanford Encyclopedia of Philosophy*. URL = https://plato.stanford.edu/entries/causation-metaphysics/.

Garfinkel, A. (1981) *Forms of explanation*: thinking the questions of social theory. New Haven: Yale University Press.

Giddens, A. (1992) Introduction. in. Weber, M. (1992) *The Protestant Ethic and the Spirit of Capitalism*. London: Harper Collins, p. vii-xxvi.

Gibbard, A. and H. Varian. (1978) Economic models. *The Journal of Philosophy*, 75, p. 664-677.

Giere, R. (1988) *Explaining science*: a cognitive approach. Chicago: Chicago University Press.

Glennan, S. (1996) Mechanisms and the nature of causation. *Erkenntnis*, 44, p. 49-71.

_____. (2002) Rethinking mechanistic explanation. *Philosophy of Science*, 69, p. S342-S353.

_____. (2009) Productivity, relevance, and natural selection. *Biology and Philosophy*, 24, p. 325-339.

Godfrey-Smith, P. (2006) The strategy of model-based science. *Biology and Philosophy*, 21, p. 725-740.

_____. (2009) Models and fictions in science. *Philosophical Studies*, 143, p. 101-116.

Gordon, R. (1993) *Macroeconomics*. 6th Edition. New York: Harper Collins College.

Green, S. and Jones, N. (2016) Constraint-based reasoning for search and explanation: strategies for understanding variation and patterns in biology. *Dialectica*, 70, p. 343-374.

Guala, F. (2019) Preferences: neither behavioural nor mental. *Economics & Philosophy*, 35, p. 383-401.

_____. (2022) Rescuing ontological individualism. *Philosophy of Science*, 00, p. 1-15.

Gul, F. and W. Pesendorfer (2008) The case for mindless economics. In A. Caplin and A. Schotter (eds.) *The Foundations of Positive and Normative Economics*. New York: Oxford University Press.

Hacking, I. (1983) *Representing and intervening*. Cambridge: Cambridge University Press.

Hall, R., Taylor, J. (1986) *Macroeconomics*. 3rd Edition. New York: Norton & Company.

Hare, R. (1952) *The language of morals*. Oxford: Oxford University Press.

Harsanyi, J. (1977) On the rationale of the Bayesian approach: comments of Professor Watkins' paper. In R. Butts and J. Hintikka (eds.) *Foundational Problems in the Special Sciences*. Dordrecht: D. Reidel Publishing.

Haslanger, S. (2015) What is a (social) structural explanation? *Philosophical Studies*, 173, p. 113-130.

Hausmann, D. (1998) *Causal asymmetries*. Cambridge: Cambridge University Press.

_____. (2011) *Preference*, *value*, *choice*, *and welfare*. Cambridge: Cambridge University Press.

Heckman, J. (2005) The scientific model of causality. *Sociological Methodology*, p. 1-97.

_____. (2008) Econometric causality. *International Statistical Review*, p. 1-27.

Hedstrom, P. and Ylikoski, P. (2010) Causal mechanisms in the social sciences. *The Annual Review of Sociology*, p. 49-67.

Heidl, S. (2016) *Philosophical problems of behavioral economics*. London: Routledge.

Hempel, C. (1965*) Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New Haven: Yale University Press.

Hempel, C. and P. Oppenheim (1948) Studies in the logic of explanation. *Philosophy of Science*, 15, p. 135-175.

Henderson, J. and R. Quandt. (1980) *Microeconomic theory*: *a mathematical approach*. New York: McGraw-Hill.

Henschen, T. (2018) What is macroeconomic causality? *Journal of Economic Methodology*, 25, p. 1-20.

Herger, N. (2019) Understanding central banks. Cham: Springer.

Hernes, G. (1998) Real virtuality. in Hedstrom, P. and R. Swedberg (eds.) *Social mechanisms*: an analytical approach to social theory. Cambridge: Cambridge University Press, p. 74-101.

Hobbes, T. (1968 [1651]) *Leviathan*. Harmonds-worth: Penguin Books.

Hodgson, G. (2007) Meanings of methodological individualism. *Journal of Economic Methodology*, 14, p. 211-226.

Holmes, D. (2018) A tractable future: central banks in conversation with their publics. in Beckert, J. and R. Bronk (eds.) *Uncertain futures*: imaginaries, narratives, and calculation in the economy. Oxford: Oxford University Press, p. 173-193.

Hooker, C. (1981) Towards a general theory of reduction. Part I: historical and scientific setting. Part II: Identity in reduction. Part III: Cross-categorical reduction. *Dialogue*, 20, p. 38-57, p. 201-236, p. 496-529.

Hoover, K. (2001a) *Causality in macroeconomics*. Cambridge: Cambridge University Press.

_____. (2001b) *The methodology of empirical macroeconomics*. Cambridge: Cambridge University Press.

_____. (2009) Microfoundations and the ontology of macroeconomics. In: Kincaid, H.; Ross, D. (eds.) *The Oxford Handbook of Philosophy of Economics*, p. 386-409.

_____. (2010) Idealization reduction: the microfoundations of macroeconomics. *Erkenntnis*, 73, p. 329-347.

_____. (2015) Reductionism in economics: intentionality and eschatological justification in the microfoundations of macroeconomics. *Philosophy of Science*, 82, p. 689-711.

Howitt, P. (1987) Macroeconomics: relations with microeconomics. in Eatwell, J. et.al. (eds.) *The New Palgrave*: a dictionary of economics. 1st Edition. London: Macmillan, p. 6107-6113.

Hull, D. (1976) "Informal aspects of theory reduction" in K. Schaffner and R. Cohen (eds.), *PSA 1976*: *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Dordrecht: Reidel, p. 653-670.

Hume, D. (1902 [1777]) *Enquiry concerning human understanding*. Oxford: Clarendon Press.

Huttemann, A. and A, Love. (2011) Aspects of reductive explanation in biological science: intrinsicality, fundamentally, and temporality. *The British Journal for the Philosophy of Science*, 62, p. 519-549.

Ihara, H. (2020) Supervenience and the mind-body problem in psychiatry. in Marková, I. and Chen, E. (eds.) *Rethinking Psychopathology*: theory and history in the human and social sciences. Cham: Springer, p. 117-127.

Jackson, F. (1995) Essentialism, mental properties and causation. *Proceedings of the Aristotelian Society*, 95, p. p. 253-268.

Janssen, M. (1993) *Microfoundations*: a critical inquiry.

_____. (2008) 'Microfoundations'. In: Steven N. Durlauf and Lawrence E. Blume (eds.) *The New Palgrave*: a Dictionary of Economics. 2nd edition. Basingstoke: Palgrave Macmillan, p. 1-10.

Jevons, S. (1911) *The theory of political economy*. 4th Edition. London: Elibron Classics.

Kahneman, D. (2011) *Thinking*, *fast and slow*. London: Allen Lane; Penguin.

Kania, A. (2017). The Philosophy of Music. *The Stanford Encyclopedia of Philosophy*. URL = https://plato.stanford.edu/archives/fall2017/entries/music/.

Kaul, N. (2022) 3Es for AI: Economics, explanation, epistemology. *Frontiers in Artificial Intelligence*, 5, p. 1-15.

Kelton, S. (2021) *The deficit myth*: Modern Monetary Theory and how to build a better economy. London: John Murray Press.

Khalidi, M. (2015) Three kinds of social kinds. *Philosophy and Phenomenological Research*, 90, p. 96-112.

Kim. J. (1988) Explanatory realism, causal realism, and explanatory exclusion. *Midwest Studies in Philosophy*, 12, p. 33-52.

_____. (1993) Concepts of supervenience. in _____. (ed.), *Supervenience and mind*: selected philosophical essays. Cambridge: Cambridge University Press.

_____. (2005) *Physicalism*, *or something near enough*. Princeton: Princeton University Press.

Kincaid, H. (1997) *Individualism and the unity of science*. Lanham, MD: Rowman and Littlefield.

_____. (1998) Supervenience. in Davis, J. et.al. (eds*.), The Handbook of Economic Methodology*. Cheltenham: Edward Elgar, p. 487-488.

King, J. (2012) *The microfoundations delusion*: metaphor and dogma in the history of macroeconomics. Cheltenham: Edward Elgar.

Kivy, P. (1980) *The corded shell*: reflections on musical expression. Princeton University Press.

_____. (1987). Platonism in Music: Another Kind of Defense. *American Philosophical Quarterly*, 24, 245–252.

Kizner, I. (1992) *The meaning of the market process*: essays in the development of modern Austrian economics. London: Routledge.

Klein, C. (2009) Reduction without reductionism: a defence of Nagel on connectability. *The Philosophical Quarterly*, 59, p. 39-53.

Klein, L. (1946) Macroeconomics and the theory of rational behavior. *Econometrica*, 14, p. 93-108.

Knnuutila, T. (2011) Modelling and representing: an artefactual approach to model-based representation. *Studies in History and Philosophy of Science*, 42, p. 262-271.

Kroedel, Thomas and Schulz, Moritz (2016). Grounding mental causation. *Synthese*, 193, p. 1909-1923.

Kuhn, T. (1962) *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lange, M. (2018) Because without cause: Scientific explanations by constraint. in Reutlinger, A. and J. Saatsi (eds.) *Explanation beyond causation*. Oxford: Oxford University Press, p. 15-38.

Lawson, T. (2001) Economics and explanation. *Revue Internationale de Philosophie*, 217, p. 371-393.

_____. (2009) Applied economics, contrast explanation and asymmetric information. *Cambridge Journal of Economics*, 33, p. 405-419.

Lebow, R. *Archduke Franz Ferdinand lives*! *A world without World War I*. New York: St. Martin Press, 2014.

Leonard, B. C., & Culture. (2019, March 8). *Political Songs | Alabama – John Coltrane*. New Frame. URL = https://www.newframe.com/political-songs-alabama-john-coltrane/.

Levinson, J. (1970) Aesthetic supervenience. *The Southern Journal of Philosophy*, 22, p. 93-109.

Levy, D. (1985) The impossibility of a complete methodological individualist: reduction when knowledge is imperfect. *Economics and Philosophy*, 1, p. 101-108.

Levy, A. (2012) Models, fictions, and realism: two packages. *Philosophy of Science*, 79, p.738-748.

_____. (2015) Modeling without models. *Philosophical Studies*, 172, p. 781-798.

Lewis, D. (1973) Causation. *Journal of Philosophy*, 70, p. 556-567. Reprinted in 1986. Page references to reprinted version.

_____. (1986) *On the plurality of worlds*. New York: Blackwell.

Lewis, P. (2013) Hayek, social theory and the contrastive explanation of socio-economic order. *Critical Review*, 25, p. 1-19.

Little, D. (1991) *Varieties of social explanation*: an introduction to the philosophy of social science. Boulder: Westview Press.

_____. (1998) *Microfoundations*, *method*, *and causation*: on the philosophy of social sciences. New Brunswick: Transaction Publishers.

List, C. and Spiekermann, K. (2013) Methodological individualism and holism in political science: a reconciliation. *The American Political Science Review*, 107, p. 629-643.

List, C. and Petit, P. (2006) Group agency and supervenience. *Southern Journal of Philosophy*, 44, p. 85-105

_____. (2011) *Group agency*: the possibility, design, and status of corporate agents. Oxford: Oxford University Press.

Longino, H. (2002) *The fate of knowledge*. Princeton: Princeton University Press.

Lucas, R. (1983 [1976]) Econometric policy evaluation: a critique. in Brunner, K. and A. Meltzer (eds.) *Theory*, *policy*, *institutions*: Papers from the Carnegie-Rochester Conference Series on Public Policy. Elsevier, p. 257-284.

_____. (1980) Methods and problems in business cycle theory. *Journal of Money*, *Credit*, *and Banking*, 12, p. 696-715.

_____. (1981) *Studies in business-cycle theory*. Cambridge: MIT Press.

_____. (1987) *Models of business cycles*. New York: Basil Blackwell.

Lukes, S. (1968) Methodological individualism reconsidered. *British Journal of Sociology*, 19, p. 119-129.

Marx, K. (1976 [1867]) *Capital*: a critique of political economy, volume 1. Harmondsworth: Penguin Books.

Machamer, P. (2004) Activities and causation: the metaphysics and the epistemology of mechanisms. *International Studies in the Philosophy of Science*, 18, p. 27-39.

Maki, U. (1992) On the method of isolation in economics. In Dilworth, C. (ed.) *Idealization IV*: *Intelligibility in Science*, *Poznan Studies in the Philosophy of the Sciences and the Humanities*, p. 317-351.

_____. (2009) MiSSing the world: models as isolations and credible surrogate systems. *Erkenntnis*, 70, p. 29-43.

_____. (2020) Puzzled by idealizations and understanding their functions. *Philosophy of the Social Sciences*, 50, p. 215-237.

Manicas, P. (2006) *A realist philosophy of social science*: explanation and understating. Cambridge: Cambridge University Press.

Mano, A. China suspends imports from two additional meat plants in Brazil amid Corona Virus fears. *The Pig Site*. URL = https://www.thepigsite.com/news/2020/07/china-suspends-imports-from-two-additional-meat-plants-in-brazil-amid-coronavirus-fears.

Mankiw, N. (1997) "An interview". In B. Snowdon and H. Vane (eds). *A macroeconomics reader*, p. 454-470. London: Routledge.

Mankiw, N.; Taylor, M. (2017) *Microeconomics*, 4th Edition. Hampshire: Cengage.

Marchionni, C. (2006) Contrastive explanations and unrealistic models: the case of the new economic geography. *Journal of Economic Methodology*, 13, p. 425-446.

_____. (2018) What is the problem with model-based explanation in economics? *Disputatio*, 9, p. 603-630.

McGeever, J. (2021) Brazil central bank hikes rates, eyes quicker return to “neutral” policy. URL = https://www.reuters.com/world/americas/brazil-central-bank-hikes-interest-rates-eyes-return-neutral-rate-2021-06-16/.

McLaughlin, C. (1996) Supervenience. In: Borchert (ed.), *Encyclopedia of Philosophy Supplement*. Macmillan, p. 558-560.

McLaughlin, B and Bennett, K. Supervenience. *Stanford Encyclopedia of Philosophy*, 2018. URL = https://plato.stanford.edu/entries/supervenience/.

McPherson, T. (2019) Supervenience in ethics. *Stanford Encyplopedia of Philosophy*. URL = https://plato.stanford.edu/entries/supervenience-ethics/.

Menger, C. (1963 [1883]) *Problems of economics and sociology*. Urbana-Champaign: University of Illinoi Press.

Menzies, P. and Price, H. (1993) Causation as a secondary quality. *British Journal for the Philosophy of Science*, 42, p. 187-203.

Mill, J. (1974 [1843-1872]) *A system of logic*. *Ratiocinative and inductive*. *Collected Works*, Vols VII-VIII. Toronto: University of Toronto Press.

Miller, T. (2019) Explanation in artificial intelligence: insights from the social sciences. *Artificial Intelligence*, 267, p. 1-38.

Morgan, M. (2002) Models, stories, and the economic world. in Maki, U. (ed.) *Fact and fiction in economics*: models, realism and social construction. Cambridge: Cambridge University Press, p. 178-201.

Morgan, J. and Patokämi, H. (2017) Contrast explanation in economics: its context, meaning, and potential. *Cambridge Journal of Economics*, 41, p. 1391-1418.

Morrison, M. (2008) "Models as representational structures". in S. Hartmann, C. Hoefer and L. Bovens (eds.), *Nancy Cartwright's Philosophy of Science*. New York: Routledge, pp. 67-90.

Nagel, E. (1935) The logic of reduction in sciences. *Erkenntnis*, 5, p. 46-52.

_____. (1961) *The structure of science*: problems in the logic of scientific explanation. Indiana: Hackett Publishing Company.

_____. (1970) Issues in the logic of reductive explanations. in H Kiefer and K Munitz (eds.), *Mind, Science, and History*. NY: SUNY Press, pp. 117–137.

Needham, P. (2010) Nagel's analysis: comments in defense as well as a critique. *Studies in History and Philosophy of modern Physics*, 41, p. 163-170.

Nelson, A. (1984) Some issues surrounding the reduction of macroeconomics to microeconomics. *Philosophy of Science*, 51, p. 573-594.

North, D. (1990) *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

Northcott, R. (2008a) Causation and contrast classes. *Philosophical Studies*, 139, p. 111-123.

_____. (2008b) Weighted explanations in history. *Philosophy of the Social Sciences*, 38, p. 76-96.

Northcott, R. (2012) Partial explanations in social science. in: Kincaid, H. (ed.) *The Oxford handbook of philosophy of social scie*nce. Oxford: Oxford University Press, p. 130-153.

Norton, J. (2012) Approximation and idealization: why the difference matters*. *Philosophy Science*, 79, p. 207-232.

Nozick, R. (2001) *Invariances*: the structure of the objective world. Cambridge: Harvard University Press.

O'Neill, R. et.al. (2017) *Inflation*: history and measurement. Cham: Palgrave Macmillan.

Ozga, S. (1965) *Expectations in economic theory*. London: Weidenfeld and Nicolson.

Paull, C. and Sider, T. (1992) In defense of global supervenience. *Philosophy and Phenomenological Research*, 52, p. 833-854.

Pearl, J. (2000) *Causality*. Cambridge: Cambridge University Press.

Phelps, E. (1970) Introduction: the new microeconomics in employment and inflation theory. in Phelps, E. et. al. (eds.) *Microeconomic foundations of employment and inflation theory*. New York: Norton & Company, p. 1-23.

Psillos, S. (2012) Regularity theories. in: Beebee, H. et.al. (eds.) *The Oxford Handbook of Causation*. Oxford: Oxford University Press, p. 131-157.

Poggi, G. (1983) *Calvinism and the capitalist spirit*: Max Weber's Protestant Ethic. Amherst: The University of Massachusetts Press.

Post, J. (1995) "Global" supervenient determination: too permissive? In: E. Savellos and Yalçin, Ü. (eds.) *Supervenience*: New Essays. Cambridge: Cambridge University Press, p. 73-100.

Pratt, R. (2000) *Closing the gap between music and history in Thomas Mann's Doktor Faustus*. PhD Thesis in Philosophy, University of London.

Rabin, M. (2002) A perspective on psychology and economics. *European Economic Review*, 46, p. 657-685.

Redman, D. (1992) *A readers guide to rational expectations*. Aldershot: Edward Elgar.

Reiss, J. (2004). Review of Kevin Hoover's methodology of empirical macroeconomics. *Economics and Philosophy*, p. 226-233.

_____. (2012) Counterfactuals. In: Kincaid, H. (ed.) *The Oxford handbook of philosophy of social science*. Oxford: Oxford University Press, p. 154-183.

_____. (2012) The explanation paradox. *Journal of Economic Methodology*, 19, p. 43-62.

Reutlinger, A. (2016) Is there a monist theory of causal and noncausal explanations? The counterfactual theory of scientific explanations. *Philosophy of Science*, 83, p. 733-745.

Ricardo, D. (1911 [1811]) *The principles of political economy & taxation*. New York: J.M.Dent & Sons.

Rosenberg, A. (1976) *Microeconomic laws*: a philosophical analysis. University of Pittsburgh Press.

Ross, D. (2014) *Philosophy of Economics*. Hampshire: Palgrave Macmillan.

Ross, L. (in Review) What is a social structure explanation? A causal account. URL = http://philsci-archive.pitt.edu/19721/.

Roznai, Y. and Kreuz, L. (2018) Conventionality control and Amendment 95/2016: a Brazilian case of unconstitutional amendment. *Revista de Investigações Constitucionais*, 5, p. 35-56.

Ruiz, N. (2021) Social ontology and model-building: a response to Epstein. *Philosophy of the Social Sciences*, 51, p. 176-192.

Saatsi, J. and Pexton, M. (2012) Reassessing Woodward's account of explanation: regularities, counterfactuals, and noncausal explanations. *Philosophy of Science*, 80, p. 613-624.

Salmon, W. (1984) *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.

_____. (1989) *Four decades of scientific explanation*. Pittsburgh: Pittsburgh University Press.

Samuelson, P., Solow, R. (1960) The problem of achieving and maintaining a stable price level: analytical aspects of anti-inflation policy. in _____. AER, 50, *Readings in Macroeconomics*. New York: Holt, Rinehart and Winston.

Sargent, T. (1979) *Macroeconomic theory*. New York: Academic Press.

Sarkar, S. (1992) Models of reduction and categories of reductionism. *Synthese*, 91, p. 167-194.

_____. (2015) Nagel on reduction. *Studies in History and Philosophy of Science*, 53, p. 43-56.

Schaffner, K. (1967) Approaches to reduction. *Philosophy of science*, 34, p. 137-147.

_____. (1993) *Discovery and explanation in biology and medicine*. Chicago: The University of Chicago Press.

_____. (2012) Ernst Nagel and reduction. *The Journal of Philosophy*, 109, p. 534-565.

Schumpeter, J. (1908) *Das Wesen und der Hauptinhalt der theoretischen Nationalökonomie. Leipzig: Duncker & Humbolt.*

_____. (1909) On the concept of the social value. *Quarterly Journal of Economics*, 21, p. 213-232.

Searle, J. (1995) *The construction of social reality*. London: The Pinguin Press.

Shaffer, J. (2005) Contrastive causation. *The Philosophical Review*, 114, p. 327-358.

Shagrir, O. (2002) Global supervenience, coincident entities and anti-individualism. *Philosophical Studies*, p. 171-196.

Shapiro, L. (2017) Mechanism or bust? Explanation in Psychology. *British Journal for the Philosophy of Science*, 68, p. 1037-1059.

Shaw, G. (1984) *Rational expectations*: an elementary exposition. Michigan: Wheatsheaf Books.

Shiller, R. (2019) *Narrative economics*: how stories go viral & drive major economic events. Oxford: Princeton University Press.

Sider, T. (1999) Global supervenience and identity across times and worlds*. *Philosophy and Phenomenological Research*, 59, 913-937.

Sklar, L. (1993) *Physics and chance*: philosophical issues in the foundations of statistical mechanics. Cambridge: Cambridge University Press.

Smith, A. (1976 [1776]) *An inquiry into the nature and causes of the wealth of nations*. Chicago: University of Chicago Press.

Snowden, B. et. al. (1994) *A modern guide to macroeconomics*. Aldershot: Edward Elgar.

Stalnaker, R. (1996) Varieties of supervenience. *Philosophical Perspectives*, 10, p. 221-241.

Stanford, P. (2021) Realism, instrumentalism, particularism: a middle path forward in the scientific realism debate. In Lyons, T. and Vickers, P. (eds.) *Scientific realism*: the challenge from the history of science, p. 216-238. New York: Oxford University Press.

Steel, D. (2004) Can a reductionist be a pluralist? *Biology and Philosophy*, 19, p. 55-73.

Stoljar, D. (2008). Distinctions in distinction. In *Being Reduced: New Essays on Reduction, Explanation, and Causation* (edited by Jakob Hohwy and Jesper Kallestrup), p. 263-279. New York: Oxford University Press.

_____. (2021) Physicalism. *Stanford Encyclopedia of Philosophy*. URL = https://plato.stanford.edu/entries/physicalism/.

Sugden, R. (2000) Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology*, 7, p. 1-31.

_____. (2016) Methodological individualism, and the foundations of the social sciences. *Journal of Economic Literature*, 54, p. 1377-1389.

Suppe, F. (1977) The search for philosophic understanding of scientific theories. in _____. *The structure of scientific theories*, 2nd Edition. Urbana and Chicago: University of Illinois Press, p. 1-241.

Tabery, J. (2004) Synthetizing activities and interactions in the concept of a mechanism. *Philosophy of Science*, 71, p. 645-664.

Takho, T. (2021) *Unity of science*. Cambridge: Cambridge University Press.

Tawney, R. (2015) *Religion and the rise of capitalism*. London: Verso.

Taylor, J. (1993) Discretion versus policy rules in practice. *Carnegie-Rochester Conferences Series on Public Policy*, 39, p. 195. 214.

Thoma, J. (2021) In defence of revealed preference theory. *Economics & Philosophy*, 37, p. 163-187.

Till Grune-Yanoff and Phillipe Verrault-Julien (2021) How-possibly explanations in economics: anything goes? *Journal of Economic Methodology*, 28, p. 114-123.

Toon, A. (2012) *Models as make-believe*: imagination, fiction, and scientific representation. Basingstoke: Palgrave Macmillan.

Tsang, E. and Ellsaesser, F. (2011) How contrastive explanation facilitates theory building. *The Academy of Management Review*, 36, p. 404-419.

Turner, S. (1983) Weber on action. *American Sociological Review*, 48, p. 505-519.

Udehn, L. (2002) The changing faces of methodological individualism. *Annual Review of Sociology*, 28, p. 479-507.

van Fraassen, B. (1980) *The scientific image*. New York: Oxford University Press.

van Gullick, R. (2001) Reduction, emergence, and other recent options on the mind/body problem. *Journal of Consciousness Studies*, 8, p. 1-34.

van Gulick, R. and R. van Riel (2019) Scientific reduction. *Stanford Encyclopedia of Philosophy*. URL = https://plato.stanford.edu/entries/scientific-reduction/.

van Riel, R. (2011) Nagelian reduction beyond the Nagelian model*. *Philosophy of Science*, 78, p. 353-375.

_____. (2014) *The concept of reduction*. Heidelberg: Springer.

Verrault-Julien, P. (2019) How could models possibly provide how-possibly explanations? *Studies in History and Philosophy of Science*, 73, p. 22-33.

Vickers, P. (2019) Towards a realistic success-to-truth inference for scientific realism. *Synthese*, 196, p. 571-585.

von, Bertalanffy, L. (1933) *Modern theories of development*. Oxford: Oxford University Press.

Walras, L. (1954 [1926]) *Elements of pure economics*. London: George Allen & Unwin.

Walters, A. (1971) Consistent expectations, distributed lags and the quantity theory. *Economic Journal*, 81, p. 273-281.

Walton, K. (1970) Categories of art. *Philosophical Review*, 79, p. 334-369.

Waters, C. (1990) Why the anti-reductionist consensus won't survive: the case of classical Mendelian genetics. In A. Fine, M. Forbes and L. Wessels (eds.), *PSA 1990*: *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1, p. 125-139. East Lansing: Philosophy of Science Association.

Watkins, J. (1953) Ideal types and historical explanation. *Readings in the Philosophy of Science*. (ed.) H. Feigl and M. Brodbeck. New York: Appleton-Century-Crofts, p. 723-743.

_____. (1968) Methodological individualism and social tendencies. in Brodbeck, M. (ed.) *Readings in the Philosophy of the Social Science*, p. 269-279.

Weber, M. (1968 [1922]) *Economy and society*, vol. 1. New York: Bedminster Press.

_____. (1976) *The protestant ethic and the spirit of capitalism*. Guildford: George Allen & Unwin.

_____. (1978 [1922]) *Economy and society*, vol. 1. Berkeley: University of California Press.

_____. (1992) *The Protestant Ethic and the Spirit of Capitalism*. London: Harper Collins.

Wilkinson, S. (2014) Levels and kinds of explanation: lessons from neuropsychiatry. *Frontiers in Psychology*, 5, p. 1-9.

Wilson, J. (2021) *Metaphysical emergence*. Oxford: Oxford University Press.

Wimsatt, W. (1972) "Complexity and organization", in K. Schaffner and R. Cohen (eds.), *PSA 1972*: *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Dordrecht: Reidel, p. 67-86.

_____. (2007) *Re-engineering philosophy for limited beings*. Cambridge: Harvard University Press.

Winther, R. (2009) Schaffner's model of theory reduction: critique and reconstruction*. *Philosophy of Science*, 76, p. 119-142.

Woodward, J. (2003) *Making things happen*: a theory of causal explanation. Oxford: Oxford University Press.

_____. (2013) Mechanistic explanation: its scope and limits. *Proceedings of the Aristotelian Society*, 87, p. 39-65.

_____. (2018) Some varieties of noncausal explanation. In Reutlinger, A. and J. Saatsi (eds.) *Explanation beyond causation*: *Philosophical perspectives on non-causal explanations*. Oxford University Press, p. 117-140.

Yablo, S. (1992). Mental causation. *Philosophical Review*, 101, p. 245-80.

Ylikoski, P. (2012) Micro, macro, and mechanisms. In: Kincaid, H. (Ed.) *The Oxford handbook of philosophy of social scie*nce. Oxford: Oxford University Press, p. 21-45.

_____. (2013) Causal and constitutive explanation compared. *Erkenntnis*, 78, p. 277-297.

Ylikoski, P. and Aydinonat, E. (2014) Understanding with theoretical models. *Journal of Economic Methodology*, 21, p. 19-36.

Zahle, J. (2016) The methodological holism in the social sciences. *The Stanford Encyclopedia of Philosophy*. URL = https://plato.stanford.edu/entries/holism-social/.