universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Understanding with economic models

verfasst von / submitted by

Arne Alexander Ruge, B.Sc.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 642
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Philosophy and Economics / Master's degree programme Philosophy and Economics
Betreut von / Supervisor:	Univ.-Prof. Tarja Knuuttila, MSc M.Soc.Sc PhD

Contents

1. Introduction 3
1.2. Structure and chapter summary 4
2. The explanation paradox 6
2.1 The explanation paradox 6
2.2 Methodological positions on economic modelling 8
2.3 (Model-based) Explanation in economics 9
2.4 Summary, observations and outlook 12
3. Scientific understanding: an overview 14
3.1 The history of scientific understanding 14
3.2 What is scientific understanding? 15
3.2.1 Explanatory understanding 15
3.2.2. Objectual understanding 17
3.3 Summary and outlook 19
4. Why traditional accounts of understanding are unsuited for theoretical modelling ... 20
4.1 A trade-off between explanation and epistemic accessibility 20
4.2 Kinds of phenomena – theoretical models investigate general phenomena 22
4.2.1 General and generic models 24
4.3 How-possibly explanations in economics 28
4.4. Summary and implications 29
5. Understanding with theoretical models 31
5.1 True causal explanation is not necessary for scientific understanding 31
5.1.2. Exemplary cases of understanding without HAE 33
5.2. A broader account of scientific understanding 35
5.3 Summary 38
6. Conclusion 40
References 42
Appendix 46
English abstract 46
German abstract 46

1. Introduction

Highly simplified and abstract theoretical models play an important role in many sciences. Economics is no exception. It relies heavily on theoretical, mathematical modelling. Yet even though the use of such models is ubiquitous, there is fundamental disagreement if and how economists can really learn from them about instances of the real world.

Theoretical models are usually approached as abstract representations of specific target systems. Often, such representations are approached through metaphors of a portrait, a map, or a mirror (Knuuttila 2011). Hence, one may use the representation, the model, to draw inferences about the represented, the real world. Usually, it is argued that such inferences are granted if the model shares the relevant features with its target system. However, such a view comes with many problems. Particularly in economics, it is hard to sustain as many theoretical models present a strongly distorted and inaccurate picture of the real world. "All models are wrong, but some are useful" (Box and Draper 1987, 424) is a well-known wisdom. If this is the case, how can inferences from the artificial model world to the real world be established? In other words, what justifies economists to rely on theoretical models as source of explanation for and scientific understanding of real-world phenomena?

In an influential essay, Julian Reiss (2012) discusses the explanatory capabilities of theoretical models in economics. Reiss argues that theoretical modelling in economics relies on three mutually inconsistent hypotheses. First, economic models are false, second, economic models are nevertheless explanatory, third, only true accounts can explain. Even though Reiss argumentation is rather unorthodox, it captures the essence of the problem well. Many economic models represent a strongly distorted picture of reality which makes it doubtful in how far explanations of real-world phenomena can be inferred from them.

This master thesis contributes a new perspective on the epistemic value of theoretical models. I will develop a position that accepts Reiss' paradox as posed but rejects his second premise. Thus, theoretical models themselves are not explanatory. Consequently, an answer must be provided to the following research question: if theoretical models are not explanatory, what does their epistemic value derive from?

I will argue that theoretical models can afford scientific understanding. However, theoretical models do not afford understanding via the usually assumed mechanism of providing explanation. Instead, a careful differentiation between how-actually and how-possibly explanations is imperative (see i.e. Grüne-Yanoff 2013). Theoretical economic models often only provide the latter. By reference to paradigmatic cases of understanding without how-actually explanation, I will argue that economic models can afford understanding of real-world

phenomena even if they lack an accurate representational relationship to singular occurrences of some particular phenomenon. Thus, the focus on the representational relationship between theoretical models and singular real-world occurrences can be wrong-headed as theoretical models aim to investigate phenomena in a more general manner. A central argument of this thesis then consists in detaching scientific understanding from the often-assumed necessity of true explanation. I will suggest that an appropriate definition of understanding needs to be chosen with respect to the kind of object that is to be understood. If this object is an individual occurrence, the traditionally assumed definition based on the conditions of how-actually explanation and epistemic accessibility is legitimate. However, understanding general phenomena requires different criteria. In these cases, how-possibly explanations providing the necessary modal information can suffice.

In terms of the initially presented explanation paradox and the research question posed above, my discussion yields two fundamental insights. First, Reiss' paradox can be resolved by rejecting his second premise: (many) theoretical models are not explanatory in the sense of providing how-actually explanation. They can, however, provide how-possibly explanations. Second, this means that Reiss' focus on explanation is short-sighted. Theoretical models are epistemically valuable also by affording understanding of general phenomena.

1.2. Structure and chapter summary

The second chapter introduces Reiss' explanation paradox[1] (2012) and embeds it into the larger philosophical framework. By reference to EP, I will systematize existing accounts that propose solutions to the question how inferences from the model world to the real world can be established. Furthermore, I will provide arguments why the explanatory capabilities of theoretical models can be quite limited. Hence, this chapter lays the foundation for the development of a position that rejects Reiss' second premise and seeks to spell out the epistemic value of theoretical models in terms of scientific understanding[2].

Chapter three introduces the concept of understanding and provides an overview of different accounts. The concepts of explanatory understanding (EU) and objectual understanding (OU) will be differentiated while highlighting that most traditional accounts of understanding conceptualize it in terms of EU. I will show that such accounts proclaim three necessary criteria for understanding: veridicality, explanation, and epistemic accessibility. Yet, if this close tie between explanation and understanding is advocated, non-explanatory models are necessarily

[1] Referred to as EP from here onwards.

[2] I will use the term understanding as short for scientific understanding throughout this thesis.

excluded from affording scientific understanding as well. Thus, if theoretical models cannot explain, spelling out their epistemic value in terms of understanding is impossible.

Chapter four questions the impossibility of gaining understanding without explanation. In this chapter, I will present three arguments which show that traditional accounts of scientific understanding are unsuited for theoretical modelling in economics. The first argument exposes potential trade-offs between the conditions of explanation and epistemic accessibility. The second argument takes a closer look at the type of phenomena that theoretical models usually investigate. In many cases, these are general phenomena, but not particular empirical occurrences. Consequently, the third argument emphasizes the explorative role of theoretical models in suggesting that they offer how-possibly explanations (HPE) of some phenomenon.

The fifth chapter substantiates these claims. By reference to paradigmatic cases of understanding without how-actually explanations (HAE), HAEs are rejected as necessary criteria for understanding. Instead, I will propose a broader account of understanding, in which the criterion of veridicality is substituted for a criterion of possibility. I will suggest choosing the constitutive criteria of understanding with respect to the kind object that is to be understood. If this object is a particular empirical occurrence, defining understanding in terms of a correct (causal) explanation and an element of grasping is appropriate. However, if the object is a general phenomenon, then the proposed broader account applies. Consequently, theoretical economic models can afford understanding of general phenomena.

The final chapter concludes the essential findings of this thesis.

2. The explanation paradox

The practice of using highly simplified theoretical models to learn about the world is ubiquitous in economics. It is both powerful and puzzling. One of the central questions surrounding this practice concerns the legitimacy of drawing explanatory inferences from the model onto the real world. How can these be justified? Standard representational approaches assume that models can give us knowledge about the world since they represent their target systems. Given that theoretical models in economics usually present strongly distorted representations of reality, accurate representation does not seem to provide the answer.

This chapter uses Julian Reiss' explanation paradox (2012) to introduce the debate on the explanatory capabilities of theoretical models and to review different attempts to solve it (part 2.1 and 2.2). Part 2.3 embeds Reiss' paradox into a larger philosophical framework. Part 2.4 summarizes important takeaways. Thereby this chapter will provide the necessary background for the discussion in the following chapters.

2.1 The explanation paradox

In an influential essay, Julian Reiss (2012) discusses the explanatory capabilities of theoretical models in economics. He investigates one specific model: 'Hotelling's law' (Hotelling 1929). Reiss' argumentation specifically concerns highly simplified theoretical models in economics for which he takes Hotelling's model as a paradigm case. In essence, Reiss observes that theoretical modeling in economics is based on three hypotheses (Reiss 2012, 49):

(1) Economic models are false
(2) Economic models are nevertheless explanatory
(3) Only true accounts can explain

Given that these three statements seem individually acceptable but are contradictory in combination, Reiss claims to have found a paradox - the explanation paradox. Reiss chose an unorthodox way to discuss this issue. Hence, it is necessary to investigate his hypotheses in more depth.

Reiss argues that economic models are false. Consequently, his argumentation implies that economic models are not true. Introducing the notion of truth in this regard is controversial. Reiss himself admits that models themselves do not have truth values and hence, cannot be true or false. Thus, hypothesis (1) is to be understood "elliptically" (Reiss 2012, 49). He argues that economic models often misrepresent their target system significantly. Therefore, one may either

say that important assumptions that define model M (and are thus necessarily true for M) are false of the target system T or that a theoretical hypothesis stating that M is like T is false (ibid., 49). This is obviously not a trivial question. Every scientific model contains some forms of simplification and/or idealization. Whether or not this model can still be considered an accurate representation of its target system (and thus be true in Reiss' terminology) is up for debate. Reiss borrows a classification of different kinds of idealizations by William Wimsatt (Wimsatt 2007, 101-102). He argues that Hotelling's model is false in all five respects of Wimsatt's classification. These concern, first, the area of application meaning that the model is applied too broadly. Second, the model uses idealizations which do not function as proper approximations. Third, the model is incomplete in the sense of leaving out causally relevant variables. Fourth, it displays a misdescription of interaction of variables. Fifth, it gives a totally wrong-headed picture of nature due to a significant number of non-existing entities and properties.

Nevertheless, Reiss claims that Hotelling's model is considered explanatory or at least that "it feels explanatory" (Reiss 2012, 49). Thus, hypothesis (2) is based on his observation that economists aim to explain real world phenomena using theoretical models. Reiss assumes a causal account of explanation in economics. He claims that this account is "widely regarded as successful and, importantly, more successful than its alternatives" (ibid., 43). Therefore, explaining an economic phenomenon requires stating the actual causes responsible for it. By reference to Nancy Cartwright (1983), he argues that causal explanations can only be successful, if they are true: "Cheap money in the early 2000s does not explain the financial crisis of the late 2000s unless money was indeed cheap (in the sense that interest rates were lower than the rate that would have been 'adequate' given the economic conditions), and unless cheap money was indeed the factor without which the financial crisis would not have occurred" (Reiss 2012, 43). The result is the hypothesis (3): only true accounts can explain.

Thus, Reiss arrives at his paradox, which questions the explanatory capabilities of theoretical models and hence, their epistemic value altogether. However, it needs to be emphasized that Reiss' argumentation is contended in the literature. For example, Mäki (2013) argues that the validity of EP strongly depends on Reiss' notions of truth and explanation and that the way Reiss uses them is at least partly ambiguous. Yet, independently of one's view on the paradox's correctness, it is a helpful tool to systematize methodological accounts on the relationship of economic models and real-world phenomena. In other words, EP helps to see how the literature has addressed the issue from different angles. The following section discusses three kinds of solution attempts.

2.2 Methodological positions on economic modelling

Existing accounts aiming to find a solution to EP can be categorized into three classes. Each class rejects one of the premises of EP.

The first class of accounts rejects that economic models are false. At least, these accounts defend the possibility that models are (approximately) true despite containing many falsehoods. Even though models misrepresent their target system in many ways, they accurately represent the important ones. Therefore, models may be considered 'true in the abstract'. Nancy Cartwright (i.e. 2009) and Uskali Mäki (i.e. 1994) are prominent authors holding such a view. In essence, they argue that the central aim of a model is to isolate a causal mechanism via idealisation. Then, a model explains real-world phenomena based on the truth of the assumptions about the causal factor it has isolated. Hence, there is a correct representation of the explanatory factors. The underlying idea is that theoretical models work like Galilean thought experiments. However, the problem is that economic models differ from these thought experiments in a significant way, as Cartwright herself has pointed out (2009). She argues that idealisations in economic models do not only eliminate possible interfering causes to study a certain tendency in isolation, but instead determine the model's overall structure, which drives its results.

The second class of accounts denies that theoretical economic models themselves are explanatory. Daniel Hausman (1992) argues that models are mainly used for conceptual exploration. Conceptual exploration means an investigation of a model's internal properties without the direct aspiration to explain a particular phenomenon. Then, models are useful insofar as inexact generalizations can be deduced from them, which can be tested empirically. Anna Alexandrova (2008) views models as open formulae. Thus, models only have a heuristic role in the formulation of causal hypotheses. In a second step, these hypotheses can be tested empirically. Till Grüne-Yanoff (2009) argues that models prove modal hypotheses. In other words, models establish possibilities of counter-dependence. By reference to Schelling's checkerboard model (Schelling 1978), Grüne-Yanoff claims that economists learn from theoretical models by establishing new causal possibilities. In Schelling's case, urban segregation can result without racial preferences. However, Schelling's model is not supposed to explain a particular instance of segregation in the real world.

The third class of accounts rejects Reiss' last hypothesis. Hence, false accounts can have explanatory power, as Robert Sugden (2000, 2009, 2013) has argued. Sugden views models as deliberately constructed parallel worlds. Thus, models are not stripped-down descriptions of the real world even though economists aspire to explain real-world phenomena with them.

Hence, in Reiss' terminology, models are false but explanatory. Sugden argues that the gap between the model world and the real world needs to be bridged by inductive inference. He suggests the following explanatory schema (Sugden 2000, 19):

E1 – in the model world, R is caused by F.
E2 – F operates in the real world.
E3 – R occurs in the real world.

Based on these observations, Sugden argues that there is reason to believe that R is caused by F in the real world as well. The confidence in such inferences is based on the model's credibility. A model's credibility is determined by its similarity to the real world and its robustness. Thus, 'false' models are explanatory by virtue of an inductive inference justified by a model's credibility. However, Sugden does not propose universal and objective criteria that determine a model's credibility. Hence, there are no necessary and sufficient conditions whether a model counts as explanatory. Inevitably, such judgements are left to the community of scientists and thereby become subjective to a certain degree.

This classification shows that EP is a helpful tool to systematize accounts that define models' relationship to the real world, whether they should or should not explain and what it takes to explain successfully. It also shows that there is significant disagreement on these issues. The following section discusses explanation in economics in more detail. The focus will be on model-based explanation.

2.3 (Model-based) Explanation in economics

Debating the explanatory power of theoretical models and the necessary requirements that need to be assigned to a model to count as explanatory quickly turn into debates about the fundamental goal of economics as a science. Some authors even questioned whether economics should pursue the epistemic goal of explanation at all. Indeed, it was Milton Friedman's methodological mission to show that the use of unrealistic assumption is not a problem if correct predictions can be drawn from them (Friedman 1953). Consequently, Friedman's instrumentalist position downplayed the importance of explanation as an epistemic goal in economics.[3]

[3] Interestingly, Friedman uses "explain" in scare quotes throughout his essay.

The decisive point is the open question whether economics should mostly pursue practical or epistemic goals. Interestingly, Reiss himself states that one of goals of EP was strategic, aiming to question the realist' project in the philosophy of economics. Hence indirectly, EP is posed in a way that it gives reason to follow an instrumentalist' account in economics like Reiss himself does (Reiss 2012b).

Today explanation is a widely acknowledged goal in economics. "Much of economics is positive: It just tries to explain how the economy works" (Mankiw 2018, 28). Given this statement from a standard textbook, it seems unquestionable that economics as a science aims to explain the actual world. The currently dominant account of explanation in economics is causal (see i.e. Reiss 2012). Thus, it is generally assumed that citing the causes of a phenomenon is necessary for explaining it. Yet even though there is a general consensus that economics should provide causal explanations, there is disagreement what form these explanations should take. In other words, the exact notion of causation is unsettled (Verreault-Julien 2021, 302). For instance, some authors defend causal explanations that refer to mechanisms (i.e. Hedström and Ylikoski 2010), while others favor causal laws (i.e. Hausman 1992). It is also unsettled whether one specific notion of causality is needed or whether multiple notions can successfully coexist (Verreault-Julien 2021, 302). Finally, it is noteworthy that new suggestions in the literature also explore non-causal accounts of explanation (Verreault-Julien 2017).

Today many methodological debates center around the topic of idealization and model-based explanation. Much of the popular criticism on conventional economics derives from the fact that highly simplified and idealized economic models are viewed as unrealistic or false. This may not pose a large threat to economics if the fundamental indicator of success is a model's predictive power as is the case in a Friedman-type instrumentalist account. If, however, models are supposed to explain actual phenomena, this may be problematic. Thus, commitment to the pursuit of (model-based) explanation as an epistemic goal in economics requires clarifying how and under which criteria explanatory inferences may be drawn.

There is a long tradition in the philosophy of science requesting that the explanans of a scientific explanation needs to be true. This is the case in the classical deductive-nomological model (see Hempel 1965) as well as more recent literature focusing on a causal explanation. A good example is provided by Strevens (2008, 297) who states: "no causal account of explanation (…) allows for non-veridical models to explain." In other words, explanation requires factivity. This requirement is not bound to a causal account of explanation. Bokulich (2016) highlights that a careful differentiation between a conception of explanation and an account of explanation is imperative. By reference to Salmon, she differentiates an ontic, a

modal and an epistemic conception of explanation. A conception of explanation is concerned with the question what explanations are. Hence, it is concerned with their ontological status. By contrast, an account of explanation spells out how explanations do their work. The factivity requirement of explanation is bound to a particular conception of explanation, namely the ontic conception. Yet, it is not bound to a particular account of explanation. Therefore, one way of solving EP is rejecting an ontic conception of explanation. This is, in essence, the solution Sugden that proposes. Sugden commits to a subjective notion of explanation, which equals an epistemic conception in Salmon's terms (see also Reiss 2013, 290-291 for this interpretation). Consequently, he pursues an instrumentalist strategy, in which the judgement of a scientists' community decides in how far a model has been successful in discovering regularities in its domain. Based on this success criterium, a model is judged as credible and consequently possesses explanatory power. Hence, Sugden claims that economic models are explanatory, but it seems that his conditions of explanatory power are less demanding.

If one commits to a conception of explanation that requires factivity, Sugden's strategy is not valid. Then, a model can only explain a real-world phenomenon by fulfilling the factivity requirement itself. It requires veridical models. Therefore, if theoretical models in economics aim to explain, there might be a need for accurate (or factive) representation of the target system. Otherwise, the explanatory power of a model is questionable in case it misrepresents the very thing to be explained. Verreault-Julien (2021, 304) underlines this by arguing that debates around EP are mostly debates in how far theoretical models satisfy the factivity requirement. Thus, EP could be reformulated as follows:[4]

(1) Models are highly idealized; they misrepresent reality
(2) Models explain phenomena
(3) Explanations are factive

Consequently, one of the fundamental challenges of the 'realism program' in economic methodology is to show that economic theory (including economic modeling) is based on and committed to (some form) of truth or factivity (see i.e. Mäki 1992). Thus, accounts rejecting premise (1) of Reiss' original paradox uphold the factivity of explanation and defend models' explanatory power based on a faithful and factive representation of decisive causal factors. Accounts rejecting premise (2) must conclude that economists are just wrong concerning the

[4] The reformulation is taken from (Verreault-Julien 2021, 304) although presented in a slightly different way here.

explanatory power of their models and need to spell out other epistemic purpose(s) that models fulfil. Accounts rejecting premise (3) will need to bear the consequences of adopting a specific, rather non-standard, conception of explanation.

2.4 Summary, observations and outlook

There are several interesting takeaways from the debate around EP. One of the central characteristics of many contributions to the debate about model-based explanation in economics is that these assume that the falsity of models' assumptions imply that they cannot provide actual explanation. Marchionni (2017) observes that such a framing disregards an important differentiation. On the one hand, there is a *conceptual issue* exploring the requirements of a model to count as explanatory and on the other hand, an *epistemological issue* concerning the justification for believing that the model possibly, probably, or certainly satisfies these requirements (ibid., 604). Concerning the *conceptual issue*, even if there is agreement that a model needs to provide the actual explanation of some phenomenon, a certain ambivalence remains. Explanations can either be complete providing all causal difference-makers or partial stating only one or some of them. Furthermore, requirements are dependent on the type of explanandum. Economics often aims to explain stylized facts and is then rather concerned with general phenomena and not specific occurrences. Furthermore, causes can be reported at different levels of abstraction. In other words, there are different levels of abstraction to explain the same phenomenon. There is considerable agreement in the literature that such abstraction, for instance to find features that different occurrences share with one another, can be valuable (ibid., 610). The *epistemological dimension* is reflected in the differentiation between actual and potential explanations. Whereas an actual explanation requires a true explanans, veridicality is not certain (or even unknown) in potential explanations. However, assessing whether explanatia are true is often difficult. For several reasons, economics might even be a particularly peculiar case. Marchionni argues that in economics, models' assumptions are often indispensable for the derivation of results. Hence, it is often not possible to de-idealize a model. Furthermore, she claims that there is a tenuous relation between theoretical and empirical models. This means that in many cases it is not possible to test how well theoretical models fit their real-world targets. Hence, Marchionni argues that "economists tend to be content with casual empiricism" (ibid., 607). Thus, theoretical modelling in economics might raise specific concerns. In particular, it is difficult to assess the explanatory capabilities in presence of false assumptions relevant to the conclusions and in absence of tested empirical confirmation (ibid., 606-607). Consequently, authors need to define what exactly they mean if they claim that a

model can (or cannot) explain a certain phenomenon. Yet, the debate goes even deeper. There is disagreement on how economics can or should explain and what role is left to models in this regard. As a result, there is no consensus on *whether* theoretical models in economics can explain real-world phenomena.

In the following chapters, I will develop a position that takes EP as it was posed and rejects Reiss' second premise. Hence, theoretical models alone cannot explain real-world phenomena. The central goal of this thesis is to spell out the epistemic value of theoretical models in terms of the concept of scientific understanding. The following chapter introduces the concept and distinguishes different approaches.

3. Scientific understanding: an overview

This chapter discusses the philosophical concept of scientific understanding. The aim of this discussion is to underline the importance of understanding[5] for the scientific endeavor by emphasizing a development that has brought understanding to the center stages of epistemology and philosophy of science. Furthermore, the concepts of explanatory understanding (EU) and objectual understanding (OU) will be differentiated. I will show that most traditional accounts of understanding are conceptualized in terms of EU which takes factive explanation as a necessary criterion. If understanding is conceptualized in such a way, it logically follows that non-explanatory models cannot afford understanding.

3.1 The history of scientific understanding

Usually, epistemology is defined as the theory of knowledge. For many years, debates about propositional knowledge ('A knows that b') have dominated the literature. Recently, however, epistemologists have shifted focus to the epistemic state of understanding and started asking about its nature, relationship to knowledge, connection with explanation and its potential status as a special type of cognitive achievement.

Catherine Elgin was one of the first to push understanding towards center stage. She claimed that knowledge alone is insufficient to make sense of many great achievements in arts and science (Elgin 1991). Instead, many great intellectual accomplishments are oriented towards a more comprehensive understanding. Even more forcefully, Jonathan Kvanvig (2003) argued that understanding is uniquely valuable from an epistemic perspective and consequently, more valuable than knowledge alone. A similar development took place in the philosophy of science. For a long time, understanding was labeled as a rather subjective and psychological notion whose importance for science is negligible. Carl Hempel's argumentation (Hempel 1965) is exemplary for that. Even though there were attempts to grant a larger scientific role to understanding (see i.e. Kim 1994), a decisive shift just occurred with Henk de Regt's distinction between phenomenological understanding and genuine understanding (de Regt and Dieks 2005). They argued that a feeling of understanding (phenomenological understanding) is neither necessary nor sufficient for genuine understanding. De Regt's work made the concept more acceptable and paved the way for much new work on the topic within the last years. Some authors now put understanding at the forefront in both epistemology and philosophy of science. For example, Potochnik (2020, 940) argues that: "I suggest that understanding, not truth, is

[5] In the following, the term understanding is simply to be taken as the short version of scientific understanding.

science's ultimate epistemic aim." Hence, the notion of understanding has gained much traction within the last years within epistemology and philosophy of science. The following section introduces different approaches to understanding.

3.2 What is scientific understanding?

Generally speaking, understanding constitutes a cognitive achievement that is ascribed to an agent (Baumgartner et al. 2017, 6). Understanding as such is a very broad topic that is highlighted by Catherine Elgin (1996, 123) who emphasizes that one may understand fields of study, particular states of affairs, institutions, other people, and many other things. This thesis concentrates on two specific uses of the term. First, the understanding of subject matters or domains of things, which is called objectual understanding (OU). Second, the understanding of why something is the case, which is called explanatory understanding (EU) (see i.e. Baumberger et al. 2017).

3.2.1 Explanatory understanding

Most traditional accounts of scientific understanding focus on explanatory understanding (EU). Within such accounts, the epistemology of understanding is strongly bound to knowledge. Pritchard (2014) terms this the "knowledge account of understanding". Understanding a phenomenon requires knowing its causes. Thus, causal knowledge is assumed to be a necessary condition for understanding. Combined with the assumption that only causal explanation can provide such knowledge, it is not surprising that understanding is often closely tied to explanation. Most traditional accounts of scientific understanding even define it in terms of explanation. Thus, most accounts of scientific understanding show the following two characteristics:[6]

1) Causal knowledge is necessary for understanding
2) Only explanation can provide this causal knowledge

However, causal knowledge is usually not taken to be sufficient for understanding. Hence, there are additional requirements enabling a transformation of knowing why into understanding why. These are often spelled out as requirements of systematicity or ability. De Regt and Dieks (2005, 150) suggest that "a phenomenon P can be understood if a theory T of P exists that is

[6] See also Verrault-Julien (2017).

intelligible (and meets the usual logical, methodological and empirical requirements)." Hence, explanation is a necessary condition for understanding. However, the ability to provide an explanation does not suffice. Instead, a further condition on theory T is employed. It needs to be intelligible. This means that scientists need to be able to "recognize qualitatively characteristic consequences of T without performing exact calculus" (ibid., 151). In other words, this condition captures the intuition that understanding requires scientists to grasp what consequences follow from T in specific situations. It requires them to grasp how predictions are generated and variables interact. The final condition of meeting the "usual logical, methodological and empirical requirements" strives to demarcate scientific theories and explanations from non-scientific ones. Thus, even though astrology may be intelligible, this condition excludes it from generating scientific understanding. In other words, de Regt and Dieks suggest that understanding a phenomenon scientifically depends on two conditions. First, it requires that there is a scientific explanation (meeting its usual requirements) and second, it makes the phenomenon epistemically accessible (framed in terms of intelligibility). These two conditions are employed in most traditional accounts of scientific understanding although the latter is usually framed in terms of grasping instead of intelligibility. Reutlinger et al. (2018) suggest an additional third criterion, which is veridicality. As discussed in the previous chapter, this criterion refers to the truth of the explanans and explanandum of the explanation. Usually, this criterion is entailed within the requirement of explanation itself. A careful differentiation is nonetheless useful as explanation might be considered a necessary criterion, however, without additionally asserting that explanation requires veridicality. Consequently, it can be stated that most accounts of scientific understanding are composed of three conditions (see also Reutlinger 2018, 1082):

(1) Explanation condition
(2) Veridicality condition
(3) Epistemic accessibility condition

Thus, many authors identify scientific understanding with grasping a correct and hence an (at least approximately) true explanation (De Regt and Baumberger 2019, 69). For instance, Strevens (2008, 3) elaborates that "an individual has scientific understanding of a phenomenon just in case they grasp a correct scientific explanation of that phenomenon. Trout (2007, 584-585) argues that "scientific understanding is the state produced, and only produced, by grasping a true explanation". The accounts given differ, however, with respect to the definition of the

success criterion of grasping. In other words, they suggest varying definitions of the epistemic accessibility condition. Stephen Grimm (2006) argues that grasping (compared to only believing) an explanation affords the scientist to answer counterfactual what-if-things-had-been-different questions. Mark Newman (2012) draws on cognitive psychology and argues that grasping differs from simply knowing an explanation in terms of the ability to draw correct inferences why the explanans explains the explanandum. He argues that scientists need to be able to develop mental models that incorporate the correct causal and logical properties of the phenomenon to be explained. Thus, similarly to De Regt and Dieks (2005), both authors define grasping in terms of the abilities that scientists derive from grasping the explanation.

However, other accounts reject the necessity of one of the three conditions. Khalifa (2012) argues that explanatory understanding simply equals knowing a correct explanation. Hence, understanding requires knowing that the explanans and the explanandum are true and knowing the correct explanatory link between them. Thus, the epistemic accessibility condition is (almost) non-existent. In his account, scientific understanding can (almost) fully be spelled out in terms of the established concept of scientific explanation. Thus, understanding as a distinct concept does not add much new to the equation.

There are also accounts that detach understanding from explanation. Lipton (2009) explores this possibility and rejects explanation as a necessary condition for understanding. He argues that understanding should rather be identified with the benefits that explanations usually provide but not with explanation itself. Therefore, he investigates whether some of these cognitive benefits which afford understanding can also be obtained via other routes than explanation. More precisely, he investigates information about causation, necessity, possibility, and unification and concludes that all of them can be acquired without explanation. Consequently, understanding does not necessarily require explanation.

3.2.2. Objectual understanding

Accounts of explanatory understanding can be distinguished from accounts of objectual understanding. Whereas the former class is focused on understanding why a phenomenon occurs, the latter class focuses on understanding holistic phenomena such as climate change. Hence, scientists seek to understand a whole system or at least a bunch of phenomena. A prominent defender of such an account is Catherine Elgin. She defines understanding as "a grasp of a comprehensive body of information that is grounded in fact, is duly responsive to evidence and enables non-trivial inference, argument and perhaps action regarding that subject the information pertains to" (Elgin 2007, 39).

Elgin's argumentation starts from the observation that science is one of humanity's greatest achievements, yet epistemology as a whole and existing conceptions of understanding (in particular) cannot do justice to its practices and achievements. The fundamental reason that she identifies lies in a wrongful overvaluation of truth in epistemology. Elgin argues that the relationship between truth and epistemic acceptability is more tenuous than usually assumed and therefore, truth-centered epistemology is wrong-headed. Instead, she argues for a "generous, flexible conception of understanding" (Elgin 2007, 33). Even though science provides understanding of the natural order, it relies heavily on laws, models, idealizations, and approximations, which are not true. Nevertheless, these are fundamental to the insights that science delivers. Thus, models help to understand reality in ways that were impossible if the tools of science were restricted to the unvarnished truth. Consequently, epistemology cannot accommodate scientific practice if models are evaluated purely with respect to the accuracy of their representation of nature. To the contrary, simplifications, idealizations and deliberate distortions are not shortcomings. They are felicitous falsehoods, which help to exemplify features that were otherwise overshadowed. In this sense, models' falsehoods are epistemically fruitful by allowing epistemic access to aspects that were otherwise inaccessible. Hence, Elgin (2004, 120) defines understanding as a non-factive cognitive success term. However, Elgin's account is not indifferent to truth. Instead, she suggests that understanding needs to be true enough. Interestingly, the epistemic acceptability of falsehoods is not necessary evaluated with respect to their 'closeness to truth'. Instead, their function dictates acceptability. For instance, there are cases, in which approximations can be accepted by reference to a negligible deviation from truth. However, in other cases, approximations may perform different functions and their deviations from truth is not a shortcoming but beneficiary.

Consequently, Elgin suggests taking felicitous falsehoods not as inaccurate statements, but as fictions. Therefore, understanding advances from fiction-based reasoning. She proposes that the gap between fiction and real-world can be bridged with the concept of exemplification. Felicitous falsehoods exemplify a feature in a specific context and the exemplification of that features contributes to understanding. Thus, exemplification affords epistemic access to a certain feature or pattern. Consequently, models and theories cannot be portrayed as mirrors of nature. Yet exactly this allows scientist to uncover systematic relationships that a direct examination of facts would not enable.

Hence, Elgin's account of objectual understanding differs significantly from traditional accounts that focus on explanatory understanding. They differ with respect to the object that is

understood. However, they also differ with respect to the role that idealizations and distortions play in science.

3.3 Summary and outlook

Given the different notions of understanding, a careful differentiation between explanatory understanding and objectual understanding is necessary. Elgin's account of objectual understanding is more flexible in accommodating practices of theoretical modelling. The traditional accounts of explanatory understanding assume in turn a tight link between explanation and scientific understanding. A true explanation is regarded as necessary criterion. The argumentation in chapter two showed that there is good reason to reject (or at least question) the explanatory capabilities of many theoretical economic models. Given the three conditions that traditional accounts share – explanation, veridicality, and epistemic accessibility – it remains doubtful whether theoretical models in economics could afford understanding. For that they would need to provide a correct explanation entailing true explanans and explanandum. In other words, if a traditional account of understanding (EU) is adopted, explanation and understanding are bound together tightly blocking the road to define the epistemic benefit of theoretical models in terms of the various affordances of understanding.

The following chapter questions this tight link. It investigates the kinds of phenomena that theoretical models target and their role in investigating them. Comparably to the differentiation between OU and EU, I will argue that a closer look at the kind of object that is to be understood is necessary. Given that these phenomena are often of general kind, traditional accounts of understanding are unsuited to make sense of theoretical modelling in economics.

4. Why traditional accounts of understanding are unsuited for theoretical modelling

As laid out in the previous chapter, traditional accounts of explanatory understanding share three fundamental conditions: explanation, veridicality, and epistemic accessibility. In this chapter, I will present three arguments why traditional criteria are unsuited for theoretical modelling in economics. First, I argue that the conditions of factive explanation and epistemic accessibility can work against each another. In other words, accurate representation can impede epistemic accessibility. The second and main argument of this chapter examines the kind of phenomena that theoretical models in economics target. I argue that these are general phenomena and not individual real-world occurrences. Thus, the object of understanding is often a general phenomenon. This impacts an appropriate account of understanding and questions the necessary criteria posed by traditional accounts. The third argument underlines the purpose of theoretical modelling in investigating general phenomena. Many theoretical models do not aim to provide actual explanations of specific facts. Instead, theoretical models are often built for exploratory purposes. Their purpose consists in suggesting possible explanations.

Considering these arguments, I conclude that the kind of understanding that theoretical models can afford is rather objectual (OU) than explanatory (EU). Consequently, traditional accounts framing understanding in terms of EU appear not to make sense of theoretical modelling in economics.

4.1 A trade-off between explanation and epistemic accessibility

Most traditional accounts of scientific understanding are composed of an explanation condition and a condition of epistemic accessibility. Consequently, one may start with a simplified working definition of understanding as follows (Reutlinger et al. 2018, 1084): scientist S understands phenomenon P via model M if and only if M explains P and S grasps M.

As elaborated in the previous chapter, different authors propose varying definitions of what it means to grasp an explanation or a model. Grasping could be characterized by knowledge of why the explanans explains the explanandum, the scientist's ability to answer what-if-things-had-been-different questions or the scientist's ability "recognize qualitatively characteristic consequences (…) without performing exact calculus" (De Regt and Dieks 2005, 151). Independently of the exact definition of grasping, and the necessary abilities of a scientist that come with it, there are severe trade-offs between conditions of veridical model-based explanation and epistemic accessibility. In other words, the conditions of explanation and epistemic accessibility can work against each other. This is particularly the case when models

are used to study highly complex phenomena as economics does. The following example substantiates this point.[7]

In 2005, sociologist Peter Hedström (2005) developed a model of unemployment for the metropolitan area of Stockholm during 1993 and 1997. It is an agent-based model, which is empirically calibrated. The model consists of 87,924 agents. Different characteristics are assigned to these agents, which include age, marital status, previous unemployment experiences, immigration background and many others. The distribution of these characteristics is supported by demographic data for twenty- to twenty-four-year-old residents in the Stockholm area in the given period. Hedström's model assigns probabilities of leaving unemployment (becoming employed) to the agents based on three sets of factors. First, their own social and demographic characteristics. Second, the level of unemployment among their neighbors. Third, the tightness of the labor market. Given that Hedström uses empirically calibrated data and assigns many empirical characteristics to the agents, I suggest the model intends to represent a particular empirical environment. In other words, the model has a lower degree of idealization than most theoretical models in economics. However, its lack of simplicity makes it much more difficult to grasp it. Hence, even though the model may fulfil the criteria for providing factive explanation, complexity diminishes its capacity to provide epistemic access to fundamental difference makers. Taking de Regt and Dieks definition of grasping as the scientist's ability "recognize qualitatively characteristic consequences (…) without performing exact calculus", it becomes apparent that increasing the complexity of models tends to impede the ability of scientists to draw qualitative consequences or make qualitative predictions for alternative scenarios. Therefore, when discussing the capacities of toy models, Reutlinger et al. (2018) argue that "the gain in complexity has an interesting effect: it diminishes the capacity of these models to provide understanding because it is mainly the simplicity of toy models that permits scientists to grasp them" (2018, 1092).

Consequently, it is worth emphasizing that modelling can serve different purposes ranging from explaining and understanding why a particular phenomenon has occurred to affording a more general understanding of (specific features of) a target phenomenon. Therefore, it requires examining what kind of phenomenon a model targets and what purpose a model serves in investigating it. In other words, this example demonstrates that the kind of understanding an empirically calibrated model can afford is categorically different from a simplified toy model. As shown in the differentiation between objectual and explanatory understanding, the definition

[7] The example is taken from Reutlinger (2018, 1092).

of understanding itself should be dependent on what exactly is to be understood. The following section takes a closer look at this finding.

4.2 Kinds of phenomena – theoretical models investigate general phenomena

To make sense of theoretical models in economics, it is important to highlight what exactly these models aim to explain and understand. In many cases, theoretical models target stylized facts. Stylized facts are "broad generalizations, true in essence, though perhaps not in detail" (Bannock et al., 1998, 396-397). For instance, some models of economic of growth try to explain the (stylized) fact that the profit rate is constant (Elgin 2004, 118)[8]. However, due to a variety of (non-)economic factors, profit rates are usually not constant. At most, they *tend* to be. Consequently, the stylized fact of constant profit rates is strictly speaking not a fact. Nevertheless, it would be wrong to conclude that economics aims to explain known falsehoods. Instead, it is often concerned with approximately true facts. Consequently, theoretical models usually do not aim to contribute to explaining specific occurrences. They are not concerned with individual instances of a phenomenon. The target of a theoretical model is often more general and described on a higher level of abstraction. The following three examples underline this point.

Akerlof's "market for lemons" (1970) is one of the most well-known theoretical models in economics. His famous model of the market for used cars aims to find a rationale for sharp prices differences between new and (only slightly) used cars. Before Akerlof's work, economists usually appealed to consumer preferences for strictly new cars to explain this difference. His paper presents a new perspective by appealing to asymmetric information between consumers and car dealers. Thus, on a broader scale, he aims to understand how asymmetric information tends to result in "a reduction in the average quality of goods and also in the size of the market" (ibid., 488). Hence, Akerlof aims to understand a more general phenomenon of adverse selection in markets with asymmetric information and how market failures could result from this. The model he develops is highly simplified. It is concerned with the automobile market. However, this market is not chosen because of its importance or realism. It is chosen for its concreteness and ease in understanding (ibid., 488). Akerlof assumes that there are only four kinds of cars – new and old and good and bad. He explicitly highlights that the model's simplifications are done "for the sake of clarity" and that he is not concerned with representing a realistic car market (ibid., 489).

[8] "The profit rate refers to the level of profits in the economy relative to the value of the capital stock" (Bannock et al. 1998, 397).

His procedure can be characterized as follows. There is a general phenomenon of reduction of product quality and market size in certain markets with asymmetric information. The aim of Akerlof's investigation is to understand the causal driver behind that phenomenon. He suggests a possible mechanism - a possible explanans: adverse selection. The model is then used to instantiate that potential explanation. Akerlof chooses the market for used cars as it exemplifies the problem well and allows for epistemic access. Therefore, the model is highly simplified and does not even aim to represent a specific automobile market accurately. Thus, Akerlof investigates how market failure could be brought about by a mechanism of adverse selection in markets characterized by asymmetric information. Most importantly though, Akerlof does not investigate a specific instance of adverse selection in a specific market. He explores the general phenomenon and instantiates it in a hypothetical car market.

Abhijit Banerjee's "simple model of herd behavior" (Banerjee 1992) is a similar case. Banerjee investigates the general phenomenon of "herd behavior – everyone doing what everyone else is doing, even when their private information suggests doing something quite different" (ibid., 798). Hence, the finding that motivates Banerjee's model is the stylized fact that people tend to use information contained in the decisions and actions of other people. Thereby, they become less responsive to their private information. Interestingly, one consequence of this pattern is that people's own actions become less informative to others which leads to a potential reduction of actions' informativeness in equilibrium. This results in a decrease in welfare which could be dampened by restricting people to use their own information only. Banerjee lists several phenomena where such mechanism could be decisive. These range from fertility choices (i.e. whether to have children or not), voters being influenced by opinion polls (voting according to polls' predictions) or academic researchers choosing to work on hot topics. Banerjee invites the reader to see such occurrences of herd behavior as instances of imperfect information in which agents have limited private information and therefore aim to infer information from the actions of others. Hence, comparably to Akerlof, Banerjee observes a general phenomenon and tries to find a rationale behind it. However, he does not investigate an individual occurrence of this phenomenon. Nevertheless, Banerjee's model is concerned with an example that could occur in the real world. He describes a choice between two restaurants based on how crowded they are. In the model, agents face a choice between restaurants A and B, which are unknown to them. Banerjee assumes a 51 percent ex-ante probability that restaurant A is better than restaurant B. Yet, 99 out of 100 agents get an additional signal that B is better. However, the agent who chooses first had a signal for A and consequently chooses restaurant A. The second mover then observes the first agent's behavior.

Even though the second agent has a private signal preferring B, the agent will know that the first agent's signal favored A. As both signals are of equal quality, they cancel out and the second agent chooses regardless of its own signal. Consequently, the second agent decides only based on ex-ante probabilities and chooses restaurant A. In the end, all agents will have chosen restaurant A even though 99 of them had a private signal for B.

Banerjee's model is highly simplified, and many decision factors are neglected. It does not intend to accurately capture the complexity of a specific instances of herd behavior in the real world. Banerjee investigates a general phenomenon by modelling a potential instantiation in the form of a sequential restaurant choice. Hence, similarly to Akerlof, one important reason for building the model is exemplification. This allows for epistemic access to the mechanisms behind herd behavior.

The third example is Schelling's checkerboard model of spatial segregation (Schelling 1978). The model provides an example how individual micro-scale behavior can have large-scale effects. It shows how minor individual preferences against being in a minority can nevertheless lead to stable large-scale pattern of segregation. The model is structured as randomly distributed agents of two kinds (A and B) on a checkerboard. These agents can either stay put or move without costs. If agents prefer not being in a minority in their neighborhood and change their location if this preference is not satisfied, it can be shown that a stable pattern of segregation emerges. To highlight the kind of phenomenon that Schelling investigates, it is important emphasize that the checkerboard model is not based on any particular city and does not aspire to represent real-world scenarios (Aydinonat 2007). To the contrary, the idea of using a checkerboard emerged from Schelling's need to find an intelligible way to represent an abstract phenomenon to his students (Elliot-Graves 2022). Again, comparable to the first two examples, the phenomenon that is investigated is abstract and of general nature. The model does not aspire to represent a specific real-world instance of segregation with its empirical details.

These three examples show that theoretical models in economics are often concerned with phenomena described on a higher level of abstraction. Thus, understanding the role of theoretical models requires discussing the nature of their target systems.

4.2.1 General and generic models

In many cases, the target system of a model is a specific empirical occurrence. Such models target individual observations or individual phenomena. However, the construction of a single model to study one specific target cannot account for all kinds of modelling practices.

Weisberg's account (2013) provides a good starting point to define further kinds of modelling practices as he also addresses different kinds of target systems.

In comparison to the model-based study of individual phenomena, Weisberg observes that theoretical research often involves exploring models that aim to understand classes of phenomena. Therefore, he differentiates between target-directed modeling and modelling without a specific target. These two categories can be differentiated with reference to the level of abstraction of their targets. Weisberg subdivides modelling without a specific target further into generalized modelling, hypothetical modelling, and targetless modelling. Generalized models are used to study phenomena such as parasitism or sexual reproduction in general. Hypothetical modelling refers to the construction of models with nonexistent targets. The last type refers to models with no specific target at all. In this case, the model itself is the only object of study without aspiration to learn anything specific about real-world systems. These three types of modelling share the feature of investigating classes of phenomena, however, they do not study individual occurrences (ibid., 114).

Theoretical models in economics are usually not target-directed. In Weisberg's terms, they fall in the category of modelling without a specific target. Akerlof's "market for lemons", Banerjee's "herd behavior" and Schelling's checkerboard model fall in this category as well. All these models are used to study phenomena in general (such as segregation or herd behavior). Weisberg defines the target of a generalized model as an abstraction over the individual instances of this phenomenon (ibid., 116). He notes two different possibilities how this target is abstracted. First, an *intersection view* in which features that specific occurrences share are aggregated to a general phenomenon. On this view, general phenomena are composed only of features that all its specific occurrences have in common. For instance, modelling the phenomenon of "sex in general" requires finding the features shared by all (or at least many) instances of sexual reproduction. In other words, the target of a generalized model to study sexual reproduction is composed only of properties that sexually reproducing species share. Second, a *construal view* in which the generalized model contains features that go beyond what is shared by its specific instances. Hence, the abstracted target contains concrete features that its specific occurrences do not exhibit. In other words, the modeler needs to make additional assumptions. Weisberg terms this a construal setting.

Within Weisberg's dimensions, theoretical models in economics usually belong to the second class. All three model examples construe a setting, and the models have concrete feature that specific instances of their target phenomenon do not have. In other words, additional assumptions, apart from aggregated features that specific instances of phenomenon share, are

needed. However, there is another interesting point that needs to be highlighted. Reconsidering the earlier introduced stylized fact of constant profit rates, it is important to note that a generalized model of an economy with constant profit rates is not simply an abstraction of specific instances of economies that share this feature. Instead, it is an abstraction of a tendency that individual economies share. Thus, the feature, or features, that subsumes specific instances of a phenomenon into a general one, need not even occur in all, some, or any of them. This makes theoretical models in economics often a particular peculiar case. Indeed, this might question whether Akerlof's, Banerjee's and Schelling's model are actually prime examples of generalized models in the way Weisberg presents them.

Weisberg uses the example of the evolution of sexual reproduction in biology (ibid., 115ff.), where generalized models are used to provide insights into the evolutionary advantage of sexual reproduction compared to asexual reproduction, without necessarily focusing on the specific manifestations of sexual reproduction in particular populations. Thus, the aim of studying a phenomenon in a general manner is not to uncover specific facts about specific populations, but rather to gain a broader understanding of the phenomenon itself and its larger implications. He argues that although there is no such thing as 'sex in general' in the world, as it only exists as a collection of specific instances of sex in different populations, there are still certain properties that all sexually reproducing species share. These common properties are considered the fundamental causal features of the general phenomenon of sexual reproduction. Therefore, by identifying the commonalities between the different instances of sex, it is possible to create a framework that the general model can represent. Essentially, 'sex in general' is an abstraction that encompasses these more specific forms of sex and consequently, it should apply to all these forms of sex. Thus, the model is applicable (or at least could be applicable) to many real-world target systems.

Even though much of this also applies to the economic models discussed in this chapter, there are significant differences as well. Take Schelling as an example as it is the most extreme case. The model is highly abstract and contains significant idealizations. For instance, it makes a very specific assumptions with respect to the structure of the neighborhood. Additionally, it assumes homogenous preferences. It ignores strong discriminatory preferences and neglects other well-known factors of segregation, such as economic conditions, welfare differences among groups and structural and organized discrimination. It also ignores psychological and sociological factors or costs of moving. Yet this would not necessarily exclude it from counting as a generalized model in the way Weisberg describes them. However, it is important to note that the base of Schelling's model is not an abstracted intersection of common properties that

different instances of segregation share. It is instead an abstract hypothetical system without direct reference to real-world cities: "I decided if I am going to teach my students, I'll have to make it all up. I wondered what to do, and decided maybe now was the time to begin playing around (…)." (Schelling, in Aydinonat 2005, 4).

Hence, Schelling does not aggregate features that individual instances of segregation share. He presents a hypothetical setting of segregation. Although both approaches are often summarized under the umbrella term of generality, these two conceptions should be distinguished more carefully. Elliot-Graves (2022) suggests differentiating these approaches into *general models* and *generic models*. On Elliot-Graves's account, the basis of generality is the identification of common properties of disparate, particular systems. Subsequent abstraction of differing factors is the fundamental pre-requisite to achieve generalization. This is close to Weisberg's account. Genericness, on the other hand, is "the absence of particularity" (ibid., 74). Thus, general and generic models differ in terms of what they are about. Generic models are not about (any) particular real-world system(s). Thus, an important distinguishing feature between general and generic models is their target system. While general models have many concrete, local, real-world target systems, the target systems of generic models are generic themselves. Hence, these target systems are hypothetical and generic modelling has much in common with investigating hypothetical systems.

However, models are not general or generic by nature. Elliot-Graves bases her differentiation on models' use, on how they should be used respectively. Thus, models are not general or generic by themselves but are either used generally or generically. General and generic models also differ in terms of their source of value. Elliot-Graves argues that generality is valuable as it enables scientists to identify particular phenomena as manifestations of general patterns. Hence, the behavior of particular phenomena can be compared to similar systems. Thus, generality refers to the ability of a model to apply to a wide range of systems or phenomena. It can be applied to many different cases. A prime example of a general model is the logistic growth model in ecology, which is used to explain and predict the size of actual populations (see i.e. Barlow 1992). This model applies to many diverse actual systems in the world, and its source of value is that it provides information about a general phenomenon in the world.

In contrast, generically used models do not apply to any particular real-world systems. Hence, the source of their value is different. Genericness refers to the ability of a model to capture the core features or mechanisms of a system in a simplified or abstract way. Thus, generic models capture essential features of a system regardless of its instances. Thereby, generic models enable scientists to explore possibilities and discover how phenomena could be

brought about without the restrictions of particular real-world scenarios. They can be used to map spaces of possibilities. Schelling's model of residential segregation is a good example. It helps to understand the underlying dynamics of a phenomenon rather than its specific manifestation in each case. While it offers a potential explanation of how segregation could occur, it does not provide information on how it occurs in real cities. The target system of Schelling's model is generic itself. Within Elliot-Graves' terminology, I suggest that Akerlof's "market for lemons" and Banerjee's "herd behavior" also fall in the category of generic models. All three examples have rather generic targets, and the models are not intended to be applied to particular real-world situations. Instead, they help to investigate and exemplify fundamental underlying dynamics for a type of phenomenon.

Hence, even though both general and generic modelling practices investigate phenomena in a general manner, there are important differences. In the following, I will refrain from using Elliot-Graves's exact terminology and instead refer to both kinds as general modelling and the investigation of general phenomena and instead differentiate them (if needed) by the terms G1 models/phenomena for Elliot-Graves' *general models* and G2 models/phenomena for her *generic models*.

The fundamental takeaway of this discussion is that many theoretical models in economics investigate phenomena *in general*. Like the three examples discussed in this chapter, most theoretical models in economics fall in the G2-category. Qua nature of the models' targets, the fundamental goal of this modelling practice is to uncover underlying dynamics and potential causal drivers for classes of phenomena. The next section underlines this explorative role.

4.3 How-possibly explanations in economics

The previous section laid focus on the object of investigation. This part highlights the nature of the explanans. In many cases, general models can be used to explore potential explanations for a phenomenon. This approach is not new. Many commentators in the literature have suggested that highly idealized theoretical models could have epistemic value by providing how-possibly explanations (Verreault-Julien 2019). Before unfolding this point in detail, it is worth reconsidering that explanations can be either *complete* or *partial*. This means that an explanation can either be complete including all the causes that make a difference to the explanandum and no others. An explanation may also be partial including some of the relevant causes. Furthermore, an explanation can be *actual* meaning that the explanans satisfies the truth condition or *potential* when it is unknown whether that is the case (Marchionni 2017). An explanation can display any combination of these two. Theoretical models in economics usually

provide *partial* and *potential* explanations.[9] For instance, Akerlof's model can answer a question like: how is it possible that asymmetric information causes market failure? Banerjee's model of herd behavior could answer a question such as: how is it possible that we often observe people choosing what other people chose? Hence, this class of models does not intend to provide *actual* explanations for individual phenomena, but *potential* explanations for general ones. Schelling's checkerboard is another paradigmatic case. The fundamental finding of the model consists in showing that it is possible for segregation to emerge without strong discriminatory preference. Consequently, commentators have suggested that Schelling's model does not provide an actual explanation (how-actually explanation) but only a possible one (how-possibly explanation) (Ylikoski and Aydinonat 2014). Thus, one could interpret Schelling's model as providing a *partial* and *potential* explanation for segregation. This potential explanation can also be reformulated in terms of necessity. Strong discriminatory preferences are no necessary criterion for segregation. In other words, racial preferences are not among the necessary causes of segregation. On a different interpretation, one may also use Schelling's results to reject an impossibility claim. Before his work, it may have been thought impossible for residential segregation to emerge without strong racial preferences. The model rejects this impossibility. In sum, Schelling's model provides modal information about a general phenomenon.

An important consequence of this finding is that the explanatory power of a model depends on the kind of explanandum it is supposed to explain. If Schelling's model is supposed to explain segregation in general, the proposed mechanism is among several possible difference-makers. If the model aims to explain a specific instance of residential segregation in a specific neighborhood, this mechanism could be a possible difference-maker, but it need not be. Evidence might simply rule out this possibility.

4.4. Summary and implications

With reference to actual modelling practices in economics, this chapter presented three arguments why traditional accounts of scientific understanding are unsuited for theoretical modelling in economics. The first argument focused on the fact that the requirements for a model to provide factive explanation and the consequent increase of its complexity can impede epistemic accessibility. Thus, there can be trade-off between the necessary conditions of explanation and epistemic accessibility for scientific understanding as presented by traditional

[9] These kinds of explanation are addressed by different terms in the literature such as „potential explanation", „possible explanation" or how-possibly explanation" and these are used interchangeably here.

accounts. The second argument highlighted that theoretical models in economics usually target classes of phenomena and not individual real-world occurrences. An interesting implication of this finding is an inappropriate high focus on the question how a single theoretical model can faithfully represent individual instances in the real-world. This focus is misguided as most theoretical models do not aim to faithfully represent a singular occurrence, but rather intend to explore general phenomena. According to the third argument the purpose of many theoretical models is explorative. Their value consists in proposing how-possibly explanations for general phenomena.

Considering these arguments, I contest that traditional accounts of scientific understanding are well suited to make sense of theoretical modelling in economics. The reason lies in the underappreciated diagnosis that characterizing scientific understanding requires considering the object, which is to be understood. At the beginning of chapter three, a differentiation between objectual understanding (OU) and explanatory understanding (EU) was introduced. In reference to this differentiation, I suggest the purpose of theoretical models rarely lies in affording direct understanding to why a specific individual phenomenon has occurred. The kind of understanding such models can generate is rather objectual. An appropriate characterization of scientific understanding able to accommodate the differentiation of different types of phenomena. In other words, the constitutive criteria of scientific understanding are dependent on the kind of object that is to be understood. Thus, if the object is a specific individual occurrence, defining understanding in terms of a factive causal explanation and a component of grasping is legitimate. However, understanding phenomena like herd behavior or segregation in general should be characterized by different criteria. In these cases, veridical explanation does not need to be set as necessary criterion. The next chapter substantiates this claim and shows that true causal explanation should not be set as a necessary requirement.

5. Understanding with theoretical models

In the previous chapter, I argued that theoretical models in economics often investigate general phenomena. The purpose of such models often lies in proposing potential explanations for these phenomena. Furthermore, I proposed that characterizing scientific understanding needs to consider the kind of object that is to be understood. With reference to the differentiation between EU and OU, theoretical models in economics can rather afford understanding of the latter kind. Therefore, factive causal explanation does not need to be among the necessary criteria of understanding.

This chapter substantiates these claims. Following Verreault-Julien (2019b), a case study of paradigmatic cases of understanding without how-actually explanation will be presented. Based on this case study, I reject the necessity of how-actually causal explanation (HAE) for scientific understanding. Instead, I propose a broader account of scientific understanding that can accommodate for cases of understanding without HAE. Furthermore, I will refine Verreault-Julien's argumentation by proposing to differentiate the necessary conditions of understanding based on the kind of phenomenon that is to be understood.

5.1 True causal explanation is not necessary for scientific understanding

As shown in the previous chapters, the traditional epistemology of understanding is strongly bound to causal explanation. It is usually argued that scientific understanding requires knowledge of causes. This knowledge can only be provided by true causal explanation. Verreault-Julien (2019b) argues that these two tenets characterize what he terms the "narrow knowledge account of understanding" (ibid., 1). Consequently, denying the necessity of explanation requires rejecting at least one of the two tenets. The following argumentation focuses on the second tenet. I will argue that knowledge of causes can be supplied by other means than true causal explanation. These other means refer to how-possibly explanations.

Verreault-Julien (2019b) names two important desiderata that an account of scientific understanding should fulfil. First, the account must provide reliable criteria to demarcate genuine from illusory understanding. It needs to consider that a sense of understanding can be a deceptive guide to genuine understanding. In other words, it needs to consider the actual possibility of misunderstanding. Second, scientific understanding should not exclude large parts of scientific practice from promoting understanding of the phenomena that are investigated: "a naturalistic outlook on science should compel philosophers to not attribute systematic and persistent error across different fields of science" (ibid., 2). Hence, if there are practices that are

prevalent across all (or at least many) areas of science, it appears suspicious to label them as inappropriate for generating understanding.

With respect to theoretical modelling, traditional accounts of explanatory understanding fare well on the first desideratum but score poorly on the second. Given that theoretical models do usually not support actual explanations, traditional accounts of EU must assess them as non-conducive to affording understanding of its targets. However, this would attribute a "persistent and systematic error" across all scientific disciplines where theoretical modelling is a prevalent practice. This should cast doubt on the constitutive criteria of understanding. In other words, the narrow knowledge account of understanding is based on conditions whose implications make its premises questionable.

Consequently, the fundamental strategy in rejecting explanation as necessary criterion for understanding consists in finding cases of genuine understanding without how-actually causal explanation (HAE). Thus, it requires finding non-explanatory sets of propositions that nevertheless convey the required information for understanding. In a similar manner, Lipton (2009) highlighted that the cognitive benefits that are typically generated through explanation, can also be acquired in other non-explanatory ways. Thus, the mistake, that traditional accounts of understanding commit, lies in identifying understanding with explanation itself, rather than with the benefits it provides. Yet, separating explanation from the cognitive benefits it provides is imperative.

According to Strevens (2013), an explanation is a set of propositions which establishes a connection between the explanans and the explanandum in the correct way. To differentiate how-actually explanations (HAE) from how-possibly explanations (HPE), one may state that while the constitutive elements of the explanans and explanandum are necessarily true in a HAE, this is not the case for an HPE. Put differently, there is no sufficient justification for the believe that an HPE is an HAE. Strevens partitions the elements of an explanation into internal and external conditions. The former refers to an explanation's logical structure, whereas the latter defines its empirical correspondence to the world. Based on this distinction, HAEs and HPEs can be demarcated as follows. While HAEs satisfy both internal and external conditions of explanation, HPEs only satisfy the internal conditions. In other words, it is simply unknown whether an HPE actually explains the explanandum. All that is known is that it could. Craver (2006, 361) defines HPEs as a description "how a set of parts and activities might be organized such they exhibit the explanandum phenomenon". HPEs show how a phenomenon could have occurred or how existing processes could lead to other outcomes. Thus, while explanantia and

explananda of an HAE are known to be true, HPEs consist of explanantia and explananda, which are either false or not known to be true.[10]

Considering this distinction, the following cases aim to demonstrate that understanding may not only be enabled by actual explanations, but also by potential ones. Hence, the cognitive benefits that afford understanding can also be supplied by how-possibly explanations (HPE).

5.1.2. Exemplary cases of understanding without HAE

Verreault-Julien (2019b) discusses two examples of theoretical models that afford understanding without supporting veridical explanation. These are Schelling's checkerboard model (Schelling 1978) and the Hawk-Dove model (i.e. Maynard Smith & Price 1973 and Maynard Smith & Parker 1976).

Schelling's checkerboard has been very influential across the social sciences. It is typically interpreted as providing an HPE. It targets the general phenomenon of segregation and identifies a potential mechanism that could explain it. More precisely, it tries to find an answer to the question of "how is it possible for segregation to emerge in a city without collective preferences for segregation?" (Weisberg 2013, 118-119). Schelling's model can show that a stable pattern of segregation can emerge even though individuals only have preferences for not living in a minority. Thereby, it refuted the prior believe that strong discriminatory preferences (i.e. racism) are among the necessary causes of segregation. This finding proved to be robust across various changes of assumptions (Muldoon et al. 2012). Consequently, many commentators concluded that Schelling's model can teach relevant findings about the real-world (see i.e. Sugden 2000). At first sight, this may seem paradoxical as Schelling refrains from making claims about specific instances of (residential) segregation and it is unknown whether the proposed mechanism is actually among the difference makers in particular real-world instances of segregation. Nevertheless, Verreault-Julien (2019b, 18) argues that the model appears to provide causal knowledge about the phenomenon: "using the model, we know that if the mechanism were true, under suitable conditions residential segregation could be brought about. We know that it could actually depend on those factors or, conversely, that it does not necessarily depend on strong discriminatory preferences". In other words, knowledge that a phenomenon could (or could not) be brought about by some causal factor, and knowledge of the conditions that are required for this to take place, is highly relevant and enhances a scientist's understanding of the phenomenon.

[10] However, one needs to bear in mind that the differentiation between HAEs and HPEs is rather a continuum. Whether an HPE is an HAE is both a question of evidence and degree of confirmation.

The Hawk-Dove model (Maynard Smith & Price 1973 and Maynard Smith & Parker 1976) is an evolutionary game theoretical model at the intersection of biology and economics. It depicts the fundamental conflict between prosocial behavior (i.e. cooperation) and antisocial behavior (selfishness) in situations of animal conflict. In essence, the model shows that individual selection processes can account for restraint from contest between individuals of the same species for limited, shareable resources. Prior to the model, it was believed that strategies of 'limited war' between animals can only be explained by a mechanism of group selection. Yet, Maynard Smith and his co-authors were able to show that such behavior can also be explained with reference to a process of individual selection: "a main reason for using computer simulation was to test whether it is possible even in theory for individual selection to account for limited war behavior" (Maynard Smith & Price 1973, 15). However, this does not mean that their model supports a HAE. It only shows that it is possible that restraint in animal combat could also be explained a process of individual selection. Nevertheless, their work is judged enormously influential by fellow practitioners and philosophers. Rohwer and Rice (2013) argue that the Hawk-Dove enhances our understanding as it justifies the true belief that a known real-world phenomenon, restraint from combat in situations of conflict over resources, cannot only be explained a mechanism of group selection but also by a mechanism of individual selection. This belief has tremendous relevance with respect to understanding why this phenomenon occurs.

With reference to these two paradigmatic cases, Verreault-Julien (2019b) argues that there are cases of understanding without actual explanation (HAE). Theoretical models may advance understanding through HPEs. The fundamental basis of this argument is the judgement of practicians - and philosophers alike – that classify these models as advancements of our understanding in their respective field. However, he remains rather short on what exactly is really understood. He argues from the rightful impression that many theoretical models have improved our understanding of their target phenomenon. Yet, he misses to define these phenomena in more detail. In both paradigmatic cases, he acknowledges that the models do not support understanding of specific real-world instances of residential segregation or animal competition. Instead, although only expressed implicitly, it is understanding of these phenomena in a more general manner. However, this is the decisive differentiation.

Evaluating the debate about the necessity of HAE for scientific understanding, arguments can be divided into two fundamental camps. Those upholding the necessity of HAE for understanding argue that it appears suspicious that theoretical models can afford understanding without truthfully depicting the empirical circumstances of the explanandum phenomenon.

Objectors point to exemplary cases trying to prove that many models have improved our understanding even without providing (or supporting) actual explanations. Yet, what is missing is an emphasize on the nature and kind of the phenomenon that is understood. Therefore, the fundamental argument of this thesis is the following: if the object to be understood is a specific individual occurrence, defining scientific understanding in terms of a correct (causal) explanation and a component of grasping is legitimate. Therefore, understanding residential segregation in a specific neighborhood in Chicago between 1993-1997 requires knowing a true causal explanation and grasping it. If, however, it is a general phenomenon, this characterization is inappropriate. Understanding phenomena like residential segregation in general does not necessarily require true causal explanation. It should be characterized by different criteria. Then, understanding should be defined by modal knowledge about counterfactual dependences, which can be supplied by HPEs as well. The following part suggests a broader account of scientific understanding that can accommodate this differentiation.

5.2. A broader account of scientific understanding

A good starting point to broaden traditional accounts of understanding is Woodward's definition of explanation (Woodward 2003). According to Woodward, "an explanation ought to be such that it enables to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various ways" (ibid., 11). Thus, explaining is essentially a matter of conveying information that display the dependence of the explanandum on the explanans, or in Woodward's words, "a matter of exhibiting systematic patterns of counterfactual dependence" (ibid., 191). Thereby, explanations enable to answer what-if-things-had-been-different questions. With reference to the Woodwardian counterfactual theory of explanation, Reutlinger (2016) attempts to provide a monist account that binds together causal and non-causal kinds of explanation. He argues that explanations, causal or non-causal, share a feature that allows to answer what-if-things-had-been-different questions. Therefore, he suggests that, for a relationship between the explanans and the explanandum to be explanatory, it must meet the following conditions (ibid., 737):

1. **Veridicality condition:** Generalizations $G_1, \ldots, G_m$, the auxiliary statements $S_1, \ldots, S_n$ and the explanandum statement E must be (approximately) true.
2. **Implication condition**: $G_1, \ldots, G_m$ and $S_1, \ldots, S_n$ logically entail E or a conditional probability $P(E|S_1, \ldots, S_n)$

3. **Dependency condition:** G_1, …Gm supports a least one counterfactual between S_1, …, S_n and E, meaning that had S_1, …, Sn been different than they actually are (in at least one way deemed possible in the light of the generalizations), then E or the conditional probability on E would have been different as well.

Hence, Reutlinger suggests that both the explanans and explanandum need to be true, the explanandum needs to be logically entailed in the explanans and finally, an explanation must answer to (at least one) what-if-things-had-been-different question. Woodward places high importance on an explanation's feature to provide systematic pattern of counterfactual dependence and the subsequent ability to answer what-if-things-had-been-different question. It is exactly this capacity of explanation (HAE) that affords understanding. This makes room to reformulate Reutlinger's conditions in a way that allows for non-explanatory propositions that fulfil the implication and dependence condition. Hence, it allows to include HPEs that convey information about counterfactual dependence. Thus, if we accept that HPEs can afford understanding, even if they contain (potentially) false explanantia and/or explananda, one may revise Reutlinger's veridicality condition and replace it by a condition of possibility. Therefore, Verreault-Julien (2019b, 18) suggests that the relationship between an explanans and an explanandum can afford understanding iff.:

1. **Possibility condition:** Generalizations G_1, …, G_m, the auxiliary statements S_1, …, S_n or the explanandum statement E are (im)possible according to the relevant modal interpretation and epistemic goal.
2. **Implication condition**: G_1, …, G_m and S_1, …, S_n logically entail E or a conditional probability $P(E|S_1, \ldots, S_n)$
3. **Dependency condition:** G_1, …Gm supports a least one counterfactual between S_1, …, S_n and E, meaning that had S_1, …, Sn been different than they actually are (in at least one way deemed possible in the light of the generalizations), then E or the conditional probability on E would have been different as well.

With this proposal, Verreault-Julien argues that while an explanation (HAE) requires an actual explanans and explanandum, understanding does not. Hence, this proposal broadens the explanation condition of traditional accounts of understanding by allowing non-explanatory propositions if these provide information of counterfactual dependence and enable scientists to answer to what-if-things-had-been-different questions. For example, Schelling's checkerboard

model proposes a possible causal driver that could cause residential segregation. However, as it is unknown whether this factor is responsible for actual real-world instances, Schelling's model is not explanatory (HAE). Nevertheless, it can afford understanding by supporting a potential mechanism and the ability to make counterfactual inferences.

Three important qualifications are needed. The first relates to Verreault-Julien's possibility clause in relevance to the "modal interpretation and epistemic goal". With regards to this clause, it must be clear what kind of possibility an HPE establishes. In many cases, this might be possibility of causal dependence. However, there can be cases of non-causal dependence as well. Second, and related to the first point, although the explanans and explanandum need not be true themselves, claims about their possibility must be. Thus, a theoretical model can only afford understanding if the possibility that it proposes is genuine. Verreault-Julien remains rather short on this matter and proposes to decide the "truth or falsehood of possibility claims (…) on the background of suitable facts, depending both on the modality and on the epistemic goal" (ibid., 14). Third, it needs to be emphasized that Verreault-Julien's case studies could not necessarily show that the checkerboard and the hawk-dove model have yielded understanding of individual occurrences of segregation and/or animal conflict. Instead, they provide information about these phenomena in general. Consequently, whether this broader account of understanding applies should be decided with reference to the kind of object that scientists aim to understand. If the object to be understood is an individual real-world occurrence, then scientific understanding should entail Reutlinger's first criterion of veridicality. Consequently, scientific understanding of specific occurrences requires:

1. **Veridicality condition:** Generalizations $G_1, \ldots, G_m$, the auxiliary statements $S_1, \ldots, S_n$ and the explanandum statement E must be (approximately) true.

If the object of understanding is of general kind (either G1 or G2), then Verreault-Julien's possibility conditions suffices.

1. **Possibility condition:** Generalizations $G_1, \ldots, G_m$, the auxiliary statements $S_1, \ldots, S_n$ or the explanandum statement E are (im)possible according to the relevant modal interpretation and epistemic goal.

This proposal is beneficial as it considers the explorative role that theoretical models often play in economics and preserves the possibility to spell out models' epistemic benefit in terms of

understanding. However, it takes into consideration the epistemic problems of referring to highly simplified theoretical models as source of understanding for specific occurrences.

Two remarks are noteworthy. First, this argument abstracts from the fact that understanding itself, and the differentiation between individual and general phenomena, comes in degrees. Hence, the situation is usually more complex than presented here. Second, and related to the first point, it requires a more precise definitions of the term generality and genericness. In terms of generality, Weisberg's account (2013) provides a helpful framework: defining a general phenomenon as some kind of abstraction over the individual instances of the phenomenon is a good starting point. In some cases, this may be spelled out in terms aggregating features that the specific instances share *(intersection view)*. However, in many cases, this is not possible. Then, general phenomena may be composed of some, but not all relevant, shared features. In other cases, these features may not be strictly shared, but only similar among the specific targets. Yet, in practice, concreteness often demands modelers to go even beyond just these features and build a *construal setting*. In these settings, the model does not only contain properties of the general phenomenon but makes additional assumptions. Particularly in the study of complex phenomena, this is very common. In such cases, it is often more fruitful to define the target system in terms of genericness rather than generality. As was shown, Schelling's checkerboard does not aggregate shared features of instances of residential segregation but instead constructs a hypothetical target system in a generic manner. In these cases, Elliot-Graves proposal (2022) to define genericness as the "absence of particularity" provides a helpful framework (ibid., 74). On her account, generic models are used to investigate phenomena regardless of their particular instances. Thus, even though general and generic models can afford understanding phenomena in general, or general phenomena, there are important differences. This thesis made a compromise by refraining from using the term genericness while nevertheless acknowledging that only using the umbrella term of generality can lead to problems of overgeneralization and misapplication of models. Hence, it will be an important task of future research to suggest more nuanced definitions.

5.3 Summary

This chapter provided an argument to reject the necessity of HAE for scientific understanding. By reference to two paradigmatic cases, I argued that theoretical models can afford understanding without providing HAE. Consequently, it requires a broader account of understanding that can accommodate for such cases. Consequently, I proposed to substitute the usually required condition of veridicality for possibility. However, whether this broader account

is suitable to characterize scientific understanding should depend on the context and be decided with reference to the object of understanding. If this argumentation is accepted, theoretical models in economics can be epistemically valuable by affording understanding of general phenomena.

6. Conclusion

This thesis has been concerned with the epistemic value of theoretical models in economics. The fundamental question that motivated this thesis was whether explanatory inferences from highly simplified models to real-world scenarios can be justified if models cannot accurately represent them.

With reference to Reiss' explanation paradox (Reiss 2012), I argued that the explanatory capabilities of many theoretical models in economics, at least in the sense of providing how-actually explanations, are limited. Consequently, economists need to be careful when referring to highly simplified models as source of explanation. This begs the question of models' epistemic status and their source of value. Driven by the impression that theoretical models can nevertheless afford important kinds of learnings about the world, an answer was sought in the concept of scientific understanding.

Chapter three introduced the concept and showed that traditionally, scientific understanding is conceptualized in an explanatory way (explanatory understanding). Understanding why something is the case is then simply presented as knowing a correct explanation and having epistemic access to it, mostly characterized by an element of grasping. Yet, if scientific understanding is conceptualized in such a way, non-explanatory models are necessarily non-conducive to understanding. However, there are other concepts of understanding. In particular, Elgin's account of objectual understanding was presented (i.e. Elgin 2004 and 2007). Elgin stresses that in order make sense of science's accomplishments, scientific understanding needs to be conceptualized more holistically. Consequently, she lays less focus on understanding the causes of singular occurrences and instead focuses on understanding subject matters as a whole. This differentiation already indicates that an appropriate characterization of understanding is dependent on what exactly is to be understood.

Chapter four analyzed what exactly theoretical models in economics usually investigate and how their role in this investigation looks like. In terms of the latter, their role is mostly explorative. I argued that theoretical models do not yield answers to why something has come about, but rather why and how some phenomenon could come about. In other words, their role lies in suggesting possible explanations. In terms of what they investigate, I demonstrated that theoretical models usually target classes of phenomena. They target general phenomena. In many cases, theoretical modelling in economics is close to what Weisberg terms generalized modelling (Weisberg 2013). Hence, theoretical models target general phenomena conceived as a kind of abstraction over its individual instances. However, I pointed out that Weisberg's notion of generality only fits imperfectly as economists often do not simply aggregate features

of individual occurrences into general phenomena. Target construction is usually more hypothetical. Consequently, Elliot-Graves' notion of genericness (2022) might be more suitable. Independently of the exact definition of generality (and/or genericness), this chapter rendered traditional accounts of understanding as unsuitable for theoretical models in economics.

Chapter five substantiated these claims and supported Verreault-Julien's case study (2019b) that presented paradigmatic cases of understanding without how-actually explanations. Consequently, I proposed a broader account of understanding which substitutes the condition of veridicality for possibility. Importantly, this thesis filled in a missing puzzle piece suggesting that constitutive criteria of scientific understanding need to be chosen with respect to the type of object that is to be understood. If this object is a particular empirical occurrence, defining understanding in terms of a correct (causal) explanation and an element of grasping is appropriate. However, understanding general phenomena should be characterized by different criteria and consequently, the broader account applies. Then, true (causal) explanation is not among the necessary conditions and understanding should be conceived as modal knowledge of counterfactual dependences, which can be supplied by how-possibly explanations as well. As a result, the epistemic value of theoretical models can be spelled in terms of understanding even if these do not support a veridical explanation.

Thereby, this thesis provided a new answer on the question of the epistemic value of highly simplified economic models. It made a suggestion how to reconcile opposing positions in the literature. While critics question in virtue of what exactly theoretical models in economics can really afford understanding to real-world phenomena, defenders continue to provide examples in which theoretical models have taught us meaningful insights about the real-world. Sometimes, a more fine-grained look shows that these positions may not be as incompatible as thought and instead precision about what is really understood is often what matters most. What is left to future research is to define this what more precisely than the scope of this thesis permitted as it is of decisive importance to develop a more comprehensive understanding of the kinds of phenomena that theoretical models target and the notion of generality that is and should be applied in economics.

References

Akerlof, George A. 1970. The Market for "Lemons": Quality Uncertainty and the Market Mechanism'. *The Quarterly Journal of Economics* 84 (3): 488–500.

Alexandrova, Anna. 2008. Making models count. *Philosophy of Science* 75 (3): 383-404.

Aydinonat, Emrah N.. 2005. An interview with Thomas C. Schelling: Interpretation of game theory and the checkerboard model. *Economics bulletin* 2 (2): 1-7.

Aydinonat, Emrah N.. 2007. Models, conjectures and exploration: an analysis of Schelling's checkerboard model. *Journal of Economics Methodology* 14 (4): 429-454.

Banerjee, Abhijit. 1992. A Simple Model of Herd Behavior. *The Quarterly Journal of Economics* 107 (3): 797-817.

Bannock, Graham, R.E. Baxter and Evan Davis. 1998. *Dictionary of Economics*. New York: Kohn Wiley and Sons.

Barlow, Jay. 1992. Nonlinear and logistic growth in experimental populations of guppies. *Ecology* 73 (3): 941-950.

Baumberger Christoph, Claus Beishart and Georg Brun. 2017. What is Understanding? An Overview of Recent Debates in Epistemology and Philosophy of Science. In Stephen Grimm, Christoph Baumberger & Sabine Ammon (eds.), *Explaining Understanding: New Perspectives from Epistemology and Philosophy of Science*: 1-34. Routledge.

Bokulich, Alisa. 2016. *Fiction as a Vehicle for Truth: Moving Beyond the Ontic Conception*. The Monist 99 (3): 260-279.

Box. G and N. Draper. 1987. *Empirical model-building and response surfaces*. New York: John Wiley & Sons.

Cartwright, Nancy. 1983. *How the laws of physics lie*. Oxford, NY: Oxford University Press.

Cartwright, Nancy. 2009. 'If No Capacities Then No Credible Worlds. But Can Models Reveal Capacities?'. *Erkenntnis* 70 (1): 45–58.

Craver, Carl. 2006. When mechanistic models explain. *Synthese* 153 (3): 355-376.

De Regt, Henk and Dennis Dieks. 2005. A Contextual Approach to Scientific Understanding. *Synthese* 144 (1): 137-170.

De Regt, Henk and Christoph Baumberger. 2019. What is Scientific Understanding and how can it be achieved? In Kevin McKain and Kostas Kampourakis (eds.), *What is Scientific Knowledge? An Introduction to Contemporary Epistemology of Science:* 66-81. New York: Routledge.

Elgin, Catherine Z.. 1991. Understanding: Arts and Science. *Midwest Studies in Philosophy* 16: 196-208.

Elgin, Catherine Z.. 1996. *Considered Judgement*. Princeton, NJ: Princeton University Press.

Elgin, Catherine Z.. 2004. True Enough. *Philosophical Issues* 14 (1): *113-131*.

Elgin, Catherine Z.. 2007. Understanding and the Facts. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 132 (1): 33-42.

Elliot-Graves, Alkistis. 2022. What are general models about? *European Journal for Philosophy of Science* 12, 74.

Friedman, Milton. 1953. The Methodology of Positive Economics. In E*ssays in Positive Economics*, Chicago, IL: University of Chicago Press, 3-46.

Grimm, Stephen. 2006. Is Understanding a Species of Knowledge? *British Journal of Philosophy of Science* 57 (3): 515-536.

Grüne-Yanoff, Till. 2009. Learning from minimal models. *Erkenntnis* 70 (1): 81-99. *of Economic Methodology* 20 (3): 255-261.

Grüne-Yanoff, Till. 2013. Genuineness resolved: a reply to Reiss' purported paradox. *Journal of Economic Methodology* 20 (3): 255-261.

Hausman, Daniel M. 1992. *The Inexact and Separate Science of Economics*. Cambridge, NY. Cambridge University Press.

Hedström, Peter. 2005. *Dissecting the Social*. UK: Cambridge University Press.

Hedström, Peter and Petri Ylikoski. 2010. Causal mechanisms in the social sciences. *Annual Review Of Sociology* 36 (1): 49-67.

Hempel, Carl Gustav. 1965. Aspects of scientific explanation. In *Aspects of scientific explanation: and other essay in the philosophy of science*: 331-497. New York: Free Press.

Hotelling, Harold. 1929. Stability in Competition. *Economic Journal* 39 (153): 41-57.

Kim, Jaegwon. 1994. Explanatory Knowledge and Metaphysical Dependence. *Philosophical Issues* 5: 51-69.

Khalifa, Kareem. 2012. Inaugurating Understanding or Repackaging Explanation. *Philosophy of Science* 79 (1): 15-37.

Knuuttila, Tarja. 2011. An artefactual approach to model-based representation. *Studies in History and Philosophy of Science* 42 (2): 262-271.

Kvanvig, Jonathan. 2003. *The Value of Knowledge and the Pursuit of Understanding*. New York: Cambridge University Press.

Lipton, Peter. 2009. Understanding without explanation. In H.W. de Regt, S. Leonelli & K. Eigner (eds.), *Scientific Understanding: Philosophical Perspectives*: 43-63. University of Pittsburgh Press.

Mäki, Uskali. 1992. On the Method of Isolation in Economics. *Poznan Studies in the Philosophy of Sciences and the Humanities* 26: 19-54.

Mäki, Uskali. 1994. Isolation, Idealization and Truth in Economics. In *Idealization in Economics*, edited by. B. Hamminga, N. de Marchi, 147-168. Amsterdam: Rodopi.

Mäki, Uskali. 2013. On a Paradox of Truth, or How Not to Obscure the Issue of Whether Explanatory Models Can Be True. *Journal of Economic Methodology* 20 (3): 268–279.

Mankiw. Gregory. 2018. *Principles of Economics,* (8th ed.). Boston, MA: Harvard University Press.

Marchionni, Caterina. 2017. What is the Problem with Mode-based Explanation in Economics? *Disputatio* 9 (47): 603-630.

Maynard Smith, John & G.R. Price. 1973. The logic of animal conflict. *Nature 246*: 15-18.

Maynard Smith, John & G.A. Parker. 1976. The logic of asymmetric conflict. *Animal Behaviour* 24 (1): 159-175.

Muldoon, Ryan, Tony Smith and Michael Weisberg. 2012. Segregation that no one seeks. *Philosophy of Science* 79 (1): 38-62.

Newman, Mark. 2012. An inferential Model of Scientific Understanding. *International Studies in the Philosophy of Science* 26: 1-26.

Potochnik, Angela. 2020. Idealization and many aims. *Philosophy of Science* 87 (5): 933-943.

Pritchard, Duncan. 2014. Knowledge and Understanding. In Abrol Fairweather (eds.), *Virtue Epistemology Naturalized*. Synthese Library, vol 366. Springer, Cham.

Reiss, Julian. 2012. The explanation paradox. *Journal of Economics Methodology* 19 (1): 42-62

Reiss, Julian. 2012b. Idealization and the aims of economics: Three cheers for instrumentalism. *Economics and Philosophy* 28 (3): 363-383.

Reiss, Julian. 2013. The explanation paradox redux. *Journal of Economic Methodology:* 20 (3): 280-292

Reutlinger, Alexander. 2016. Is there a monist theory of causal and noncausal explanation? The counterfactual theory of scientific explanation. *Philosophy of Science* 83 (5): 733-745.

Reutlinger, Alexander, Dominik Hangleitner and Stephan Hartmann. 2018. Understanding (with) Toy Models. *The British Journal for the Philosophy of Science* 69 (4): 1069-1099.

Rohwer, Yasha and Collin Rice. 2013. Hypothetical pattern idealization and explanatory models. *Philosophy of Science* 80 (3): 334-355.

Schelling, Thomas. 1978. *Micromotives and Macrobehaviour*. New York. Norton.

Strevens, Michael. 2008. *Depth: an account of scientific explanation*. Cambridge: Harvard University Press.

Strevens, Michael. 2013. No understanding without explanation. *Studies in History and Philosophy of Science Part A* 44 (3): 510-515.

Sugden, Robert. 2000. Credible worlds: the status of theoretical models in economics. *Journal of Economics Methodology* 7 (1): 1-31.

Sugden, Robert. 2009. Credible worlds, Capacities and Mechanisms. *Erkenntnis* 70 (1): 3–27.

Sugden, Robert. 2013. How Fictional Accounts Can Explain. *Journal of Economic Methodology* 20 (3): 237–43.

Trout, J.D.. 2007. The psychology of scientific explanation. *Philosophy Compass* 2 (3): 564-591.

Verreault-Julien, Philippe. 2017. Non-causal understanding with economic models: the case of general equilibrium. *Journal of Economic Methodology* 24 (3): 297-317.

Verrault-Julien. 2019. How could models possibly provide how-possibly explanations. *Studies in History and Philosophy of Science Part A* 73: 22-33.

Verreault-Julien, Philippe. 2019b. Understanding does not depend on (causal) explanation. *European Journal of Philosophy of Science* 9: 18.

Verreault-Julien, P. 2021. Explanation in Economics. In C. Heilmann, & J. Reiss (Eds.), *The Routledge Handbook of Philosophy of Economics:* 300-315. Routledge Taylor & Francis Group.

Weisberg, Michael. *2013. Simulation and Similarity: Using Models to Understand the World*. New York, US: Oxford University Press.

Wimsatt, William. 2007. *Re-engineering philosophy for limited beings: piecewise approximations to reality*. Cambridge, MA: Harvard University Press.

Woodward, James. 2003. *Making things happen: a theory of causal explanation*. New York: Oxford University Press.

Ylikoski, Petri and Emrah N. Aydinonat. 2014. Understanding with theoretical models. *Journal of Economics Methodology* 21 (1): 19-36.

Appendix

English abstract

Theoretical models in economics can at best represent a highly idealized version of the real economy. This raises the question in how far such models can still help to understand and explain real-world phenomena. In an influential essay, Julian Reiss (2012) argues that economic models rely on three mutually inconsistent hypotheses. First, economic models are false, second, economic models are nevertheless explanatory, third, only true accounts can explain. Reiss' paradox reflects that explanation is usually assumed to require factivity (true explanans and explanandum) and it is doubtful in how theoretical models can provide that.

This master thesis contributes a new perspective on the epistemic value of theoretical models by developing a position that rejects Reiss' second premise. Thus, theoretical models alone are not explanatory. They are used to explore potential explanations of general phenomena. Nevertheless, theoretical models can afford scientific understanding. However, this understanding is not generated via the traditionally assumed mechanism of providing explanation. Hence, a central argument of this thesis consists in detaching scientific understanding from the often-assumed necessity of true explanation. Instead, I will suggest that an appropriate definition of understanding needs to be chosen with respect to the kind of object that is to be understood. If this object is an individual occurrence, the traditionally assumed criterion of true explanation is appropriate. However, understanding general phenomena requires different criteria. In these cases, potential explanations providing the necessary modal information can suffice. Consequently, theoretical economic models can afford understanding of general phenomena without providing true explanation.

German abstract

Theoretische Modelle in der Wirtschaftswissenschaft können bestenfalls eine stark idealisierte Version der realen Wirtschaft darstellen. Dies wirft die Frage auf, wie solche Modelle dennoch zum Verständnis und zur Erklärung realer Phänomene beitragen können. Julien Reiss' „Erklärungsparadox" (2012) ist beispielhaft für diese Frage. Reiss argumentiert, dass ökonomische Modelle auf drei inkonsistenten Hypothesen beruhen. Erstens, ökonomische Modelle sind falsch, zweitens, ökonomische Modelle besitzen dennoch Erklärungskraft, drittens, nur wahre Darstellungen können erklären. Reiss' Argumentation ist exemplarisch für die Annahme, dass Erklärungen Faktizität (wahres Explanans und Explanandum) voraussetzen und es fraglich ist, inwieweit theoretische Modelle dies leisten können.

Diese Masterarbeit argumentiert für eine neue Perspektive in Bezug auf den epistemischen Wert theoretischer Modelle. Diese beruht auf einer Ablehnung der zweiten Prämisse des Paradoxons. Demnach sind theoretische Modelle allein nicht erklärend. Sie dienen der Erforschung möglicher Erklärungen allgemeiner Phänomene. Dennoch können theoretische Modelle wissenschaftliches Verständnis ermöglichen. Dieses Verständnis wird jedoch nicht über den traditionell angenommenen Mechanismus der tatsächlichen Erklärung erzeugt. Ein zentrales Argument dieser Arbeit besteht somit darin, das Konzept des wissenschaftlichen Verstehens von der traditionell unterstellten Notwendigkeit einer wahren Erklärung zu lösen. Stattdessen sollte eine angemessene Definition wissenschaftlichen Verständnisses unter Berücksichtigung der Art des zu verstehenden Objekts gewählt werden. Handelt es sich bei diesem Objekt um ein individuelles Ereignis, so ist die traditionell angenommene Definition, welche eine wahre Erklärung voraussetzt, angemessen. Das Verständnis allgemeiner Phänomene erfordert jedoch andere Kriterien. In diesen Fällen können potenzielle Erklärungen ausreichen. Folglich können theoretische Wirtschaftsmodelle zum Verständnis allgemeiner Phänomene beitragen, ohne den Kriterien einer wahren Erklärung Genüge zu tun.