Philippe Verreault-Julien

Understanding With Models

PhD thesis—Erasmus University Rotterdam

SUPERVISORS
Julian Reiss
Jack Vromen

UNDERSTANDING WITH MODELS

Hoe modellen onze de wereld helpen begrijpen

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on

Thursday the 17th of January 2019 at 15.30 hrs
by

Philippe Verreault-Julien
born in Québec, Canada

Erasmus University Rotterdam

Doctoral Committee:

Promotors:

Prof.dr. J.J. Vromen

Prof.dr. J. Reiss

Other members:

Prof.dr. F.A. Muller

Prof.dr. T. Grüne-Yanoff

Dr. H.C.K. Heilmann

Philippe Verreault-Julien
Understanding With Models

TITLEBACK
ISBN:
Cover design by Andrée-Ann Cloutier.

CONTACT
☞ pvj@pvjulien.net

Empirical science, in all its major branches, seeks not only to *describe* the phenomena in the world of our experience, but also to *explain* or *understand* them.

–Carl G. Hempel (1965b, p. 297)

To the loving memory of *grand-mère*.

1926 – 2014

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

EIPE Erasmus Institute for Philosophy and Economics

ER-Problem epistemic representation problem

FInfR factive inferentialist account of representation

HAE how-actually explanation

HPE how-possibly explanation

InfBUn inferentialist-behavioural account of understanding

KnUn knowledge account of understanding

w-questions ‘what-if-thing-had-been-different’ questions

w-inferences ‘what-if’ inferences

1 INTRODUCTION

During a meeting with one of my supervisors, he asked me a question which basically had the following meaning: "Why do you want to defend economics that much?" The innuendo was that a lot of my research appeared to be about finding loopholes that would salvage the epistemic credibility of highly idealized models. He was of course right. One implication of my work is that we often mix up genuine understanding or knowledge with epistemic frivolousness. Another implication is that I am lending a hand to the social science that is already at the top of the disciplinary hierarchy (Fourcade et al. 2015). For someone whose initial impetus to study philosophy of economics was a (very) critical attitude towards 'the dismal science' and who cited Kropotkin in the students' association journal, it may look as if I took a bad turn somewhere along the way.[1]

Indeed, an important motivation of mine when beginning these studies was to understand better on what grounds economists were justifying their claims. I just had a hard time fathoming why economists gave credence to their unrealistic models and followed them for policy-making. It seemed to me that not only their models were methodologically questionable, but that they were also morally perverse.

My undergraduate professor of Aristotelian logic used to say that ideas are like punching bags: you need to hit them hard to test their robustness. If it was in a sense the ethics of economics that prompted me to study the discipline, it is with the tools of philosophy of science that I wanted to analyze and peel its theoretical carapace. Understanding how economists rationally justify the foundations they give to their science and assessing whether or not we should grant our credibility to these justi-

1 For those who might be interested: " 'Struggle so that all may live this rich, overflowing life. And be sure that in this struggle you will find a joy greater than anything else can give.' This is all that the science of morality can tell you. Yours is the choice" (Kropotkin [1897] 2002, p. 113).

fications is a task that, I believe, partly lies within the realm of philosophy. What brought me to philosophy of economics was by and large politically motivated, yet understanding—and, why not, criticizing—economics nevertheless requires an honest epistemological appraisal.

My honest epistemological appraisal quickly led me—after long hours trying to make sense of F.A. Hayek's at times inconsistent epistemological prescriptions (Verreault-Julien 2010)—to economic models. If you want to examine the work of blacksmiths, you better look at how they handle hammers. The economist's toolbox is full of models of all shapes and sizes.

> Today, if you ask a mainstream economist a question about almost any aspect of economic life, the response will be: suppose we model that situation and see what happens. It is important, then, to understand what a model is and what it is not (Solow 1997, p. 43).

If I wanted to understand economics, I had to understand models. I quickly realized—after long hours scratching my head over articles on scientific representation—that this was easier said than done. As I had developed interests in issues concerning scientific explanation and about the instrumentalism/realism debate, I was immediately drawn to Reiss's (2012b) "explanation paradox" when I had to read it for a class. I viewed it as a very fruitful way of approaching the problem of the epistemic import of models in economics, but also more generally. 'We have good reasons to think that models explain, yet this is a conclusion that we also have good reasons to resist. What should we do?' Since I did not believe that our theories of explanation were the problem and since it seemed to me that the model-world relationship was often very tenuous, then the only alternative was to conclude that economic models did not explain. Despite my prima facie preference for this verdict, Reiss yet again phrased one colourful objection it was facing.

> Why do economists build complex, mathematically sophisticated models rather than, say, resort to creativity and intuition, crystal balls, hypothesis-generating algorithms or consciousness-enhancing drugs? All of

> these sources of inspiration would be a lot easier to come by, and some of these would be more fun, than doing the hard work of constructing and solving a model (Reiss 2013c, p. 282).

There are probably a number of reasons why economists like so much their models. Proficiency in mathematics has for some time been considered very important for success in the discipline (Colander and Klamer 1987; cf. Colander 2007). Mathematical modelling is also often considered to be a mark of scientific rigour (Backhouse 2010, 99ff.). And perhaps there is some truth to Paul Krugman's (2009) post-crisis pronouncement that "economists, as a group, mistook beauty, clad in impressive-looking mathematics, for truth". Regardless of the plausibility of each of these explanations, it seemed to me it did not capture everything that was going on and that I needed to give economic models, like peace, a chance.

Thus, another explanation may just be that "simple models of the type that economists construct are absolutely essential to understanding the workings of society" (Rodrik 2015, p. 11). *Understanding*, that was it. At the time when I started to work on that topic, there had been a renewed interest in the notion of understanding and in its relationship with its sibling/parent, explanation. Since Hempel (1948; 1965a), the epistemology of understanding had been for all practical purposes reduced to the epistemology of explanation. Theories of scientific explanation all claim explanations provide understanding (Achinstein 1983; Michael Friedman 1974; Hempel 1965a; Kitcher 1981; Salmon 1984; Strevens 2008; Woodward 2003), but they view it solely as a by-product of explanation and not as an object of study in its own right. However, the year 2000s saw an important rehabilitation of understanding as a notion worthy of inquiry (de Regt 2004, 2009; de Regt and Dieks 2005; de Regt, Leonelli, et al. 2009; Grimm 2006, 2008, 2010; Khalifa 2011).[2] It seemed to me that

2 Of course, 'understanding' has a long history in the philosophy of the social sciences where the Weberian '*Verstehen*' (interpretative understanding) was distinguished with '*Erklären*' (explanation). The current debate on understanding has rather its roots in the more general debate on the nature of scientific explanation. One underlying premise is that the sciences, natural and social, all aim at producing understanding of empirical phenomena.

understanding could be the key to offer a novel solution to the explanation paradox.

The guiding idea behind my research has been that *perhaps models do not explain, yet afford understanding*. This solution, I thought and still think, has maximum benefits for limited costs. What I purported to do with this thesis is to lay the ground for an epistemological framework that reconciles plausible philosophical accounts of representation, idealization, and explanation while also taking seriously the economists'—and other scientists'—claim that we understand the world better because of the models they build. Not only do I think it is reasonable to believe that the epistemic judgments of whole disciplines are not wholly misguided, but I also think that it is what an 'honest epistemological appraisal' calls for. Let us make the strongest case for models and let us see where it leads us.

1.1 UNREALISTIC ASSUMPTIONS AND THE EXPLANATION PARADOX

Do businesspeople act in a way to maximize profits? Do individuals have a perfect knowledge of the economy? Are there no transaction costs? Are markets efficient? Are people only self-interested? Is competition perfect? The short answer to all these questions is a simple "No". We know these statements, taken as universal generalizations, to be false descriptions of the world we live in. Yet, various economic models contain idealizations of this sort. The long answer, as the vast literature devoted to discuss the methodological implications of unrealistic assumptions, is more complicated.

Indeed, the use and status of unrealistic assumptions has a long history in economics. One of the first methodological treatise on economic methodology was Mill's (1844) *Essays on Some Unsettled Questions of Political Economy*. Mill argued that the claims of economics were true "*in the abstract*" (Mill 1844, 144-45, emphasis in original). To know what would actually happen also requires to know all the disturbing factors and other countervailing forces as well as how they combine together.

The problematic status of unrealistic assumptions came under light during what has been called the 'marginalist controversy' (Backhouse 2009; Mongin 1997). Basically, the controversy stems from a series of studies which showed that, contrary to what the marginalist theory of the firm implied, firms were not actually behaving in profit-maximizing fashion. Most famously, Hall and Hitch (1939) used survey data to conclude that businesspeople were not fixing prices using information about marginal cost and revenue, but were rather following simple rules of thumb like the 'full cost principle'. This principle involves calculating ex ante the costs of producing a unit of output and then fixing its price by adding a margin of profit to the costs.

The marginalist controversy sparked a debate on the status of unrealistic assumptions. There are many ways to respond to what appears to be deviating price-setting behaviour. One can conclude that the profit-maximization theory shall be rejected, one can simply reject the validity of the evidence, or one can amend the theory so that full cost price-setting can be reconciled with the data. At any rate, facing data that prima facie contradict key theoretical assumptions, economists felt the need to provide a methodological response.

The single most famous response to the controversy is Friedman's (1953) *The Methodology of Positive Economics*.[3] In this essay, Friedman advocates what has often been interpreted as an instrumentalist position (Blaug 1992; Boland 1979; Caldwell 1994; Reiss 2012a).[4] According to him, "the only relevant test of the *validity* of a hypothesis is comparison of its predictions with experience" (1953, 8-9, emphasis in original). In short, he claimed that the realisticness of the assumptions was irrelevant to a theory's appraisal and that only its predictive power mattered. Friedman's text came to be "one of the most influential texts in the methodological literature of twentieth-century economics" (Blaug 2009, p. 351; see also Boland 1989; Mayer 2009).

Despite efforts to legitimize the role of unrealistic assumptions, the underlying concern raised by the marginalist controversy was not completely dispelled. Indeed, issues concerning the sim-

3 Although similar to Friedman's, but arguably philosophically sounder, Machlup's (1955) defence was not as influential.

4 For a dissenting interpretation, see Mäki (2009b).

plifications, abstractions, and other idealizations are "*the* central questions in economic methodology" (Hausman 2018, emphasis in original). To what extent the idealizations of economics are harmful or helpful is still to a large extent an open question.

A reason why a position à la Friedman can become questionable is when we consider the goals of economics. Menger (1985), for instance, recognized three aims to economics: explanation, prediction, and control. If one only cares about the narrow predictive success of economics, perhaps there is no reason to "look under the hood" (cf. Hausman 1992a, pp. 70-73) of theories. Perhaps one can only judge a theory by its predictive capacity. This is why Hausman (1998) argues that what is really at stake in the instrumentalism/realism debate in economics is a question concerning its goals: should they be purely practical (instrumentalism), or should they also be cognitive? If they are also cognitive, like explanation, then determining the status of the assumptions becomes a more pressing issue. For if the cognitive goal of explanation requires truth, then false assumptions are at first sight an obstacle to the achievement of this goal. One key challenge unrealistic assumptions pose thus concerns the explanatoriness of economic models.

But, one may ask, do economists actually aim to explain economic phenomena? It appears they do. Economists regularly ask explanation-seeking questions like 'Why are there business cycles?', 'Why did the 2007-08 financial crisis occurred?', 'Why is there unemployment?', or 'Why did the 1970s suffered from an episode of so-called 'stagflation'?' All these questions call for an explanation: we want to know *why* the events occur. We can also find the language of explanation in textbooks or seminal academic papers.

> To understand how the economy works, we must find some way to simplify our thinking about all these activities. In other words, we need a model that explains, in general terms, how the economy is organized and how participants in the economy interact with one another (Mankiw 2015, p. 22).

> A competitive equilibrium model was developed and used to explain the autocovariances of real output

> and the covariances of cyclical output with other aggregate economic time series for the post-war U.S. economy (Kydland and Prescott 1982, p. 1368).

> The proposed explanation correspondingly focuses on generalized external economies rather than those specific to a particular industry (Krugman 1991, p. 485).

> [O]ur point is that many aspects of herd behavior can be explained quite plausibly without invoking these kinds of gains from association (Banerjee 1992, p. 801).

But despite economists' self-avowed goal of providing explanations, whether they actually achieve it is a point of contention. Can they explain the world if they rely on assumptions that we *know* are descriptively false of the world we live in? Do shocks in technology or productivity explain business cycles if there are no such shocks? Would the greed of bankers explain why the Great Recession occurred if bankers had not actually been greedy? Is unemployment a voluntary decision if in fact people preferred to work? In a nutshell, can unrealistic assumptions explain?

1.1.1 The explanation paradox

A contemporary characterization of this long-established problem has been called the "explanation paradox" (Reiss 2012b). The explanation paradox can be formulated this way:

1. Models are highly idealized; they misrepresent reality.
2. Our best philosophical theories of explanation require faithful representation.
3. Yet, models appear to explain phenomena.

Reiss claims that there are good reasons for independently holding these three propositions. However, since they are mutually inconsistent, this leads to a paradox; something has to give. What I have always found interesting with the 'paradox' is that it neatly showcases how various strands of the literature deal with the overarching issue of accounting for the explanatoriness of economic models. Indeed, what Reiss's explanation paradox highlights is that there are already accounts that attempt to tackle this issue and that these accounts are perhaps not as successful as it seems.

There are basically three sets of solutions. The first attempts to show that despite models using idealizations and misrepresenting reality, they still capture true parts of the world we live in. They simplify reality in order to 'isolate' the relevant explanatory factors. If some parts of a model faithfully represent, the argument goes, then these may explain and thus the paradox vanishes. Proponents of this solution are, for example, Cartwright (1989) and Mäki (1992, 2009a). However, whether this is an appropriate characterization of what models do is disputed. For instance, Grüne-Yanoff (2011) argues that modellers do not aim to isolate causal factors and that even causal factors of interest tend to be idealized. Cartwright (2009) argues that since there are few robust economic principles, then models must cope with a host of structural assumptions which then also drive the results. Thus, it becomes unclear what models are isolating in the first place.

The second set of commentators maintains that it is the requirement of faithful representation which is inadequate. According to proponents of these solutions, philosophical theories are wrong and models that misrepresent can still explain phenomena. If misrepresentation and explanation are in fact compatible, this may solve the paradox. While there were various theories of explanation in the general philosophy of science literature which relaxed this requirement (e.g. van Fraassen 1980; Kitcher 1981), they have not been as influential as the causal account (Reutlinger 2017b). In the context of modelling, Sugden (2009, 2013), for one, submits that models are akin to fictions and that judgments of similarity between model and target underscore their explanatoriness. According to him, these judgments are to some extent subjective. But if a model needs to be similar to its target in order to explain, then one may ask whether the model really fails to faithfully represent. Unless one adopts a pragmatic concept of similarity, it seems it does not solve the paradox. And if one does adopt such a concept, there is still the problem that theories of scientific explanation typically require (approximate) truth or faithful representation.

The final set simply denies that models actually explain phenomena. If we trust that our best philosophical theories are correct and that models indeed misrepresent reality, then the logical conclusion is to deny that they explain phenomena. Perhaps models serve other purposes, e.g., heuristic ones, but do not directly explain. When economists say that their models explain, it should therefore be taken with a grain of salt. They do not explain in the philosophical sense of the term. They may serve other purposes, but explanation is not one of them. This solves the paradox by showing that, in fact, practitioners' judgments are mistaken and that therefore the third proposition is false.

Advocates of this position are, for instance, Hausman (1992b), Alexandrova (2008), Alexandrova and Northcott (2013), or Grüne-Yanoff (2009, 2013a). For one, Alexandrova (2008) denies that models per se explain. Instead, she claims, it is their material concretization in the form of, for example, experiments that carry the explanatory work. Grüne-Yanoff pursues a different strategy. He denies that model actually explain, but claim they provide so-called 'how-possibly' explanations. These how-possibly explanations, he argues, can help us learn about the world under certain circumstances. One problem Reiss (2012b) sees with this line of defence is that it does not account for economists' judgments of explanatoriness. Denying that models actually explain may solve the paradox, but it also comes at a hefty cost, viz. attributing systematic error to practitioners.

1.1.2 A novel solution

From the time of my research master onwards, it seemed to me there was another way of solving the paradox. Without claiming that other answers are in all cases incorrect or fatally flawed, I thought that a different—and in my mind better—solution was available. In a sense, the following thesis can be viewed as an attempt to 'solve' the explanation paradox. It builds on two trends in contemporary philosophy of science and epistemology.

Firstly, whether causal explanation is the one and only type of legitimate explanation is more and more put into question

(Lange 2017; Reutlinger and Saatsi 2018a). If explanations do not have to be causal, this is germane to the discussion insofar as a lot of theoretical economics appears to be preoccupied with achieving mathematical results more than establishing causal knowledge. This suggests that "blackboard economics" (Coase 1992) and the widespread use of formal methods that is so often decried (e.g. Blaug 2002; Chick 1998; Lawson 2009; Rosenberg 1992) may not be epistemically futile, after all.

Secondly, as I said above, there has been in the past fifteen years or so a revival of taking the notion of 'understanding' as a legitimate object of philosophical study. This was inaugurated in work by, e.g., Trout (2002), de Regt (2004), and de Regt and Dieks (2005). Historically, even though the close connection between understanding and explanation has long been recognized (Michael Friedman 1974), 'understanding' has seldom received independent attention, to the benefit of its parent term 'explanation'. Recent research studies if and to what extent the epistemology of understanding can be reduced to the epistemology of explanation.

Within this literature, one can find two distinct projects: 1) providing an account of understanding that tells us under what conditions explanations provide us with understanding, and 2) providing an account of understanding such that understanding can be achieved without an explanation.[5] The second project is of particular interest because traditionally having an explanation was considered to be a sufficient and necessary condition for understanding (Khalifa 2012). However, if understanding can be had without having an explanation, that may the key to eating our proverbial cake and eat it too; models may fail to explain, yet afford understanding.

All that being said, what is more precisely this solution and how does it fit in terms of the explanation paradox? Basically, the overarching project of this thesis is to show that models may provide epistemic benefits in the form of understanding even when they do not actually explain or when they do not provide causal knowledge. This solution challenges the third proposition of the paradox, viz. that models explain. In this respect, I therefore side with those who claim that economic

5 I owe this distinction to Nounou and Muller (2015).

models do not typically explain phenomena by identifying its true causes. However, the novelty of my approach is that it does not come at the costs usually associated with this set of solutions. These costs are of two kinds.

First, an undesirable implication of denying that models explain is that it seems to entail that economists are under a systematic and deep delusion about their own work. Economists may think that they explain and provide understanding of the world, but they are wrong and achieve neither. I consider that this blanket claim does not readily fit with a broadly construed naturalistic philosophy of science. Practitioners can of course be wrong. But it is also undesirable for a philosophical account to conflict with sentiments that are widespread in the practice of a science. As Hausman argues, methodologists "should be suspicious of accounts that attribute to economists egregious and persistent errors" (2009, p. 40). That is not to say that philosophy can't criticize economics, but rather that we should be careful before dismissing a scientific practice that appears to afford genuine understanding for economists.

An important motivation of mine is to take seriously what practitioners think of their work and how they reveal their epistemic preferences by choosing to engage in theoretical modelling. Despite all the valid criticisms one can raise against economics and the practice of theoretical modelling, I believe that there is a philosophical mistake in not taking scientists judgments more at face value. Science is *hard*. Mistakes are unavoidable. And if the past is any indication of the future, theories that we have currently all the reasons to believe are true will turn out to be mistaken, if not flat out wrong. Yet, we sent probes with golden records outside the solar system. We eradicated smallpox. I can call family and friends in Québec for free over the Internet while walking in the streets of Rotterdam.[6] We designed algorithms that help people receive the kidney transplant they so direly need. The burden of proof shall be on those who dismiss the epistemic success of science.

The second cost is that denying the explanatoriness of models seems to imply denying they afford epistemic benefits tout court. If models do not explain, then what are they doing for us?

6 I should probably do that one more often.

Hausman (1992b) and Alexandrova (2008), for instance, argue that models suggest causal hypotheses and thus have a heuristic function. But even though this may be an important role of models, models also appear to afford us genuine insight into why the world is the way it is. Reiss expresses a similar point:

> To warrant their existence, models must do more than to provide hypotheses. They must have some genuine epistemic benefit. EP (Reiss 2012b) asks but does not assume that this benefit is explanation. If not explanation, however, there must be something else [. . .] (Reiss 2013c, p. 282).

This 'something else', I contend, is understanding. According to the proposal I develop, it is possible to deny that models explain while also granting that they afford understanding. It hence responds to Reiss's (2013c) demand that a philosophical account of models grant them a genuine epistemic benefit. This is a promising strategy insofar as economics seldom seem to provide fully-fledged explanations of phenomena, yet appear to offer some insight about the world. Hence, perhaps economists do not explain the world, but this may not come at the cost of not improving our understanding of it. We may have understanding without an explanation.

But, one may object, am I not committing the same mistake, i.e., going against practitioners' judgments of explanatoriness? I do not think I am, for two reasons. First, as Alexandrova and Northcott (2013) acknowledge, denying that economic models explain implies the need for an error theory: what may explain that economists are mistaken? They suggest three possibilities:

1. 'Explanation' is an ambiguous term and economists may not use it properly.[7]

2. People are prone to all sorts of cognitive biases, most notably hindsight and overconfidence (see Trout 2002).

3) Explanatory intuitions are not conducive to reliable judgments.[8]

7 Grüne-Yanoff (2013b, p. 258) proposes an explanation along these lines.
8 See Grimm (2009) for a more positive outlook on the reliability of these intuitions.

Reiss (2013c) finds these three hypotheses lacking. While I basically side with Reiss concerning 2) and 3), I think that 1) deserves more consideration.[9] Following a distinction due to Salmon (1984) between ontic and epistemic explanations, Alexandrova and Northcott (2013) view the former type as explaining by virtue of identifying the objective features of reality responsible for the phenomenon, typically causes. By contrast, according to Alexandrova and Northcott, epistemic explanations explain by making the phenomenon expected, or less surprising. Since they believe epistemic explanations to be inherently subjective, they reject their validity and thus economists' judgments. Reiss argues—rightly—that this presumes that there is only one correct type of explanation (ontic).

But, it seems to me, the relevant confusion that would explain mistaken explanatory judgments is not so much between epistemic and ontic explanations, but rather, as Grüne-Yanoff (2013b, p. 258) points out, between 'how-possibly' and 'how-actually' explanations. Whereas a how-actually explanation (HAE)—or explanation simpliciter—provides an account of how a phenomenon actually occurred by citing its (approximately) true explanans and explanandum, a how-possibly explanation (HPE) does not provide such an account. As Reiss argues, "[a] 'how-possible "explanation" ' is not an explanation. It is *possibly* an explanation" (2013a, 111, emphasis in original).

Even though there is now a substantial literature (e.g. Craver 2006; Forber 2010; Reydon 2012; Grüne-Yanoff 2013a; Bokulich 2014; Ylikoski and Aydinonat 2014; Rice 2016) likening theoretical models to HPEs, there is no established account that provides clear criteria on how to demarcate them from HAEs. Philosophers do not even agree on whether HPEs are a genuine species of explanation (see Dray 1968; Forber 2010). If this is the case, then it should hardly come as a surprise that scientists, economists in-

9 Without entering in the details, I consider that whereas 2) and 3) may be plausible at the individual level, they are not at the level of whole epistemic communities like economics. Science is organized such as to provide checks and balances on the possible hubris of individual researchers. There may be sociological reasons for why scientists adopt inadequate epistemic values or methodologies, but I doubt these come from cognitive biases or the phenomenology of explanation. Perhaps individual economists can be misled about the explanatoriness of its models, but can a whole discipline be?

cluded, mix up HAEs with HPEs. I thus agree with Grüne-Yanoff (2013b, p. 258) who says that "[...] committing this conceptual confusion is hardly a crime, as no generally accepted account of this distinction is extant". In this sense, it is not so much that economists are mistaken, simply that they use the same concept of 'explanation' to describe two different exercises. And since we do not have a clear account we can't blame them for doing so. In other words, I do not think that economists typically mistake HPEs for HAEs (even though they might do so in specific cases), but simply that they use the same words to describe the two.

The second reason I do not think I am going against practitioners' judgments is, in fact, quite similar to the first. When economists tell us their models explain, I take the main contribution of the models to be that they afford *understanding*. The crucial point is that even if I deny that models provide explanations—in the HAEs sense—I do not deny the sort of epistemic contribution they are making, i.e., understanding. Hence, if I am going against the judgments of practitioners—and perhaps I am not at all considering the ambiguity between HAEs and HPEs—, it is only in a very limited way. Furthermore, it is also a potentially far less important mistake to attribute because the cognitive benefits are categorically similar; economists are right to hold their models afford understanding. Once more, if philosophers themselves do not distinguish explanation from understanding, why should we expect scientists to do so? As I do not think the failure is with our account of explanation, I also do not think the failure is with economists. Instead, the failure is with our account of understanding.

In light of these two reasons, my solution does not imply that economists are under a collective blameworthy delusion. It offers a simple error theory—the lack of a suitable conceptual framework—that does not depend on attributing biases, cognitive or institutional, to practitioners. We can't blame economists for lacking in conceptual precision when philosophers commit the same sin. If anything, philosophers should take responsibility for not having cleared-up the muddy conceptual waters of explanation and understanding. This thesis endeavours to participate in the cleaning effort.

It also remains within the realm of a naturalistic philosophy of science, but one that still has some bite. If philosophy of science should be careful before systematically condemning scientific practices, it should also not condone them at all costs (Hausman 2009). The solution I propose strikes a reasonable balance between taking what economists do seriously, while also calling attention to conceptual inaccuracies.

My project is therefore both philosophy and science-directed (Currie 2015). It is philosophy-directed in that it looks at science in order to claim and support philosophical theses. Examining science may raise various challenges to traditional accounts of explanation and understanding. For instance, theoretical modelling suggests, among other things, that many significant epistemic benefits models provide fall short of being (actual) causal explanations. Looking at how science is conducted can pinpoint where philosophy fails in reconstructing and appraising the results of science.[10] It is also science-directed in that philosophy can also inform our appraisal of what science actually achieves and thus change our perspective on the world.

This is why I believe a revised epistemology of understanding, which my thesis contributes to, not only sheds light on scientific practice, but can also provide new evaluative tools to practitioners and methodologists alike. Notably, this allows to take seriously the economists' claim that their models improve our understanding of the world. But it also allows to downplay, when justified, their actual achievement.

1.2 FORETASTE

The main goal of this thesis is to examine and provide novel ideas on how to map out the epistemic contribution of theoretical models. The case of economics is interesting because of its peculiarity as a highly mathematized social science and because its empirical success is, to put it mildly, contested. Physics can always point to, among other things, the Voyager probes drift-

10 Note that I wrote '*re*construction', not 'constructing'. The purpose of philosophy is not foundational, science provides the foundations of science. Rather, it is to judge to what extent we should give credence to its results.

ing in space more than 15 billion kilometres away from earth to vouch for the helpfulness of their idealized models. Biologists can forecast the dynamics of populations of organisms and help save endangered species.

Economics does not have such clear-cut achievements. If anything, it might contribute to endangering said species. The recent—and looming—international financial crisis was a dreadful reminder that the models need some adjustments. Famously, even the Queen of England did not understand how possibly could the whole discipline miss the mark by that much (Stewart 2009). Economists were also divided in their solutions to the crisis. While some countries went for stimulus packages—and thus espousing a Keynesian perspective—, others went down the austerity route and cut down on public spending.[11] At any rate, economics should be less divided about some of the solutions it offers to the public. Here, I do want to say that there is no empirical success, simply that it is clearly less obvious than in other natural sciences.

Furthermore, the fact that economic models typically depict worlds populated with agents no sane person would like to have as friends casts doubt on their epistemic import. How possibly could one understand actual economic phenomena using these highly idealized models that so badly misrepresent the world we live in?

Although I have talked a lot about economics so far, I have to warn the reader that this thesis is not uniquely or particularly about economics. At least, not how one might expect. There are two reasons for this. First, many difficulties we have with appraising the epistemic contribution of models are not specific to economics. Models are pervasive across the sciences and whether or to what extent highly idealized models may explain is a general problem (Batterman 2009; Batterman and Rice 2014; Bokulich 2011; Graham Kennedy 2012; McCoy and Massimi 2018; Morrison 2015; Potochnik 2017; Rice 2015; Wayne 2011; Weisberg 2013). As I said above, some lines of defence are not readily available to economics, for instance empirical adequacy. However, economic modelling nevertheless shares common philosophical

11 It appears austerity was a mistake (Guajardo et al. 2014; Jordà and Taylor 2016; Ostry et al. 2016).

issues with other disciplines. Therefore, I developed my epistemological framework with generality in mind. Economics is the motivation, not the final destination.

Second and relatedly, some problems we face when assessing the contribution of specific models require taking a step back. This means questioning and examining the underlying epistemology. This is a strategy I pursue during much of the thesis. As a matter of fact, the most detailed case study—from economics—will only come at the end in chapter six. In a sense, the previous chapters lay the logically prior conceptual ground work. In practice, however, these previous chapters are heavily indebted to my study of the Arrow and Debreu (1954) model of general equilibrium. I partly lacked the language that would have helped me articulate better my analysis, so I took on myself to solve this. While looking at economics was instrumental in reaching some of my views, it was ultimately the more general epistemological framework that was in need of revision.

The five main chapters of this thesis were written as self-standing articles. A modified version of chapter 4 is forthcoming in the *European Journal for Philosophy of Science*. A slightly altered version of chapter 5 is forthcoming in *Studies in History and Philosophy of Science Part A* and a modified version of chapter 6 was published in the *Journal of Economic Methodology*. Even though they are therefore in principle independent of each other, they are united in their aim to contribute to the varied aspects of the same larger project. In the following I introduce the chapters as well as the issues they address and, hopefully, solve.

1.2.1 Chapter 2

Chapter 2 analyses what I call the inferentialist-behavioural account of understanding (InfBUn). Proponents of InfBUn suggest that understanding is constituted by having knowledge of relations of dependence and being able to use that knowledge to answer 'what-if-thing-had-been-different' questions (w-questions). InfBUn is behavioural because it considers that actual inferential performance is the key criterion for attributing understanding.

InfBUn, which is developed in a series of studies (see Ylikoski 2009; Kuorikoski and Lehtinen 2009; Ylikoski and Kuorikoski

2010; Kuorikoski 2011; Ylikoski 2013; Ylikoski and Aydinonat 2014; Kuorikoski and Ylikoski 2015), offers an insightful and extensive defence of the epistemic import of theoretical modelling. One reason why I devoted a whole chapter to it is because I think it is to a large extent on the right track. I do believe that knowledge of relations of counterfactual dependence and the ability to put that knowledge to use are closely related to understanding phenomena. And I also believe that models improve our understanding precisely by allowing to make more and better counterfactual inferences. What motivated me to pursue this analysis was not that I thought the account was wrong-headed. To the contrary, it was that I felt compelled to identify what were the potential weaknesses of the account and then look at ways it could be improved.

I consider that InfBUn identifies many correct aspects pertaining to understanding, namely that understanding can only be achieved when certain facts of the matter obtain.[12] I also agree that understanding demands a sort of cognitive grip, or grasp, that are not necessarily implied by common doxastic attitudes like knowledge. It is not because Bob knows that the windmill caused the polder to drain that he really understands what happened. He also needs to have some sort of grasp of why this is the case.

That said, drawing on literature in epistemology and philosophy of science, in this chapter I argue that InfBUn is inadequate. First, I show that the behavioural concept of understanding can't properly distinguish illusory from genuine cases of understanding. This is because it places excessive emphasis on actual inferential performance as a criterion for understanding rather than the ability proper. The ability to understand is best understood in dispositional terms and that is precisely what their behavioural notion inclines to reject. Second, I contend that it is not necessary to have a behavioural concept of understanding in order to retain compatibility with viewing models as extended cognition. Since this is one motivation for adopting a behavioural notion, I thus effectively defuse the two main reasons—distinguishing genuine

12 Whether understanding requires knowledge or simply true belief is a matter of contention (see Hills 2016). However, no one argues that mere belief is sufficient for understanding.

understanding from the sense of understanding and compatibility with extended cognition—to adopt a behavioural notion of understanding.

I the last section of this chapter I propose that InfBUn should be better viewed as an *evaluative* account of understanding and not as a *substantive* one. In short, I argue that while InfBUn may be lacking as a substantive epistemological account, it may nevertheless identify a relevant dimension according to which we should attribute understanding (see Wilkenfeld 2017).

1.2.2 Chapter 3

While the previous chapter looked at the inferentialist-behavioural account of *understanding*, this chapter examines what I call the factive inferentialist account of representation (FInfR). FInfR claims it is immune to the difficulties the traditional inferentialist account of representation (e.g. Suárez 2004) has with accounting for the explanatoriness of models. In particular, FInfR aims at showing how highly idealized models can provide explanations given that explanations are supposed to be factive. In a nutshell, FInfR holds that the correct 'what-if' inferential affordances of a model are a sufficient condition on establishing *explanatory* representation.

In the third section of the chapter, I argue that FInfR does not, in fact, allow to demarcate merely phenomenological models from explanatory ones. This is because of a crucial indeterminacy within FInfR. What are *correct* 'what-if' inferences (w-inferences) is either too ambiguous or too liberal. Basically, even if we grant that correct counterfactual inferential affordances are sufficient for (faithful) representation, not all counterfactual inferences are explanatory. For instance, predictively successful models will in a sense offer *w*-inferences. However, these models will not be explanatory.

I then expound a programmatic dilemma FInfR faces. The first option is to double down on deflationism, viz. dismissing any substantive criteria about successful representation. The second one is instead to substantiate under which conditions *w*-inferences are explanatory. I argue that the first horn is undesirable because it implies abandoning FInfR's main motivation, namely its commitment to the factivity of explanation and to realism. I then show how the second horn is preferable and discuss two possibilities. The first is to make a list of the sort of correct *w*-inferences they consider to be explanatory. For instance, FInfR could tie the notion of correct *w*-inferences along interventionist lines (Woodward 2003). The second possibility is to sever the link between explanation and understanding that FInfR assumes. While it is plausible that all *w*-inferences may afford understanding, as I argue they are not all explanatory. The benefit of this last line of defence is that it provides an adequate solution to the puzzle of model-based *understanding*, but at the cost of progress on the puzzle of model-based *explanation*.

1.2.3 Chapter 4

The fourth chapter examines two sets of views one can commonly find in the literature about the relationship between understanding and explanation:

1. Causal knowledge is necessary for understanding.
2. Only explanations can provide that knowledge.

Following a terminology I borrow from Pritchard (2014), I call the conjunction of these views, or tenets, the *narrow* knowledge account of understanding (KnUn). This chapter has three aims. The first is to show that narrow KnUn faces descriptive and normative issues because it can't account for scientific practices that do not actually explain (e.g., theoretical models that offer how-possibly explanations) and for those that do not explain causally (e.g., mathematical explanations). Narrow KnUn does not have the resources to describe these practices and has the undesirable normative conclusion of ruling out these practices as being able to afford understanding.

The second aim is to debunk the tenets of narrow KnUn. The strategy I employ is to show that significant parts of the existing literature already and implicitly imply that the tenets of narrow KnUn are false. In one section I observe that current and numerous accounts of *mathematical* explanations constitute counter-examples to the necessity of having causal knowledge for understanding. I consider objections, but ultimately reject them and come to the conclusion that this tenet of narrow KnUn is unwarranted.

In another section, I address a perhaps less evident issue, namely that actual explanations of phenomena are not necessary for understanding. To support my claim, I bring to the fore the literature on how-possibly explanations, 'explanations'[13] that that they do not explain actually, in the sense that they do not cite, for instance, what actually caused a given phenomenon. If how-possibly explanations may afford understanding, as we have good reasons to think they do, then we should also reject the second tenet of narrow KnUn.

The third aim is to propose an alternative to narrow KnUn, one that I coin 'broad' KnUn. Broad KnUn, contrary to its narrow sibling, makes the following two propositions:

1. Causal knowledge is not necessary for understanding.
2. Having an (actual) explanation is not necessary for understanding.

In particular, in the last section of this chapter I propose an account of broad KnUn that builds on Reutlinger's (2016) counterfactual theory of explanation. I show that by amending the veridicality condition of his theory in favour of what I call the *possibility* condition, we are left with an account of understanding. This account can accommodate, for instance, both mathematical and how-possibly explanations. My proposal for broad KnUn does not have the same descriptive and normative issues that narrow KnUn has.

Notwithstanding its contribution to our comprehension of the relationship between explanation and understanding and the specific proposal of broad KnUn that I make, this chapter plays a special role within the thesis. Indeed, the two propositions of broad KnUn lay the foundation for the rest of the dissertation. It serves as a bridge from the current state of the literature towards more specific proposals on HPEs and economic models.

13 As I explain in the chapter, whether HPEs are a species of explanation is an open question. The important point is simply to note that they do not explain actually.

1.2.4 Chapter 5

In chapter 5, I propose and develop a new account of HPEs. HPEs are usually contrasted to HAEs. HAEs provide explanatory accounts of why or how phenomena actually occurred. The concept of HPEs has a long history in the literature on scientific explanation. Yet, as I show, it is subject to various interpretations and other not always fortunate relabelling.

Using a distinction Strevens (2013) makes between the internal and the external conditions for explanatory correctness, I classify accounts of HPEs into two broad types: the Dray-type and the Hempel-type. Roughly, the Dray-type regards HPEs as a different species of the same genus 'explanation'. According to this type, HPEs are genuine and complete explanations in their own regard. HPEs differ from HAEs in that they do not have the same internal conditions; HAEs basically answer a different sort of questions than HPEs. The Hempel-type instead holds that HPEs diverge from HAEs not by their internal conditions, but by the external ones.[14] Hempel-type HPEs and HAEs have the same structure, but the former lack sufficient empirical confirmation. Were they (approximately) true, HPEs would be HAEs.

Considering the significant differences between the two types, I then show that accounting for the epistemic import of theoretical models by viewing them as HPEs can be misleading. For if we do not have a proper account of what HPEs are and can do, then the notion loses most of its philosophical usefulness. On top of that, I also show that neither the Dray nor the Hempel-type are adequate characterizations of models that offer HPEs. There is therefore both a philosophical and practical need for a novel account of HPEs.

This is precisely what the remainder of the chapter purports to do. Basing my proposal on what various commentators take HPEs to do—but without developing the idea—I suggest that the fundamental feature of HPEs is that they provide knowledge of possibility. In terms of internal conditions, I propose that HPEs have different ones than HAEs, the latter being best characterized as propositions of the form '*p* because *q*' whereas the former

14 Hempel-type HPEs technically are, as I explain in more details in the chapter, 'potential explanations' (Hempel 1965a).

include a possibility operator and thus have the form '$\Diamond$(p because q)'. This is different from the Dray-type in that it does not characterize HPEs as answering a different type of questions or as identifying necessary conditions for a phenomenon. It is also different from the Hempel-type in that it acknowledges that HPEs indeed have a different form and are not only false/not known to be true HAEs.

As for the external conditions, propositions of the form '$\Diamond$(p because q)' can have a truth value; they can be true or false. Evaluating their truth requires at least a minimal amount of relevant background knowledge. One has to assess the possibility of the explanans, of the explanandum, and of the explanatory relationship. Contrary to the Hempel-type, this implies that only satisfying the internal conditions is not sufficient to qualify as a HPE. A proposition is not a HPE only in virtue of its form. Contrary to variants of the Dray-type, my characterization does not require to identify any actual element, nor can it be satisfied without at least minimal empirical support.

In the last section I come back to the initial motivation of this chapter, namely to provide a characterization of HPEs that clarifies issues pertaining to the epistemic import of theoretical modelling. I argue that models should not be viewed stricto sensu *as* HPEs; rather, models provide reasons to believe propositions of the form '$\Diamond$(p because q)'. Models are therefore not explanations, but they enable or justify the beliefs we have in certain real-world propositions like HPEs and HAEs. Acknowledging this evidential role of models elucidates what is their relation with respect to explanatory propositions and how they could, in a sense, provide HPEs. However, what this does not settle is whether HPEs should be regarded as a genuine species of explanations. That said, this issue can be partly set aside since regardless of whether HPEs are explanations, there is strong evidence that they may grant similar epistemic benefits.

1.2.5 Chapter 6

This chapter is in a sense a culmination of what I discussed so far. It draws on many insights (account of understanding, broad KnUn, HPEs as possibility claims, etc.) and applies them to a specific case in economics, the Arrow and Debreu (1954) model of general equilibrium. It is interesting to look at the Arrow-Debreu model because it has been both very influential—it laid the foundations for much of economics going forward and earned each of its authors a Nobel prize—, but also very much criticized. From the economists' widely held judgment that the model improved our understanding of economic phenomena to some methodological appraisals (e.g. Blaug 2002) to the effect that the model may have *decreased* understanding, there is, to say the least, a slight discrepancy.

A potential solution to this problem is to draw on the literature on HPEs. Perhaps the Arrow-Debreu model is similar to other models (e.g., Schelling's checkerboard model (1978) or Banerjee's (1992) model of herd behaviour) and provides a HPE of general equilibrium. However, one difficulty with this interpretation is that the contribution of the Arrow-Debreu model is first and foremost a mathematical result. Whereas our current accounts of HPEs in economics regard models as providing some sort of causal knowledge, this can hardly be the case with Arrow-Debreu. The main claims of this chapter is that their model of general equilibrium provides a *mathematical* HPE and that this HPE affords understanding not only of the model, but also of the world.

To support these claims, I first present the historical background of the general equilibrium problem. I show that the Arrow-Debreu model provided a solution to a problem whose origin can be traced back to Smith's ([1776] 1904) famous invisible hand. The modern interpretation of the invisible hand hypothesis—the two fundamental theorems of welfare economics—indicate that a competitive equilibrium is Pareto optimal. However, whether such an equilibrium could actually exist was a separate but related question. While economists trusted the equilibrium could exist, there was no conclusive evidence, not even Walras's (1954) model.

I then argue for the main claim of this chapter, viz. that the Arrow-Debreu model is a mathematical HPE that affords understanding of the world. I first present some defining features of mathematical explanations and show what sort of dependence is involved in these cases. I then submit that the Arrow-Debreu model is not an actual mathematical explanation, but that it instead provides a mathematical HPE. In particular, I argue that even though the model can in principle receive a causal interpretation, the how-possible problem economists were trying to tackle was properly mathematical.

In the final section, I address the crucial issue of whether the Arrow-Debreu model affords understanding not only of the model, but also of the world. To do so, I draw an analogy with causal HPEs as discussed by Ylikoski and Aydinonat (2014). After presenting the distinction between understanding the world and the model, I show that proving the existence of the general equilibrium was motivated by the fact that it could inform economists, albeit minimally, about its actual existence. This is because mathematical claims can serve an evidential role with respect to causal claims. For if the equilibrium would be mathematically impossible, then this would be evidence that it is causally impossible. Therefore, there is an important sense in which the Arrow-Debreu model is also an inquiry about the world.

Since this is an important aspect of the account of model-based understanding that I use, I finish this section by showing what sort of *w*-questions the Arrow-Debreu model allows to answer. By establishing relations of mathematical dependence between certain assumptions and the theorems of existence, the model constrains the economic space of possibility. We can infer that if some conditions were not satisfied, then there would be no general equilibrium. This is even truer if we consider the cluster (see Rodrik 2015; Ylikoski and Aydinonat 2014) of models that use similar assumptions and that test their robustness.

I end by defusing two objections: 1) that since the model fail to identify robust conditions, then it can't afford understanding and 2) that any mathematical model could afford understanding, which is an undesirable conclusion. I respectively argue that failure of robustness may be a significant achievement in that it shows what conditions should be satisfied for the general equilibrium to actually exist and that not all mathematical HPEs are born equal, some fail to establish relevant relations of dependence.

1.3 FINAL PROLEGOMENA

One general implication of my research is that understanding is an epistemic benefit that can be had in the absence of an actual explanation and that, as a consequence, it has less demanding requirements than explanation. Have I made economists' and other scientists' life too easy by providing them with a framework within which most of what that they do could be viewed as affording understanding? I do not think so for three reasons.

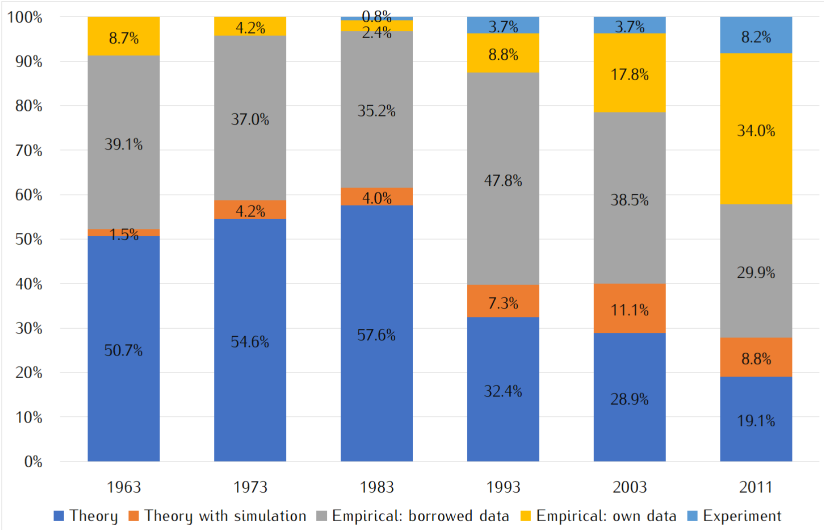
The first is that while I believe there are good arguments to the effect that we need to be more liberal in our attributions of understanding, of course not anything goes. We need criteria to distinguish genuine from illusory understanding, spurious science from fruitful one. I have explicitly discussed a form or another of this objection in all the chapters. As I see it, research on the boundaries of understanding has just begun. Perhaps I have been too generous, but one conclusion I am sure of is that understanding is more prevalent than some of our philosophical accounts would have us believe.

The second reason is that nothing that I said implies that what Northcott and Alexandrova (2014) call an 'efficiency analysis' can't or should not be done. This analysis consists in asking whether particular exercises of theoretical modelling pay off in terms of solving actual and important problems. That is indeed a very important question and one about which my research is agnostic. As far as economics is concerned, figure 1.1 shows that there is already an important trend towards more empirical work. The majority of resources appear to be put on approaches that deal directly with the collection and analysis of data.[15]

That said, whether there is still too much theoretical modelling in economics or other sciences is of course an open question. It may indeed be the case that, for instance, too much energy, time, and money was invested in proving the existence of an equilibrium. Perhaps the benefits in terms of understanding were not worth the costs. As in everything, it is likely there are bad investments in understanding. Having said that, this question is orthogonal to the problem of formulating an epistemological

15 This is even more the case if we assume that empirical research is more expensive than theoretical work.

Figure 1.1: Changing methodology of economics



Source: Hamermesh (2013)

framework that allows to conceptualize in a fruitful manner the epistemic benefits models may give. The task I undertook was to work on such a framework, but it is of course important that we carry further research on the efficiency of theoretical modelling. I do not consider that my work overestimates the benefits and conceals the costs of modelling. Rather, it simply allows for a more fine-grained assessment.

The third reason is that I believe that a new account of the relationship between understanding and explanation allows us to assess better the epistemic contribution of various epistemic practices. Traditionally, the problem of the epistemic contribution of models was like a switch, on or off. Either models explain and thus afford understanding, or they do not. But since "understanding admits of degrees" (Elgin 2007, p. 36), we need to be able to distinguish cases where models afford us limited understanding to those where it is more deeper or broader. In particular, this raises the question about where, on the spectrum of understanding, explanation begins. For instance, economists appear to mix-up providing *actual* explanations with providing HPEs (Grüne-Yanoff 2013b). So they may be right that their models afford understanding, but they may be too hasty in claiming they explain. And if explanations are the ideal of understanding (Khalifa 2013b), it may show that economics often fall short of that ideal. Economic models may afford understanding, but ultimately what we expect from them is that they provide explanations.

Opening the door of understanding does not imply we can't judge its quantity or quality. It just means we need to roll up our philosophical sleeves and figure out what is understanding with models.

4 UNDERSTANDING DOES NOT DEPEND ON (CAUSAL) EXPLANATION

4.1 INTRODUCTION

The epistemology of understanding has traditionally been related, if not reduced, to the epistemology of causal explanation. A prominent view on the epistemology of scientific understanding (understanding hereafter) and explanation submits that one has understanding of a phenomenon only if one has an explanation of it. Furthermore, explanations are usually taken to be of a specific sort, namely *causal*. Causal explanations provide understanding in virtue of the causal knowledge they provide. I characterize the combination of these two views as the narrow knowledge account of understanding (narrow KnUn). Narrow KnUn has two tenets: 1) causal knowledge is necessary for understanding and 2) only explanations can provide that knowledge.

I show that these tenets are descriptively and normatively inadequate because they can't account for scientific practices that do not actually explain (e.g., theoretical models that offer HPEs) and they can't account for non-causal explanations (e.g., mathematical explanations). I argue for a broadening of narrow KnUn, the *broad* knowledge account of understanding (broad KnUn), which does not face these problems.

In section 4.2 I give a brief overview of narrow KnUn. Section 4.3 spells out the descriptive and normative issues narrow KnUn gives rise to. I then argue in section 4.4 that causal knowledge is not necessary for understanding and in section 4.5 that having an explanation is also not necessary. In section 4.6 I present an alternative, broad KnUn.

4.2 THE TWO TENETS OF NARROW KNUN

One can find two sets of views concerning the relationship between understanding and explanation in the literature, viz. that:

1. Causal knowledge is necessary for understanding.
2. Only explanations can provide that knowledge.

Proponents of the latter set consider that explanations are the only legitimate source of understanding, for instance:

> The resulting objectivist, ontic, account, in generic form, states that scientific understanding is the state produced, and only produced, by grasping a true explanation (Trout 2007, pp. 584-585).
>
> [U]nderstanding amounts to (*a*) knowing that the explanans is true, (*b*) knowing that the explanandum is true, and (*c*) for some *l*, knowing that *l* is the correct explanatory link between the explanans and the explanandum (Khalifa 2012, p. 26).
>
> An individual has scientific understanding of a phenomenon just in case they grasp a correct scientific explanation of that phenomenon (Strevens 2013, p. 510).[1]

In a nutshell, what these accounts say is that one can't understand without an explanation. According to Trout, it is important to separate the sense of understanding, which can be a misleading phenomenology, from the genuine understanding one obtains when being in possession of a true explanation. Khalifa maintains that a distinct concept of understanding adds nothing to what he calls the "Explanatory Model of Understanding" (EMU). For Khalifa, having scientific understanding is a matter of having explanatory knowledge. Similarly, Strevens argues that what he calls the "simple view" adequately depicts the connection between explanation and understanding. For him, the epistemology of explanation precedes and guides the epistemology of understanding. One understands *why* something is the case, according

1 See also Strevens (2008, p. 3).

to Strevens, when one not only grasps a state of affairs, but also its correct explanation.

It is important to bear in mind that Strevens and Trout do not claim that knowing an explanation is *sufficient* for understanding.[2] Indeed, the 'grasping' condition may require a state or ability on top of knowledge. This is why Strevens (2013, p. 510) does not reduce understanding to explanation. However, grasping can be related to knowing. Indeed, as Strevens (2013, fn. 6) indicates, his account is compatible with the view that knowledge is necessary for grasping. Trout is not as explicit as Strevens, but nevertheless suggests that one might "treat grasping as a kind of knowing" (2007, p. 585). In short, perhaps knowing an explanation is not sufficient—one may need to grasp it—, but to grasp a true explanation one may need to know it.[3] That said, the important point is that one needs to stand in the appropriate epistemic relation—e.g., knowing, grasping, believing, etc.—with an explanation to understand. Explanations are the bearer of the information that affords understanding and without an explanation there is no understanding.

Trout's, Khalifa's, and Strevens's accounts thus all assert that an explanation is necessary for understanding. If this is true, one may then ask what *sort* of explanation may provide understanding. In point of fact, one characteristic that competing theories of explanation share is their lack of agreement over what affords understanding. For this reason, discussions about the nature of scientific explanation have largely focused on what information affords understanding: e.g., for Hempel (1965a) it is nomic expectability (given by laws of nature), for Salmon (1984) it is information about the causal history of a phenomenon, for Michael Friedman (1974) and Kitcher (1981) it is unification, and for Woodward (2003) it is patterns of causal counterfactual dependence. These accounts started, in a sense, with the basic notion

2 Khalifa (2013c) argues that the crucial cognitive ability involved in understanding, which we may call grasping, is the ability to conduct a reliable explanatory evaluation. See also Khalifa and Gadomski (2013).

3 There is also a debate over whether understanding, unlike knowledge, can be 'lucky' (see, e.g, Hills 2016; Khalifa 2013c; Morris 2012; van Riel 2016; Rohwer 2014). More minimally, we could say that what is required is to have true beliefs about the explanation. Nothing here hinges on how this question is settled.

of understanding and then used it to explicate their notion of explanation.

Despite the existence of alternative theories of explanation (e.g. Achinstein 1983; Batterman 2002; van Fraassen 1980; Kitcher 1981; Ruben 1990), the causal account of explanation has emerged as the currently dominant view (Lange 2017; Potochnik 2015). As Reutlinger and Saatsi (2018b, p. 2) observe, "[t]he state of the field after six long decades suggests that something close to a consensus was reached: scientific explanation is a matter of providing suitable information about causes of the explanandum phenomenon". As such, accounts of causal explanation received tremendous philosophical attention and analyses of causation have been a major topic of philosophical research. On the contrary, "non-causal scientific explanations have been largely neglected by philosophy of science" (Lange 2017, p. xii). One can find many proponents of the causal account of explanation in the literature, for instance:

> Here is my main thesis: *to explain an event is to provide some information about its causal history* (Lewis 1986, 217, emphasis in original).

> Scientific explanations, from which such understanding derives, are often, if not always, causal (Salmon 1998, 2, emphasis in original).

> Understanding is not some sort of super-knowledge, but simply more knowledge: knowledge of causes (Lipton 2004, p. 30).

As maintained by the causal account of explanation, the nature of the information that affords understanding is causal. Put differently, a causal explanation is one that provides information about the causes, processes, or mechanisms that bring about the phenomenon to be explained. Pritchard (2014), from whom I borrow the terminology, calls this underlying epistemology of understanding the 'knowledge account of understanding' (KnUn). In short, "the idea is that understanding is essentially a type of knowledge–*viz.*, knowledge of causes" (Pritchard 2014, p. 315). A consequence of the dominance of the causal account of explanation is that having causal knowledge is de facto viewed as necessary for understanding. Since causal explanations provide this knowledge, knowing a causal explanation thus becomes *the* pathway to understanding.

According to the first set of views discussed, having an explanation is necessary for understanding. According to the second set of views, causal knowledge is necessary for understanding. These are the two tenets of what I call *narrow* KnUn. According to narrow KnUn, one has understanding only if:

1. One has knowledge of causes;
2. That knowledge is provided by an explanation.

As shown, narrow KnUn incorporates two common views about the relationship between understanding and explanation. The qualifier 'narrow' emphasizes that within narrow KnUn only knowledge of causes is a source of understanding and that only explanations can provide that knowledge. In other words, it is narrow because it states that knowing a *causal explanation* is a necessary condition for understanding. Of course, each of these tenets can in principle be held independently. For instance, one can consider that having an explanation is necessary and be a pluralist about the sorts of explanations there are (e.g. Khalifa 2012). In fact, as we will see, they are also not equally controversial. Hence, narrow KnUn is not so much an actual thesis than a useful characterization of widespread—at least taken individually—positions in the literature on explanation and understanding. Using it as a foil allows to assess better the costs and benefits of holding these tenets, conjointly or individually. Furthermore, it also allows to see in which direction we need to amend our current accounts to relax the unduly restrictive nature of narrow KnUn.

4.3 NARROW KNUN: ITS LIMITS AND DIAGNOSIS

Holding narrow KnUn does not come without costs. A significant one is that it has difficulty making sense of the epistemic value of actual scientific practices. Debates on scientific representation, idealizations, and the explanatoriness of models are striking evidence that our epistemology of understanding suffers from important lacunae. For if it could straightforwardly accommodate how science is conducted, then these debates could be settled much more easily. As a matter of fact, an important motivation to revise the relationship between explanation and understanding is precisely this difficulty of reconciling our epistemology of understanding with actual science. Narrow KnUn is descriptively and normatively inadequate. Descriptively, narrow KnUn can't account for scientific practices that:

1. Do not provide explanations, yet provide causal knowledge.
2. Provide explanations, but not causal ones.

Normatively, it has the undesirable implication of ruling out large parts of science as improving our understanding.

Firstly, there are cases in which scientists cite causes that are not known to actually obtain, but where they nevertheless consider citing them to provide understanding. Following common usage, these 'causes' do not explain because they are not the *actual* causes of a given phenomenon. Yet, as we will see in section 5, scientists think that they afford understanding. Not citing actual causes therefore does not appear to be necessary for understanding. Scientists are regularly engaged in activities, theoretical modelling being a prime example, that they consider to afford understanding, but do not explain. Thus, from the perspective of narrow KnUn, it is rather unclear—if not outright inconceivable—how highly idealized models could improve our understanding of the world. For if knowledge of (actual) causes is necessary to explain a phenomenon, and if models do not always identify actual causes, then how could they afford understanding? Narrow KnUn has no room for practices whose goal is not to provide causal knowledge of that sort.

Secondly, narrow KnUn can't account for explanations of the non-causal sort. Analyses of scientific understanding have put emphasis on knowledge of causes since providing this knowledge is a hallmark of what science does. However, as section 4.4 will show, there are cases of non-causal explanations, for example mathematical explanations. Since they do not provide causal knowledge, narrow KnUn would therefore commit us to conclude that mathematical explanations are not, in fact, explanations and for this reason can't afford understanding of empirical phenomena. However, the point of contention in debates over mathematical explanations is not so much whether they actually explain empirical phenomena—commentators and practitioners alike reckon they do—but rather, how we could account for them? Narrow KnUn is of no help here.

In consequence, if we give due credit to the cases I will present in the following sections, then narrow KnUn can't provide a normative account of science. Due to a lack of conceptual resources, it appraises actual scientific practices—e.g., mathematical explanations and theoretical modelling—as not affording understanding. One important desideratum of an account of understanding is that it provides evaluative criteria for assessing when understanding has been achieved. The sense of understanding (see Trout 2002; Ylikoski 2009) can be a misleading cue for genuine understanding, the state of actually understanding. We want to be able to demarcate cases of genuine understanding from illusory ones. As such, it must be recognized that *mis*understanding is a possibility.

However, another important desideratum is that the account does not rule out entire areas of science as not being conducive to understanding. A naturalistic outlook on science should compel philosophers to not attribute systematic and persistent error across different fields of science. While narrow KnUn scores very well on the first desideratum, it does poorly on the second. It provides clear and explicit criteria for judging whether understanding is genuine, but at the cost of making illusory understanding dubiously prevalent. Indeed, one implication of narrow KnUn is that we have to conclude that a very important part of what scientists do does not afford understanding because it either does not provide causal knowledge or does not provide an explanation.

An account of understanding with such implication is suspect. The problem is that causal knowledge and explanations are simply very hard to obtain. Restricting understanding only to cases where one has a causal explanation appears to place unduly constraints on what counts as genuine understanding. Perhaps having knowledge of causes can be seen as an ideal (Khalifa 2013b), but surely there must be some epistemic benefits in the form of understanding for results that fall short of that achievement. In other words, achieving something other than knowledge of causes or of an explanation should not necessarily entail a complete lack of understanding. We may sometimes have good reasons to doubt that scientists are overly optimistic concerning some of their results. However, the fact that our epistemology of understanding would rule out a large part of what scientists do as not promoting understanding is still an unwelcome conclusion.

Having identified two problems with narrow KnUn, a question remains: what features of narrow KnUn generate these problems? As we have seen, narrow KnUn states that only causal knowledge provided by an explanation can afford understanding. Narrow KnUn has these problems precisely because it:

1. Restricts the type of knowledge to knowledge of causes.
2. Restricts the source of (causal) knowledge to explanations.

In other words, narrow KnUn considers that having a causal explanation is necessary for understanding. We need an epistemology that is both stringent, but also in line with how science is actually conducted.

Now that we have the diagnosis, what is the treatment? I claim that narrow KnUn needs to be *broader*. Consequently, I will present a version of what I call 'broad KnUn'. Broad KnUn contradicts narrow KnUn in two ways. It asserts that, for understanding:

1. Causal knowledge is not necessary;
2. Having an (actual) explanation is not necessary.

In other words, broad KnUn broadens the knowledge—i.e., not only knowledge of causes—that affords understanding and broadens the ways this knowledge can be acquired.[4]

4 Whether narrow KnUn states sufficient conditions for understanding is a different question. While it appears prima facie plausible, here I remain agnostic over this issue.

What I think a better fleshed out notion of understanding can do is precisely to help us solving these problems in two different ways. Firstly, it allows for the possibility of such a middle ground between non-explanatoriness and understanding. Secondly, it helps to address the assessment problems above by spelling out more clearly what is the nature of the benefits provided. It also gives space for non-causal approaches to understanding.

One can already find philosophical positions that, as we shall see, are in agreement with broad KnUn. Broad KnUn, as I will show later on, combines existing philosophical insights. Therefore, broad KnUn is not as controversial as it may seem. Rather, it is in a sense the natural development of the latest work on explanation and understanding. In the following sections, I will show how some current positions already presuppose broad KnUn and then propose an epistemology based on them.

4.4 IS CAUSAL KNOWLEDGE NECESSARY?

The first tenet of narrow KnUn is that causal knowledge is necessary for understanding. Even though it has been the de facto position in philosophy of gscience for the past decades (Reutlinger 2017b), it is also the more controversial of narrow KnUn's two tenets. Indeed, there is now a burgeoning literature on non-causal explanation (e.g. A. Baker 2012; Baron et al. 2017; Batterman 2002; Batterman and Rice 2014; Lange 2013, 2017; Pincock 2015; Reutlinger and Saatsi 2018a**Saatsi:2016aa**) whose aim is to show that there are cases of explanations which can't count as being properly causal. The literature thus already proposes many counterexamples to narrow KnUn's first tenet.

Even current prominent accounts of causal explanation are now more liberal. For instance, despite the fact that their main focus is on causal explanation, Woodward (2003, pp. 220-221) and Strevens (2008, sec. 5.7) briefly indicate that the criteria they propose for explanatory relevance could also be applied to non-causal explanations.[5] There is thus a prima facie strong case for rejecting that causal knowledge is necessary for understanding.

5 Woodward's insight, in particular, has recently received some attention (e.g. Grimm 2010; Pincock 2015; Reutlinger 2016).

In fact, the case for the "liberal consensus" (Reutlinger 2016) is so strong that I will simply rehash important aspects of the debate and consider some objections.

An important class of plausible non-causal explanations are what Lange (2013) calls 'distinctively mathematical explanations', i.e., explanations of physical phenomena that can properly be seen as mathematical, but not because they employ mathematics. Many causal explanations are formulated using mathematical expressions, but distinctively mathematical explanations are explanatory because of the mathematical facts they cite. These explanations have received a lot of philosophical attention lately (e.g. Baron 2016; Baron et al. 2017; Batterman 2010; Bueno and Colyvan 2011; Lange 2013; Pincock 2015; Saatsi 2011). Even though there are still debates over how exactly these explanations work, there is clear support for the idea that science sometimes explain physical phenomena with mathematical facts. And if this is the case, then distinctively mathematical explanations are a direct counterexample to the first tenet of narrow KnUn.

One simple but convincing example is the bridges of Königsberg case (Lange 2013; Pincock 2007; Reutlinger 2016). Leonhard Euler (1741) provided an explanation of why it is impossible to cross all seven bridges of Königsberg exactly once through a single path. He proved that the mathematical structure instantiated by the bridges, a graph, made it impossible because:

1. not all parts of Königsberg were connected to an even number of bridges, and;

2. nor were exactly two parts connected to an odd number of bridges.

In other words, there was no Eulerian path—a path that visits every edge in a graph exactly once—one could follow to achieve the desired result. It is a mathematical impossibility. In virtue of the isomorphism between the actual layout of the bridges in Königsberg and the mathematical structure, Euler's proof showed on what the (im)possibility of crossing the bridges only once depended. The Königsberg example is a simple and intuitive case where mathematical relations of dependence carry explanatory work. Knowing the mathematical facts established by Euler plus the particulars of Königsberg afford understanding despite the absence of causal knowledge.

Different interpretations of why this constitutes an explanation can be given. Lange (2013) argues that it explains because it shows, in a modally stronger sense than causal impossibility, why the explanandum is the case. The specific causal structure instantiated does not make a difference to the explanandum the way the mathematical facts do. These facts necessitate the explanandum in a stronger way than the causal factors. The mathematical explanation goes beyond contingent causal facts about Königsberg. As for Reutlinger (2016), he argues that the generalization used by Euler is mathematical and intuitively non-causal. That generalization, conjoined with details about the arrangement, entails the explanandum, i.e., the impossibility to cross all the bridges once. The generalization, Reutlinger says, also supports counterfactuals. Indeed, we know that if all parts of Königsberg were connected to an even number of bridges, then there would be an Eulerian path. Regardless of which interpretation is correct between Lange and Reutlinger, there is agreement that there is a class of explanations that does not appeal to causal knowledge and which should be considered mathematical.

Proponents of narrow KnUn can respond to cases of distinctively mathematical explanations in two ways. First, they can deny that non-causal generalizations may afford understanding. They can do so by, for example, challenging that these cases are genuine explanations. Second, they may simply retort that what appear to be non-causal generalizations are, in fact, causal. Both avenues are implausible.

The first objection is the simplest to counter. First, there is very few opposition that these cases represent genuine instances of explanations (cf. Strevens 2018). Second, denying that they explain and afford understanding may oppose practitioners' judgments of explanatoriness (Pincock 2015). Third, as I argue in more detail in section 6.5.1, it is plausible that they indeed do explain. The question is thus not so much *if* they explain, but in virtue of what.

The second objection is the most serious. One could argue, as Strevens (2018) does, that the mathematical proof within distinctively mathematical explanations only serves to grasp better the real causal difference-makers, namely the physical setup. For him, the mathematics illuminates the reasons why the differences-makers effectively make a difference. Euler's proof, for example, would help in grasping the explanation, but would not explain itself. The explanation, according to Strevens, is causal, not mathematical.

There is certainly a sense in which the causal facts act as difference-makers. Had all parts of Königsberg been connected to an even number of bridges, an Eulerian path would have been possible. One could then go on and build bridges so as to make the walk possible. But if the mathematical proof is necessary to grasp which physical details matter, why should the latter receive all the explanatory credit? By citing the causal facts and omitting the mathematical proof, we would not gain understanding. Whereas causal facts are sufficient for understanding in ordinary causal explanations, they are not in the case of distinctively mathematical explanations. It is only because we grasp the mathematical facts that we can understand why the physical set up matters in the particular way that it does. Ultimately, it is thus not the causal facts that afford understanding. If this is the case, then there is really a class of explanations that do not appeal to causal knowledge to afford understanding. Rather, the information that conveys understanding is a sort of non-causal dependence.

But even if we accept that Strevens is right—despite the evidence to the contrary—concerning cases such as the Königsberg bridges, we would still need an argument for why distinctively mathematical explanations are *in principle* always causal. However, there is no such argument. At the very least, these examples put the burden of the proof on the advocates of narrow KnUn to argue why we can't regard them as mathematical explanations.

In light of cases of distinctively mathematical explanations, we have to conclude that the first tenet of narrow KnUn, namely that causal knowledge is necessary for understanding, is unwarranted.

4.5 IS EXPLANATION NECESSARY?

If the first tenet of narrow KnUn is more controversial, the second, namely that one needs to have an explanation to understand *why*, is more widely accepted. Under this view, explanations are the sole providers of knowledge that afford understanding. An explanation is essentially just a set of propositions that connects an explanans to an explanandum in the right way (Strevens 2013). While it is uncontroversial that explanations provide the right kind of propositions and structure, what reasons do we have to believe that non-explanatory propositions conveying suitable information cannot provide understanding?

The strategy I follow here is very similar to Lipton's (2009).[6] According to him, understanding should be identified with the benefits that explanations provide rather than with the explanations themselves. His strategy is thus to identify ways of getting the same type of information that explanations typically provide without passing through an explanation. In a similar fashion, I would like to show that the kind of benefits that constitute understanding—e.g., having causal knowledge—can be acquired without having an explanation.

One important source of knowledge conducive to understanding phenomena is referred to in the philosophy of science literature as HPEs (Bokulich 2014; Forber 2010; Grüne-Yanoff 2013b,a; Rohwer and Rice 2013; Ylikoski and Aydinonat 2014). A terminological disclaimer is now essential. As their names suggest, it may seem that HPEs and HAEs are simply different species of the same genus 'explanation'. If it is the case, then arguing that having an explanation is not necessary because HPEs may afford understanding would be misleading. Having an HPE would amount to having an explanation and thus it would say nothing about the necessity of explanation for understanding. So we first need to find out whether HPEs are explanations or not.

Without entering too much in the details since chapter 5 discusses HPEs at length, HPEs are to be contrasted with HAEs. HAEs give an account of how phenomena actually occurred.

6 Strevens (2013) explicitly disagrees with some of Lipton's examples, but says nothing about the general approach. See Khalifa (2013b) for an in-depth analysis of Lipton's strategy, its success, and limitations.

HPEs have the form of HAEs, but do not provide the same sort of actual accounts. In other words, HPEs satisfy internal conditions of adequacy—the explanation's structure—whereas HAEs satisfy both internal and external—correspondence to the world—ones (Strevens 2013).[7] For instance, I could, using phlogiston theory, provide an internally correct explanation of the phenomenon of combustion. However, the theory does not actually explain combustion because there is no such entity as phlogiston. The explanation is false and does not meet the external conditions of adequacy.[8] To give another example, Ptolemaic astronomy provided an internally correct explanation of the motion of the planets using a geocentric cosmology and epicycles. But since the theory is false because it depicts, among other things, the earth at the centre of the solar system and planets as moving along epicycles, it is not externally correct and thus not a HAE.

For the purpose of my argument, what is important to bear in mind is that what narrow KnUn requires for understanding is to have a HAE, not merely a HPE. The textual evidence of the previous section makes this plain. Strevens explicitly denies that potential explanations—or HPEs—afford understanding of phenomena. He indicates that grasping a correct explanation "requires grasping that the propositions expressing a relevant model's explanatory content are true, or in other words, understanding that the states of affairs represented by those propositions obtain" (2013, p. 512). This requirement is clearly not satisfied by most cases of HPEs discussed in the literature since they depict either false—or not known to be true—explanantia or explananda. Their explanatory content is false and therefore

7 What I say here is in principle orthogonal to the scientific realism debate. As Khalifa (2011) argues, to say that explanation is necessary for understanding does not imply the factivity of the explanans. It only requires the explanation to be correct according to some metric, e.g., empirical adequacy. A HPE thus can simply be an explanation that is not correct or not known to be correct. But, of course, many accounts require the (approximate) truth of the explanans for understanding (e.g. Strevens 2013; Trout 2007; Woodward 2003).

8 There is also a distinction to be made between understanding *with* a theory or model, also sometimes called objectual understanding (see Khalifa 2013a; Kvanvig 2003), and understanding *why*. We are here only concerned with the latter. It is thus possible to understand combustion *with* phlogiston theory while not understanding *why*. To understand *why* some externals conditions of adequacy need to be fulfilled.

one can't, according to narrow KnUn, reap understanding from them. The issue therefore does not hinge on whether HPEs are a species of explanation. The issue is rather whether a HAE is necessary for understanding. According to proponents of the second tenet of narrow KnUn, it is.

But if HPEs may afford understanding, as the literature suggests, then it would imply that having a HAEs is not necessary, notwithstanding narrow KnUn. That HAEs are not necessary for understanding is precisely what contemporary discussions of scientific modelling show. The philosophical challenge is to make sense of a widespread practice—theoretical modelling—across the sciences that seemingly afford understanding without explaining in the HAE sense. Scholars took up the challenge and developed accounts of modelling as HPEs in response.

An oft-discussed example in the literature is Schelling's (1971, 1978) checkerboard model of residential segregation. According to Sugden, an economist himself, the checkerboard model "tells us something important and true about the real world" (2000, p. 2). W. A. V. Clark and Fossett (2008, p. 4109), both social scientists, consider the model "was critical in providing a theoretical basis for viewing residential preferences as relevant to understanding the ethnic patterns observed in metropolitan areas". Prior to the checkerboard model, social scientists believed that only strong discriminatory preferences–i.e., racism–could lead to residential segregation (Aydinonat 2007; W. A. V. Clark and Fossett 2008; Grüne-Yanoff 2009; Sugden 2000). The model showed that it was possible that preferences for not being in a minority status could also produce the same pattern of segregation, a result that has proven to be very robust across changes of assumptions (Muldoon et al. 2012).

The checkerboard model is often interpreted as having provided a HPE of residential segregation (e.g. Grüne-Yanoff 2013a; Weisberg 2013; Ylikoski and Aydinonat 2014). The model does not make any specific claim about the actual mechanism producing instances of residential segregation. More precisely, it is *not* a HAE of segregation since we do not know whether it explains any actual instance of that phenomenon. Instead, it answers a general how-possibly question, namely "how is it possible for segregation to happen in a city without collective preferences for segregation?" (Weisberg 2013, pp. 118-119).

Even though the model represents phenomena in a highly stylized manner and even though the mechanism it depicts is not known to be actual, it still 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 (Grüne-Yanoff 2009; Reiss 2008; Ylikoski and Aydinonat 2014).

Knowing that some causal factors may bring about residential segregation improves our understanding of the phenomenon even though we do not know what actually causes it. It may do it in various ways. It can expand our 'menu' of possible explanations (Ylikoski and Aydinonat 2014). It can also license *w*-inferences about phenomena (Kuorikoski and Ylikoski 2015; Ylikoski and Aydinonat 2014). Or it can contradict impossibility theses people hold (Grüne-Yanoff 2009). All these accounts suggest the checkerboard model, read as providing a HPE, can afford understanding or can teach us about real world phenomena.

Another widely discussed example in this literature on HPEs (e.g. Sugden 2011; Rohwer and Rice 2013; Rice 2016), this time at the intersection of biology and economics, is the use of game theory models of animal competition (e.g. Maynard Smith and Price 1973; Maynard Smith and G. A. Parker 1976). These models helped investigate the phenomenon of restraint in animal combat. For instance, it may be expected that individual members of the same species would develop weapons or strategies and fight to death in order to gain selective advantages like mates or resources. However, that is not what happens. Rather, individuals typically display restraint and solve their conflicts conventionally, viz. without significant cost to the participants. This appeared to go against individual selection. Group selection was thus believed to be the adequate explanation. For if it may not be in the interest of the individual to show restraint, it certainly is beneficial for the species if its members are not regularly gravely injured. Maynard Smith and Price were interested, using a computer model, to examine "whether it is possible *even in theory* for individual selection to account for 'limited war' behaviour" (Maynard Smith and Price 1973, p. 15). They concluded that their analysis is "sufficient to show that individual selection can explain why potentially dangerous offensive weapons are rarely used in intraspecific contests [...]" (1973, p. 17). In other words, the model shows that restraint in contest between individuals of the same species benefit not only the species as a whole (the group), but also the individuals. Contrary to what was expected, a 'limited war' strategy can also benefit individuals' fitness. To explain that behaviour, it is not necessary to resort to group selection. Individual selection is sufficient.

One interpretation of that model is that it "produces some understanding of how individual selection could possibly lead to restraint in situations of animal conflict" (Rohwer and Rice 2013, p. 341). As such it aims to answer a how-possibly question, namely how individual selection could bring about restraint in combat. Commenting on this type of modelling, Maynard Smith (1978, p. 52) said that "[t]he role of optimization theorizing in biology is not to demonstrate that organisms optimize. Rather, they are an attempt to understand the diversity of life". In a similar vein, biologists Arnott and Elwood (2008, p. 529) note that our "understanding of this variation [of forms of contests for resources] was boosted by the application of game theory (Maynard Smith and Price 1973; G. A. Parker 1974), which examined how different strategies might be used by each contestant and how the winner is determined". Even though they do not satisfy the typical empirical external conditions associated with HAEs, both the judgement of practitioners and philosophical analyses of specific cases of theoretical modelling lead to the conclusion that HPEs can provide understanding. Proponents of the second tenet of narrow KnUn could dispute this conclusion on two grounds. They could contend that HPEs are, in fact, explanations in the required sense of narrow KnUn. Or, they could deny that HPEs can afford understanding. However, neither horn of the dilemma is readily available to them.

Firstly, it is implausible to regard all HPEs as being HAEs. Reydon (2012), for instance, argues that what Forber (2010) calls global HPEs are actually genuine explanations of type-level phenomena.[9] Since the point of narrow KnUn is precisely that *only* HAEs can provide understanding, this would be a successful way of defusing the claim that HPEs can afford understanding. However, the fact that some HPEs should perhaps rather be considered as HAEs does not exclude that others are genuine HPEs. As I already pointed out, Strevens (2013) stresses that an explanation must be externally correct, that is, it must contain a true explanans, in order to afford understanding. Reutlinger's (2016) veridicality condition requires that both the explanandum and the explanans must be true in order for the relationship to be explanatory. HPEs, by definition, do not satisfy these criteria because either the explanans or the explanandum is false or not known to be true. Furthermore, there are clearly cases—e.g., the ones discussed above—that do not satisfy them. Insofar as empirical support is lacking for the explanans, models similar to the checkerboard or the Hawk-Dove examples can't qualify as HAEs. So even though some HPEs could better be viewed as HAEs, not all can. For this reason, narrow KnUn can't account for them.

Secondly, if one rejects the claim that HPEs afford understanding, this would imply that exemplary cases of theoretical modelling are epistemically suspect. That many contemporary philosophical accounts as well as practitioners' hold that HPEs afford understanding is strong evidence that they actually do so. Insofar as practice is correctly described, the burden of proof should be on those philosophical accounts that want to deny HPEs can afford understanding, not on practitioners.

In a related but separate discussion, Fumagalli (2016) argues that both the checkerboard and the Hawk-Dove models, interpreted as 'minimal models'—that is, models that supposedly lack any representational features—can't justify a change of confidence in necessity or impossibility theses (see Grüne-Yanoff 2009). The checkerboard model, for instance, affects our confidence in the thesis that only strong discriminatory preferences can bring about residential segregation. Fumagalli may be right as far as minimal models thus defined are concerned. But this does not imply that HPEs need be minimal models. Actually, the mistake seems to rest in regarding the checkerboard and the Hawk-Dove

9 Forber (2012) claims nothing really hinges on that distinction because for him global HPEs are a kind of explanations.

models as minimal. The checkerboard model may afford understanding precisely in virtue of *some* similarity or resemblance between the world and the model (Sugden 2009; Ylikoski and Aydinonat 2014). To maintain that the mechanism depicted by the checkerboard model is causally possible indeed appears to require at least a minimal assessment of its similarity with the actual world. One could nevertheless argue, as Resnik (1991) and Reydon (2012) do, that HPEs are in fact HAEs or merely serve heuristic purposes. But what is at stake is not whether HPEs are a species of explanations or not, but whether they afford understanding. Reydon, however, does not specifically address this issue. The fact that HPEs lack full empirical support does not necessarily imply they can't afford understanding. Furthermore, HPEs may sometimes serve heuristic purposes, but other times they may also afford understanding. The checkerboard model seems to do both. It suggests a novel empirical hypothesis that can orient future research, while also allowing to answer various questions about residential segregation. The two functions are not necessarily mutually exclusive.

It might be the case that scientists are sometimes too optimistic about results from HPEs or that they mistake some HPEs for HAEs. There might thus be cases of HPEs that do not improve our understanding. However, such negative readings do not imply that HPEs can't, out of principle, afford understanding. Descriptively, denying this capacity to HPEs is infelicitous as actual practitioners consider they afford understanding. However, the normative point may still hold, viz. that practitioners are in fact mistaken. That said, proponents of narrow KnUn would need to offer a plausible argument for why we should consider they are indeed mistaken. Arguments of that kind are currently lacking.

If HPEs may afford understanding, as it is plausible they sometimes do, then narrow KnUn faces a serious objection: it appears that having an explanation, in the sense of a HAE, is not necessary for understanding. Whereas HAEs of course afford understanding, HPEs may also.

4.6 BROAD KNUN

In the two preceding sections I argued that the two tenets of narrow KnUn are false. *Broad* KnUn, I contend, provides an alternative epistemology of understanding that fulfils the desiderata set forth in section 4.3. In particular, accounting for scientific practice should not come at the cost of blurring the difference between illusory and genuine understanding. It should also make salient the relationship between explanation and understanding. I would now like to provide a more positive characterization.

One fruitful way of advancing towards a more formal characterization of broad KnUn is to look at Reutlinger's (2016) theory of counterfactual explanation. This is because it already embraces one element of broad KnUn, namely it welcomes non-causal knowledge. Reutlinger's theory aims at capturing the 'common element' of causal and non-causal types of explanation without necessarily being tied to an interventionist interpretation of counterfactuals. Reutlinger's strategy is to stay as close as possible to the Woodwardian (2003) spirit of causal explanation, while making room for non-causal generalizations to serve as explanantia. An important motivation of his is precisely to accommodate mathematical explanations such as the widely discussed Königsberg bridges case (see, e.g., Pincock 2007). According to Reutlinger, a relation between an explanans and an explanandum is explanatory iff it satisfies the following conditions (2016, p. 737):

VERIDICALITY CONDITION Generalizations $G_1, \ldots, G_m$, the auxiliary statements $S_1, \ldots, S_n$, and the explanandum statement E must all be (approximately) true or be well confirmed.

IMPLICATION CONDITION $G_1, \ldots, G_m$ and $S_1, \ldots, S_n$ logically entail E or a conditional probability $P(E|S_1, \ldots, S_n)$.

DEPENDENCY CONDITION $G_1, \ldots, G_m$ support at least one counterfactual between $S_1, \ldots, S_n$ and E.

Since Reutlinger's theory allows for non-causal generalizations, it already incorporates the first tenet of broad KnUn. The conditions he states are those that explanations (i.e., HAEs) must satisfy. Accordingly, his theory leaves out the second tenet of broad KnUn, viz. that an explanation is not necessary for understanding.

The first step that will allow us to filter out understanding from explanation is by identifying in virtue of what explanations provide understanding. To put it in Lipton's (2009) terms, we have to separate the benefit explanations provide—understanding—from the explanations themselves. Reutlinger is not explicit about this, but since his theory is an extension of Woodward's (2003), we can find indications there. For Woodward, explanations provide understanding because they convey information that is relevant to answering *w*-questions about a phenomenon of interest. One understands when one obtains information about counterfactual dependence that allows to answer these questions. We thus see that what is key to understanding is the *information* some propositions provide, information that is closely related to the satisfaction of the dependency condition.

Broad KnUn expands on the idea that it is essentially information about counterfactual dependence that contributes to understanding, regardless of whether it is causal or not, and, crucially, regardless of whether it is obtained through an explanation or not. Reutlinger's theory already accommodates non-causal dependence by modifying the dependency condition. Having an explanation implies that certain relations of dependence actually obtain. This information that explanations provide allows to answer *w*-questions. But is it possible to answer some 'what-if-things-had-been-different' questions about a phenomenon even if the relations of dependence are not actual? The challenge for broad KnUn, therefore, is to show that having an explanation is not necessary for satisfying the dependency condition. Put differently, how could the dependency condition, which appears to be essential for understanding, be satisfied without the veridicality condition?

Reutlinger proposes an account of explanation and explanations are usually taken to be factive, i.e., they give true accounts of the facts. The function of the veridicality condition is precisely to ensure the factivity of explanation. False generalizations would not explain an explanandum and true generalizations would not explain an explanandum that is known to be false. For a set of propositions to count as an explanation, both the explanans and the explanandum must be true. But what if the veridicality condition is not satisfied? What if the generalizations are false or not known to be true? What if the explanans or explanandum are merely possible? Put differently, what if we have a HPE? According to Reutlinger's account, this would imply the relationship is not explanatory. However, if we accept the compelling evidence that HPEs may provide understanding despite the fact that they contain false explanantia or explananda, then this suggests that the veridicality condition is not necessary for understanding. Since we can obtain understanding from false explanantia or explananda, the veridicality condition may be necessary for explanation, but not for understanding.

This indicates that we can disentangle the condition for actual explanation in Reutlinger from the core constituents that specifically concern understanding. We can achieve this result, I submit, simply by amending the veridicality condition. Accordingly, we can modify the theory of explanation to achieve a theory of understanding. I propose that the relationship between an explanans and an explanandum affords understanding iff:

POSSIBILITY CONDITION The 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.

IMPLICATION CONDITION Generalizations $G_1, \ldots, G_m$ with the auxiliary statements $S_1, \ldots, S_n$ logically entail E or a conditional probability $P(E|S_1, \ldots, S_n)$.

DEPENDENCY CONDITION $G_1, \ldots, G_m$ support at least one counterfactual between $S_1, \ldots, S_n$ and E.

In other words, I propose to amend the veridicality condition for what I call the *possibility condition*.[10] Essentially, it relaxes the explanatory requirement that the explanans and explanandum be actual. What is actual is possible, but what is possible is not necessarily actual. Explanations require that all its constituents are actual, but not understanding.

10 Reutlinger and others (Reutlinger 2017a; Reutlinger, Hangleiter, et al. 2018) themselves suggest that HPEs do not satisfy the veridicality condition. The possibility condition I put forward is a novel proposal.

This fits very well with accounts of HPEs that view them as describing "how a set of parts and activities *might be organized* such that they exhibit the explanandum phenomenon" (Craver 2006, 361, my emphasis). They show how an event could possibly occur or how known processes can lead to different outcomes (see also chapter 5). This, in turn, affords understanding of real world phenomena. For instance, the checkerboard model exhibits a possible causal mechanism—possible causal generalization—that can bring about residential segregation, which is an actual phenomenon. However, the model itself does not explain residential segregation because we do not know whether it is actually that mechanism that produces segregation. Yet, the checkerboard model affords understanding by virtue of showing how it *could* be brought about. It supports counterfactuals of the form 'If individuals had not strong discriminatory preferences, then residential segregation could still occur'. More generally, it allows to make various counterfactual inferences about the phenomenon of residential segregation (Kuorikoski and Ylikoski 2015; Ylikoski and Aydinonat 2014). Despite the fact that it does not satisfy the veridicality condition, it does satisfy the possibility and the dependency conditions. If it were the case that preferences for not living in a minority status cause the phenomenon, then we would have an explanation. We would also of course understand. But we can also understand even when we lack an explanation insofar as the dependency condition is satisfied.

The "(im)possible according to the relevant modal interpretation and epistemic goal" clause within the possibility condition makes explicit that 1) there are various ways we can deem a proposition to be (im)possible (Lange 2009; Kment 2012) and 2) that understanding is achieved on the background of a specific epistemic goal. Many HPEs, for instance the checkerboard model, display possible *causal* dependence. However, other cases may appeal to different modalities (e.g., epistemic, nomic, metaphysical, etc.). In mathematical explanations, the relevant modality is mathematical or logical. The truth or falsehood of possibility claims is reached on the background of the suitable facts, depending both on the modality and on the epistemic goal. To say that something is logically possible does not imply that it is also causally possible. The same constraints do not apply. For the purpose of scientific understanding, causal, epistemic, or nomic possibility are perhaps the most relevant types. Broad KnUn, however, does not rule out a priori the sorts of possibilities or relations of dependence that may afford understanding.

To make this clearer, let us use an example discussed by Strevens (2013), that of the young earth creationists. They purportedly explain the formation of the Grand Canyon by citing one massive flood occurring over a short period of time. The flood would have laid down most of the different rock layers and the flood would have dug the canyon itself. Strevens says that he "cannot think of any conversational context in which it is correct to say, without frantic hedging, that the young earth creationists understand the formation of the Grand Canyon" (2013, p. 513). That they do not understand follows from Strevens's requirement that one needs to grasp a correct explanation in order to understand. The great flood explanation is false, i.e., its explanans is not true in the sense that it is not actual. The Grand Canyon was actually formed by other geological processes, namely by sediment accumulation, plate movement, and slow erosion. Recent evidence suggests large parts of the canyon may be as old as 70 million years (Flowers and Farley 2012), a far cry from the young creationists' claim.[11]

But does relaxing the veridicality condition imply that comparable cases may satisfy the possibility condition and thus afford understanding? If so, it may imply that the possibility condition is too liberal since utterly wrong-headed HPEs could afford understanding. Since one desideratum of an account of understanding is to allow demarcation of illusory from genuine understanding, this would be an unwelcome implication of broad KnUn. Again, offering a descriptively adequate epistemology of understanding should not make it too easy to obtain.

I agree with Strevens that the young earth creationists' great flood explanation does not afford understanding. Here, it is important to take into consideration what is the relevant modal interpretation for a given epistemic purpose. This allows to clarify

11 The age of the Grand Canyon is an on-going debate. Scientific advocates of a 'young' canyon date its origin to about 5-6 million years. In any case, we are still far from the young earth creationists who claim the earth was formed between five and ten thousand years ago.

how exactly a HPE contributes, or not, to understanding. If one asks ‘How possibly could have the Grand Canyon been formed?’, the question can most appropriately be interpreted as a request for causal information about the Grand Canyon. One wants to know how the Grand Canyon could have possibly been causally formed. Hence, not only is the young earth creationists’ great flood explanation actually false, but according to what we know about geological processes, this could *not* have happened the way the young earth creationists claim. In other words, the great flood explanation does not qualify as being a HPE and is not an appropriate answer to the how-possible question above.[12]

By that, I mean that the causal generalization linking the flood to the canyon is not only inconsistent with what we know about the causal history of this specific case, but also with other more general causal facts. It conflicts with the fact that geological processes forming canyons operate over long periods of time, not during one year. And while floods may produce certain geological formations (e.g., the Channeled Scablands in the US, see V. R. Baker and Bunker (1985); Waitt (1980)), the floods are local, not global. It also contradicts scientific facts about the age of the earth and of the Grand Canyon. Or it can hardly account for the presence of fossils in the different rock layers. What makes the young earth creationists’ explanation an inadequate answer to the how-possible question is not so much the fact that floods can play a causal role, but the precise way they claimed it happened and how inconsistent it is with everything else we know about geological processes. In a nutshell, it just could *not* have happened the way the young earth creationists claim. The great flood explanation thus does not fulfil the possibility condition since the how-possibly request calls for information about causal possibility, which the young earth creationists do not provide.

It also does not satisfy the dependency condition because the generalization linking floods and the Grand Canyon does not support the right kind of (true) counterfactuals. For instance, the

12 Of course, young earth creationists would most probably disagree with this statement. The flood causing the formation of the Grand Canyon is consistent with their other background beliefs. Here I am assuming that the geological sciences provide reliable empirical information.

young earth creationists' explanation implies that the following counterfactual is true: 'Had there been a great flood, the Grand Canyon would have been formed as a result'. However, that counterfactual is false. Perhaps if other geological processes would also have been sufficiently different the flood could have formed the canyon, but this qualification is not part of the young earth creationists' story. Absent changes to these other processes, the flood would not have brought about the Grand Canyon. It is at best unclear what sort of *w*-questions the great flood explanation may answer.

All that said, since the possibility condition allows for *im*possible explanantia and explananda, does this mean the great flood generalization would, on reflection, afford understanding? It could, but only in a specific set of circumstances. Let us suppose that we did not know that a flood during a very short period of time could have dug the canyon and laid down the rock layers, but that we did know that the actual explanation involved erosion, sediment accumulation, and plate movement. One question we may ask ourselves is whether other causes, like the flood, could have brought it about. We may thus build a model or simulation in which we try to generate the Grand Canyon with an intense flood as the main causal driving force. In order to achieve this result, we would most probably need to assume very different geological conditions and processes than the ones currently existing. In doing so, we would thus learn about the contingency, or necessity, of the actual explanation. We would also learn to what extent the actual world would need to be different for the flood to have the capacity to form the canyon. As Weisberg argues in discussing similar examples of impossible explananda, modelling the impossible is a sound scientific practice which may afford understanding.

> Why should theorists who are primarily interested in studying what is actual try to understand what isn't actual? The answer to this question cuts deep into the heart of theoretical practice: Theorists ultimately aim to partition the space of possibilities. They aim to understand what is possible, what is impossible, and why (Weisberg 2013, p. 128).

What this shows is that whether a given HPE affords understanding or not depends on what is exactly achieved and on the context of enquiry. Whereas the young earth creationists' specific 'explanation' brings nothing to the table in terms of understanding, we can imagine slightly different scenarios where investigating the conditions under which a flood could have formed the Grand Canyon would yield understanding. We understand better the Grand Canyon if we establish, say, that it is impossible to generate it with known initial conditions and a great flood as a possible process. Or that it is possible that a flood forms the Grand Canyon, but only if the world would have been a very different place. These generalizations may for instance support the counterfactual that 'Had there been a flood, the canyon would not have been formed as a result'. The crucial difference between this illustrative case and the young earth creationists' one is that the latter does not identify relevant possible causal generalizations and that these generalizations do not support counterfactuals.

Another way of illustrating this point is to briefly consider the cases of phlogiston theory or Ptolemaic astronomy. As I said in section 4.5, both theories rely on false generalizations and therefore do not explain. Looking at these cases with our contemporary eyes, shall we say they afford understanding?[13] Our account of understanding should view these cases as offering at best a limited understanding of nature. As with the young earth creationists' case, much depends on what we take the relevant *w*-questions to be. Both theories are famously unable to answer some questions that concern the phenomena they are supposed to explain. For instance, phlogiston theory has trouble supporting counterfactuals concerning the combustion of metals since some gain weight instead of losing it. Moreover, not only are their generalizations false, they provide a wrong-headed picture of the world.

So perhaps we might say that they depict *im*possible explanantia—e.g., the earth at the centre of the universe—, but whether phlogiston theory or Ptolemaic astronomy would satisfy the possibility

13 I want to set aside the question whether phlogiston theory, for instance, increased understanding of its contemporaries. This is out of the scope of this chapter.

and dependency conditions again depends on the epistemic goal and what sort of counterfactuals are under investigation. For example, one might be interested in exploring what difference it would make if the earth were at the centre of the universe. This may help understand why the world is the way it is. However, that is a different epistemic goal than what Ptolemaic astronomy purported to do.

My proposed account of the conditions under which the relationship between an explanans and an explanandum affords understanding, therefore, has the resources to discriminate between genuine and specious cases of understanding. While broad KnUn seeks to be as accommodating as possible in order to reflect actual scientific practice, it also imposes limits on what can count as affording understanding. The possibility and dependency conditions provide criteria according to which we can assess whether or not a relationship affords understanding.

One final objection may be the following: does discarding the veridicality condition imply that understanding is not factive? Reutlinger's (2016) veridicality condition originates from the usual requirement that explanations are factive. We may not need factive explanations for understanding, but we still want understanding to be factive. How can we have factive understanding without (factive) explanation? We can because the function of the veridicality condition, as it is formulated, is simply to ensure the relationship could be part of an actual explanation, not that it would afford understanding. An explanation is constituted by an actual explanandum and an actual explanans. A possible or false explanans does not explain an actual explanandum. That being said, a relationship that does not satisfy the veridicality condition may afford understanding. Indeed, it should not obscure the fact that generalizations can still support counterfactuals, and thus satisfy the dependency condition, even if the explanans or the explanandum are merely possible.

This is, in fact, the determining component of understanding. As we have seen above, it is information that allows to answer *w*-questions that is relevant for understanding (Grimm 2010; Woodward 2003). Explanations provide that kind of information, as HPEs can do. We can truthfully answer some *w*-questions even when we lack the actual explanation. For example, the checkerboard model shows that residential segregation could, in suitable conditions, be brought about if individuals had certain preferences. Since we consider the mechanism to be causally possible, we can use that information to answer 'what-if' questions. It therefore satisfies the dependency condition. To know that an explanandum may depend on a possible explanans is genuine knowledge of dependence and is, in that sense, factive. Likewise, that a possible explanans supports at least one counterfactual implies that we can evaluate its truth. Understanding is therefore still factive because it relies on matters of facts about possibility and counterfactuals. One would not understand without knowledge of possibility or without truthful assessment of counterfactuals. Knowing that something is possible responds to facts about possibility. This makes understanding factive. Broad KnUn therefore does not necessarily come at the cost of having to reject the factivity of understanding.[14]

4.7 CONCLUSION

Neither tenet of narrow KnUn is necessary. In other words, understanding does not depend on causal explanation. Amending Reutlinger's (2016) theory of counterfactual explanation as suggested above yields a sound basis for an epistemology based on broad KnUn. Upholding narrow KnUn comes at the price of having an epistemology of understanding that can't appraise neither descriptively nor normatively some actual scientific practices. We do not have to pay that price. Broad KnUn avoids all the problems that beset its narrow counterpart only at the cost of pushing to their logical conclusions various positions one can already find in the literature. Within broad KnUn we can easily view HPEs as affording understanding even though they do not actually explain. We can also see mathematical explanations as having the capacity to afford understanding of the world. Broad KnUn thus has a lot of normative appeal since it coheres with the actual conduct of science.

14 Some (e.g. Elgin 2007; Kvanvig 2003) maintain that understanding is non-factive. Whether it is or not has no direct implication for the account just presented.

Therefore, it fulfils an important desideratum of an account of understanding, namely that it should not regard vast areas of science as being mistaken about what they achieve. Broad KnUn is simply the formal and explicit recognition of both current scientific practice and of existing philosophical insights that provides a sound, unifying, and much needed epistemology of understanding.

5 HOW COULD MODELS POSSIBLY PROVIDE HPES?

5.1 INTRODUCTION

One proposal to solve the puzzle of model-based explanation is to view these models as providing HPEs (e.g. Craver 2006; Forber 2010; Reydon 2012; Grüne-Yanoff 2013a; Bokulich 2014; Ylikoski and Aydinonat 2014; Rice 2016). They usually contrast HPEs to HAEs, which are accounts of how phenomena actually occurred.

This response, however, raises two sets of issues. First, although the distinction between HPEs and HAEs is widely acknowledged, what it precisely amounts to is ambiguous since existing views attribute different features to HPEs. There are two important families of accounts, which I call the Dray-type and the Hempel-type. While the Dray-type (e.g. Dray 1968; Forber 2010) considers HPEs and HAEs to be a different species of explanation, the Hempel-type (e.g. Hempel 1965a; Brandon 1990; Bokulich 2014) distinguishes them by their degree of empirical support. These distinctions, I contend, do not appropriately capture the real contrast between them. To consider models simply as HPEs therefore does not straightforwardly solve the puzzle. Rather, it raises an important question about the nature of HPEs, namely what demarcates them from HAEs?

The second issue concerns the relationship between highly idealized models and HPEs. Many models appear to not depict possibilities, but rather *im*possibilities (van Riel 2015). If this is the case, then how could models possibly provide HPEs if their idealizations can't possibly be true? Spelling out the epistemic contribution of models in terms of HPEs does not constitute genuine progress unless we have a clear idea of how models can indeed provide them.

I aim to provide an account of HPEs that clarifies their nature in the context of solving the puzzle of model-based explanation.

In section 2 I introduce the two types of HPEs. Section 3 shows why they are inadequate in the context of solving the puzzle of model-based explanation. I develop the account of HPEs as providing knowledge of possible explanations in section 4. In section 5 I spell out in more details what is the relationship between HPEs, models, and explanation. Section 6 concludes.

5.2 TWO TYPES OF HPES

The notion of 'how-possibly explanation' was introduced by Dray (1957) in his discussion over the adequacy of the deductive-nomological (D-N) model of explanation (Hempel and Oppenheim 1948) in history.[1] Since Dray, the notion has been revisited many times and the concept was recently used in some discussions over the epistemic contribution of scientific models. Unfortunately, there is now a multitude of concepts on offer and it is now harder than ever to cash out a precise distinction between HAEs and HPEs. Table 5.1 gives a sense of the various terminology used when the terms are classified into three broad categories.

Table 5.1: Types of explanations

'How actually'	'How-possibly'	Misc.
why-necessarily (Dray 1957)	how-possibly (Dray 1957)	would-be explanation (Hindriks 2013)
true (Hempel 1965a)	potential (Hempel 1965a)	pseudo-explanation (Resnik 1991)
how-actually (Dray 1968)	possible explanations how-possibly (Dray 1968)	how-plausibly (Machamer et al. 2000)
potential how-actually (Reydon 2012)	global how-possibly (Forber 2010)	genuine explanations in need of explananda (Reydon 2012)

1 Dray (1954) did not yet use the terminology of 'how-possibly'.

'How actually'	'How-possibly'	Misc.
possible explanation why-necessarily (Dray 1968)	local how-possibly (Forber 2010)	more or less strongly confirmed (Hempel 1965a)

The literature thus far demonstrates that there is a relevant distinction between HPEs and HAEs. However, what this distinction amounts to remains a source of contention (Bokulich 2014). How are we to make sense of this?

One possibility is to use the distinction between what Forber (2010) calls, following Sober (2003), Hempel's problem and Peirce's problem. The former is the problem of whether there are different types of explanations and of what constitutes an ideally complete explanation and the latter is the problem of comparing evidential support and accordingly choosing the best explanation. According to Forber, the dividing line between HAE and HPEs concerns Hempel's problem, not Peirce's. He uses the distinction to argue that since Brandon (1990) views the difference between HPEs and HAEs as one evidential support—Peirce's problem—, then what Brandon sees as HPEs are not, in fact, genuine ones. Rather, Forber says that Brandon's HPEs should "count as incompletely confirmed or tentative how-actually explanations [...]" (2010, p. 36). Reydon agrees and call them "potential how-actually explanations" (2012, p. 307). These discussions indicate, as Reydon explicitly emphasizes, that HPEs and HAEs are sometimes believed to have a different logical structure. According to this view, HPEs and HAEs do not have the same form. A 'potential how-actually explanation' differs from a HPE in terms of form and also lacks the appropriate evidential support to be a genuine or complete HAE.

I find this discussion of Hempel's and Peirce's problems illuminating in that it suggests two dimensions along which to organize the accounts of HPEs on offer. As Strevens (2013, pp. 512-513) notes, all accounts of explanation place two types of conditions on explanatory correctness: internal and external conditions.[2] Internal conditions state what *form* a set of propositions should take to count as an explanation. It is about the internal structure of the explanation. For instance, internal conditions for a D-N explanation indicate that the explanation needs to have the form of a deductive argument which takes a lawlike statement and initial conditions as its premises and an explanandum as its conclusion. The lawlike statement must be an essential part of the derivation.

External conditions concern the empirical match between the explanation and what is to be explained. An externally correct D-N explanation, for example, is one whose lawlike statement, initial conditions, and explanandum are all true. It is important to note that whether these conditions are satisfied is in principle independent. An explanation can have the form of a D-N explanation, thus satisfying the internal conditions, yet be false. An explanation can also identify an externally correct explanans and explanandum, yet fail to connect them properly—e.g., the explanandum does not deductively follow from the explanans—and therefore does not satisfy the internal conditions.

Making this distinction allows us to see that there are in fact two main classifications in the literature. One class of HPEs states that the crucial difference between HPEs and HAEs is that they differ in their internal conditions. In the tradition of Dray (1957), these accounts suggest that HAEs and HPEs are, in virtue of a different logical structure, different species of the same genus explanation. Another class of HPEs instead states that the essential difference is in whether and how the external conditions are satisfied. Following Hempel (1965a), these accounts state that HPEs are internally similar to HAEs, but they do not satisfy the external conditions for explanatoriness. To make this clearer, let us look at, respectively, what I call the Dray and Hempel types.

5.2.1 Dray-type: HPEs as a different species of explanation

Dray (1957) argued that there was a type of explanation that did not obey the strict logic of the D-N model. D-N explana-

2 De Regt (2009) makes a similar distinction between the "logical" and "empirical" requirements for scientific understanding as well as Hempel and Oppenheim (1948) with the logical and empirical conditions of adequacy.

tions reveal why an explanandum *had to* happen, hence why Dray calls them 'why-necessarily'. In explaining how-possibly, however, "we rebut the presumption that it [the explanandum] *could not* have happened, by showing that, in the light of certain further facts, there is after all no good reason for supposing that it could not have happened" (161, emphasis in original). According to Dray, the need for a HPE arises in the specific circumstances where one does not believe the explanandum could have happened. The HPE rebuts this belief by showing how it could possibly have happened. If one believes it was impossible for something to happen, a HPE makes clear to her that it was, in fact, possible. But it does not imply that it had to happen. For Dray, a HPE reveals one or a few necessary conditions for the explanandum, though it does not show the sufficient conditions.

However, this does not mean that HPEs are incomplete D-N explanations (cf. Brandon 1990). Rather, Dray says they fulfil a different explanatory task, which is the refutation of the belief that some state of affairs is impossible. In this sense, HPEs can be complete. In D-N cases, if a belief is rebutted, it is the belief that the explanandum need not have happened. Put differently, what Dray calls why-necessarily explanations answer the question 'Why it did so?' whereas HPEs answer 'How it could be?' (Dray 1968). For him, then, what distinguishes HPEs from HAEs is their different internal conditions: the "two kinds of explanation are logically independent" (Dray 1957, p. 167).

In response to some critics, Dray (1968, p. 399) raises an interesting distinction between 'possible explanations how-possibly' and 'possible explanations why-necessarily'. This makes clear that not all how-possibly explanations may count as true explanations. HPEs can also be false. Thus, a 'possible explanation how-possibly' is one that meets the internal conditions, but not the external ones. Likewise, a HAE can also only satisfy the internal conditions. In his terms, it is a 'possible explanation why-necessarily'. 'Possible explanations how-possibly' and 'possible explanations why-necessarily' differ in terms of their internal conditions, but are similar insofar as they both do not meet the external conditions. For Dray, a HPE has to offer a true necessary condition. A HPE that would offer a false or irrelevant one would only be a 'possible explanation how-possibly'.

Forber (2010), who is interested in explanations in evolutionary biology, follows Dray and frames the problem similarly by identifying three different species of explanations: 1) global how-possibly, 2) local how-possibly, and 3) how-actually explanations. Like Dray, Forber argues that each sort of explanation answers a different question:

> The first type [global HPE] answers the question: could some potential process produce evolutionary changes in idealized populations? The second [local HPE]: could some known process produce, in a way consistent with the local information set for a real population, an observed evolutionary outcome or pattern? And the third [HAE]: why, exactly, did some target evolutionary outcome or pattern occur (Forber 2010, p. 34)?

For Forber, these explanations all have different internal conditions. Where Forber differs significantly from Dray, however, is when he claims that "we can successfully explain how-possibly without *any* empirical support [...]; we need only show that some outcome or pattern is consistent with a specific set of information" (Forber 2010, 39, emphasis in original). Indeed, Dray's account requires that rebutting the impossibility presumption identifies an actual necessary condition for the explanandum. This of course demands some sort of empirical support. Relatedly, Reydon (2012), in a critique of Forber, also notes that his global and local HPEs do not exactly have the same structure as Dray's HPEs. For Dray, HPEs serve to refute the belief that a certain state of affairs is impossible. This is not necessarily the case for Forber's conception of HPEs. However, his project is similar to Dray's in that he maintains that there are different *types* of explanations and that for him HPEs are not just incomplete or partial HAEs.[3]

3 In a reply to Reydon (2012), Forber (2012) says this is the main feature which makes his project similar to Dray's.

5.2.2 Hempel-type: HAEs lacking confirmation

Hempel denied HPEs were a different type of explanation. He held that 'how-possibly' concerned the pragmatics of explanation (Hempel 1965a, p. 428). The fact that an explanation may have to be presented differently depending on one's prior beliefs and explanatory requests does not imply that there are two different types of explanation. Accordingly, information may have to be presented differently in order to make it intelligible to different persons, but that does not imply that the explanations are fundamentally different.

Hempel, in contrast with Dray and Forber, rather argued that all explanations have to satisfy the same internal conditions, namely those specified by the D-N model. Yet, Hempel differentiated explanations according to whether they are 'potential', 'more or less strongly supported or confirmed by evidence', or 'true' explanations (Hempel 1965a, p. 338). A 'potential' explanation is defined as "any argument that has the character of a D-N explanation except that the sentences constituting its explanans need not be true" (ibid.). That is to say, it is an explanation that satisfies the internal conditions, but not the external ones—i.e., the explanans is false or not known to be true. A 'true' D-N explanation is one that meets both the internal and external conditions.

Hempel rejected the assertion that HPEs were a distinct species of explanation and therefore did not believe it was necessary to characterize them further. Even though some accounts (e.g. Salmon 1989; Brandon 1990; Craver 2006; see also Bokulich 2014) have equated Dray's HPEs with Hempel's potential explanations, it is in a sense misleading since for Hempel 'explaining how-possibly' belonged to the pragmatics of explanation and not to its logic. As it stands, Hempel's potential explanations can be viewed as another way of spelling out the distinction between HAEs and those 'explanations' that do not satisfy the conditions of HAEs. For the sake of the present discussion, what I call a Hempel-type HPE is thus a potential explanation.

Contrary to Dray, Hempel held that all explanations have the same internal conditions. Therefore, he rejected that there were different species of explanations. A potential explanation is not an explanation per se given that it does not actually explain. For a set of propositions to explain, the external conditions also need to be fulfilled. We can thus regard 'potential', 'more or less strongly supported by the evidence', and 'true' explanations as a continuum. Where they differ is in the degree to which they each satisfy the external conditions. The less evidence we have to warrant belief in the actual (true) explanation, the closer we are to a potential explanation. Conversely, if the explanation is for all practical purposes confirmed, then we are closer to the actual (true) explanation. Were a Hempel-type HPE to be true—or be empirically confirmed—, then it would satisfy the external conditions and thus explain.

5.3 DRAY AND HEMPEL TYPES IN PRACTICE

Dray-type HPEs are thus distinct from Hempel-type potential explanations. A Dray-type HPE has both different internal and external conditions. Internally, it is a different species of explanation, whereas Hempel-type potential explanations are not. Externally, the strict Dray-type also needs to identify true necessary conditions for the explanandum. Potential explanations precisely do not have these truth requirements. It is thus rather surprising—and confusing—that some of the accounts in recent literature have adopted Dray's language, but Hempel's concept.[4] For instance, Resnik says that "the difference between how-possibly and how-actually explanations is quantitative–a difference of degree–since empirical support comes in degrees" (1991, p. 143). To give another example, Craver maintains that how-possibly models "are purported to explain, but they are only loosely constrained conjectures about the mechanism that produces the explanandum phenomenon" (2006, p. 361). In these two cases, the use of HPEs is closer to the Hempel-type than to Dray. HPEs are not so much different types of explanation then ones not meeting the external

4 Bokulich makes a similar point: "Subsequent discussions of how-possibly explanations typically treat them instead as *potential* explanations" (2014, 322, emphasis in original). See also Reydon (2012).

But is it not only a language problem? What is really at stake if scholars misidentify 'potential' explanations as 'how-possibly' explanations? The problem is that the confusion occurs not only at the semantic level. In many debates on theoretical models, the concept of HPE is mobilized to account for their epistemic contribution (e.g. Brandon 1990; Cooper 1996; Craver 2006; Aydinonat 2007; Grüne-Yanoff 2013b,a; Rohwer and Rice 2013; Bokulich 2014; Ylikoski and Aydinonat 2014; Rice 2016). Unfortunately, the lack of a clear and shared account of HPEs makes resorting to them a shaky strategy. The following example illustrates the kind of issues we often encounter.

For instance, as we have seen in the previous chapter, Rohwer and Rice (2013) argue that the Hawk-Dove model (Maynard Smith and Price 1973; Maynard Smith 1982) is "explanatory" and provides "understanding". It does so, according to them, even though it is *not* an explanation. It is explanatory and provides understanding, they say, because "the model is still able to answer certain key how-possibly questions (Resnik 1991; Forber 2010; Reydon 2012)" (Rohwer and Rice 2013, p. 349). In all fairness, Rohwer and Rice do not themselves develop extensively what they mean by 'how-possibly'. Yet, the three authors Rohwer and Rice cite to support their claim all have substantially different views over what HPEs are. For Resnik, the main purpose of HPEs is heuristic. Fruitful HPEs can help to develop new and better theories. However, Resnik is silent with respect to the relation between HPEs and understanding. Forber, on the contrary, precisely argues against views such as Resnik by defending that HPEs, local and global, are a distinct type of explanation.[5] This also contradicts Rohwer and Rice's point that HPEs are not proper explanations. In fact, that HPEs *are* explanations in their own right is Forber's main contention. Finally, Reydon argues, among other things, that Forber's global HPEs are explanations, but not how-possibly. Without digressing into the details, as Rohwer and Rice describe the Hawk-Dove model it would clearly be a global HPE, not a local one. Thus, far from supporting Rohwer and Rice, this goes against their claim that the Hawk-Dove model is not an explanation and is only in the 'how-possibly' business.

5 He considers Resnik has a similar account as Brandon (1990), who he argues directly against.

As we can see, it is not possible to rely straightforwardly on the notion of HPEs to make sense of scientific modelling. Depending on one's favourite account, HPEs can provide fruitful heuristics or be viewed as complete and genuine explanations to specific types of questions. This is of course making a significant difference when the question is precisely to assess the explanatoriness of models. If HPEs are to play a role in our appraisal of models, we better agree on what we take the features of HPEs to be and what is their epistemic purpose. Simply saying that models provide HPEs is unsatisfactory in the absence of an appropriate account.

Of course, that there are competing views over the appropriate notion of HPEs in the literature does not imply that all existing accounts of HPEs are mistaken. However, what I want to show is that neither the Dray-type nor the Hempel-type are fully adequate. To do so, it suffices briefly to consider the case of Schelling's (1971, 1978) checkerboard model of residential segregation. The most popular interpretation of the model is that it provides a sort of HPE (Aydinonat 2007; Grüne-Yanoff 2013a; Weisberg 2013; Ylikoski and Aydinonat 2014). Weisberg (2013, pp. 118-119) summarizes as follows the question the model raises and answers:

> In other words, how is it possible for segregation to happen in a city without collective preferences for segregation? The answer is that this is possible when every individual has a small preference for similar neighbors and tries to satisfy this preference.

The model is not strictly speaking a Hempel-type potential explanation because it goes *beyond* meeting the internal conditions. The model does not only tell us how certain consequences can be derived. When considering the model, "we see Schelling's checkerboard cities as *possible cities*" and see the similarity between the model and the world "by accepting that the model world *could be* real - that it describes a state of affairs that is *credible* [...]" (Sugden 2000, 25, emphasis in original). The model appears to tell us something more than just the fact that segregation can be obtained from given rules of behaviour.

However, it is also not clear whether the model actually explains the general phenomenon of residential segregation or specific instances of it. For this reason, Aydinonat (2007, p. 430) considers we should view the checkerboard model as offering a "partial potential (theoretical) explanation". What the model shows is that it is possible that preferences for not living in a minority status bring about segregation. But that it is possible does not imply that it actually happens in this way. This is therefore different from offering a merely internally correct explanation and from providing a HAE. Furthermore, there is a prima facie puzzle on the standard reading of the Hempel-type since it was not developed to deal with the idealizations models typically contain. Laws of nature can be true or false, but we can't really say the same of models (Reiss 2012b). It is therefore unclear how the external conditions of the Hempel-type relate to models.

It is also not a Dray-type HPE for two different reasons. Firstly, what the model refutes is not the belief that residential segregation could not happen. The impossibility does not concern the explanandum, but rather the explanans. It was considered unlikely, if not impossible, that something other than strong discriminatory preferences could bring about residential segregation (W. A. V. Clark and Fossett 2008). As Grüne-Yanoff (2009) argues, the checkerboard model contradicts this belief. But no one had any reason to disbelieve that segregation exists or that it could happen. Secondly, since the checkerboard model does not identify actual necessary conditions for residential segregation, it would constitute a 'possible explanation how-possibly' (see Dray 1968), not a HPE. Since the causal factors identified in the checkerboard model are at best sufficient for segregation—not necessary—and are not known to be actual, then it would only satisfy the internal conditions for a Dray-type HPE, but not the external ones. There might be other instances of segregation where these causal factors are not present.

5.4 THE INTERNAL AND EXTERNAL CONDITIONS OF HPES

What I take to be the distinguishing characteristic of HPEs is the modal information they convey. This appears to be a feature that both proponents of Dray-type and Hempel-type HPEs emphasize, albeit while not explicitly appealing to it. Consider the following quotations from commentators on diverse sides of the debate (emphases in original).

> What we know seems to rule out the possibility of the occurrence which is to be explained. The explanation consists in showing that in spite of appearances to the contrary, it is not an impossible one after all (Dray 1957, p. 161).

> We use the notion of a potential explanation, for example, when we ask whether a novel and as yet untested law or theory would provide an explanation for some empirical phenomenon [...] (Hempel 1965a, p. 338).

> What good is a speculative how-possibly explanation? The short answer is: it shows how known evolutionary mechanisms *could* produce known phenomena (Brandon 1990, p. 180).

> They [how-possibly models] describe how a set of parts and activities might be organized such that they exhibit the explanandum phenomenon (Craver 2006, p. 361).

> In contrast, how-possibly explanations aim to explain how some event could possibly occur (Forber 2010, p. 33).

> The Hawk–Dove game is intended to show how individual selection could *possibly* produce this behavior in a wide range of populations (Rice 2016, p. 92).

All the preceding quotes suggest that HPEs have something to do with *modality*. The use of words like 'might', 'could', and 'possibility' are all modal terms. It is interesting to see that regardless of one's specific position in the debate over HPEs, a common idea is that HPEs provide modal information. What I take to be the defining feature of HPEs is therefore not the type of question they answer nor their degree of empirical confirmation, but rather that they contribute to our knowledge of the possibility of certain states of affairs. HAEs, on the contrary, contribute to our knowledge of what is actually the case. Using this demarcation criterion, we can spell out in more detail a general characterization of the internal and external conditions of HPEs.

5.4.1 Internal conditions

The first question we may then ask is whether HPEs and HAEs have the same internal conditions, viz., do they have the same structure or form? In its general form, an explanation "is a set of propositions with a certain structure" (Strevens 2013, p. 510). The question is thus whether HPEs and HAEs have the same propositions with the same structure. I propose that they do not. Following a useful characterization by van Riel (2015), I first take HAEs to express propositions of the following form:

HAE *p* because *q*

Here, 'because' is simply a shorthand for one's favourite relation of explanatory entailment. We might say that '*p* because *q*' on causal grounds, i.e., that '*q*' is a cause of '*p*'. But '*p* because *q*' could also be the case if a law of nature '*q*' would be subsuming '*p*'. Nothing hinges on the specifics of what constitutes the relation. Whether explanations should have the form of a deductive argument or invariant counterfactual generalizations is not a relevant issue here. This general form is therefore flexible and can accommodate various substantial views over the nature of explanation. A HAE is simply a set of propositions such that they are amenable to a formulation '*p* because *q*'.

One thing that is important to mention is that '*q*' can be divided into several propositions, say the generalizations or laws of nature '*q*' and the initial conditions '*c*'. A more precise characterization may thus be '*p* because *q* and *c*'. This will prove important when discussing specific accounts and can also make a difference when assessing, for instance, the epistemic contribution of a model. I will generally use the shorter version for the sake of brevity and simplicity. This characterization of HAEs should be fairly uncontroversial. The contentious issue is the following: How do HPEs differ from HAE with respect to form? Still following van Riel (2015), I suggest we view HPEs as propositions of the following form:

HPE $\Diamond(p \text{ because } q)$

Propositions such as this mean "It is possible that 'p because q' ". This captures a key desideratum of HPEs that virtually all accounts share, namely that HPEs express possibility claims. Propositions '$\Diamond(p$ because $q)$' and 'p because q' do not have the same form and do not express the same content. Whereas knowing a HAE implies knowing what is *actual*, knowing a HPE entails knowing what is *possible*. The possibility operator is placed on the explanation proposition in order to allow maximum flexibility in assessing specific cases. As the following discussion should make clear, a general account of HPEs should accommodate cases where either or both the explanans and explanandum are possible.

In this formulation, how the modality of the possibility operator should be interpreted is deliberately left open. This is an essential feature of the characterization I propose. Modality comes in various sorts, e.g., epistemic, metaphysical, causal, logical, nomic, etc. (see Kment 2017). For instance, it is, according to classical first-order logic, logically impossible that '$p \wedge \neg p$'; 'p' can't be both true and false at the same time. Or, it is nomically impossible for a solid uranium sphere to be more than a kilometre in diameter because it could not ever reach that diameter without triggering a chain reaction (see van Fraassen 1989). It is not an accidental empirical generalization, but impossible according to the laws of nature. By contrast, it is nomically possible for a solid sphere of gold to have that diameter even though we may never find one in the whole universe.

Quite often, the possibility of scientific interest is causal and typical HPEs thus establish that 'p because q' is causally possible. Logical possibility, on the other hand, is usually not very valuable insofar as logic rules out very few empirical possibilities. To know a logical possibility may not be informative with respect to learning about the empirical world. That said, as the burgeoning literature on mathematical explanation attests (e.g. Baron et al. 2017; Lange 2013, 2017; Pincock 2015), in some cases the relevant modality may well be mathematical. My goal here is not to restrict what counts as acceptable modality in all contexts, but rather to propose a general characterization. As I will show below, many accounts of HPEs rely on an implicit or explicit notion of modality, e.g., epistemic or causal possibility.

Leaving the interpretation of the possibility operator open has two main advantages. Firstly, it provides a flexible and unifying characterization. The same general schema can accommodate various HPEs. Causal or mathematical HPEs may have the same basic form, only different modal requirements. Secondly, it does not a priori assert what may be the relevant types of modality. In some contexts, scientists may be interested in nomic possibility. In others it may be mathematical possibility. My account does not preclude particular modalities.

How the possibility operator is interpreted has of course significant consequences on the external—or truth—conditions of a HPE. I will discuss it in more detail below, but here it is simply important to note that a thoroughly specified HPE states the modality of the operator. What matters is that the relevant modality ultimately needs to be selected in order to asses the possibility claim.

At first sight, it may appear that my proposal is simply a more formal characterization of Dray-type HPEs. Is this to say that, at least with respect to form, the Dray-type is the correct one? The Dray-type proponents are right to emphasize that the propositional content of HAEs and HPEs is different. But it is neither because HPEs answer a different type of question nor because they identify only necessary conditions.

Firstly, for Dray (e.g. 1968), HPEs only reveal necessary conditions and do not appeal to generalizations. Hence, the explanatory entailment relation—the 'because'—is for him categorically different between HPEs and HAEs. In Dray's (1968) parlance, if 'p because q' states sufficient conditions for 'p', then it is not a HPE, but rather an 'explanation why-necessarily'. Dray only requires the identification of appropriate actual initial conditions 'c'. A strict Dray-type could thus be formulated as follows: '$\Diamond p$ because c'. The form is therefore quite different.

But to show that an explanandum is possible, it is not necessary to identify necessary conditions. Suitable sufficient conditions may do the trick. As Salmon (1989, p. 137) notes, "any potential explanation not ruled out by known facts is a suitable answer". At the very least, some how-possibly questions can be successfully answered without citing necessary conditions. Moreover, as Reiner (1993) argues, finding necessary conditions is methodologically difficult, if not outright impossible. Indeed, the set of necessary conditions is considerably smaller since some causes may be individually sufficient for an explanandum. Establishing necessary conditions puts very high demands on our knowledge. Perhaps there are HPEs that identify necessary conditions, but it seems we do not have good reasons to be monists about this. On the contrary, my account leaves open what is the set of possible interpretations of the explanatory 'because' relation.

Secondly, my account downplays the difference between HPEs and HAEs. One critical claim of the Dray-type is that HPEs and HAEs are different species of explanation. Both Dray (1957) and Forber (2010) construe HPEs as answers to different types of questions. For Dray, HPEs rebutted the belief in the impossibility of the explanandum. Similarly, as I showed in section 2, Forber frames his different type of HPEs as answers to different questions. However, as was previously objected (Pitt 1959; Hempel 1965a), quite often one can answer a how-possibly question with a HAE. Put differently, a HAE implies its corresponding HPE. In fact, a HAE may be the best answer to a how-possibly question since one would not only know why something is possible, but why it is actually the case. Dray and others may be right to point out that a HAE is not necessary to answer a how-possible question. But if a HAE can also in principle answer these questions, it suggests there is not a one-to-one mapping between types of questions and their corresponding explanations. Therefore, whether how-possible questions require a different *kind* of explanation is implausible. A HPE can be sufficient for answering a how-possible question, but it is not necessary.

More concretely, why should we limit what questions can be asked in what contexts? Grüne-Yanoff (2013a) shows that in the course of constructing HPEs, scientists sometimes start with different questions or background information. One example he gives is Ainslie's (2001) feedback model of self-control. The model indicates that preferences displaying hyperbolic discounting are compatible with the phenomenon of moderate impulsiveness through a process of recursive self-prediction. Here Ainslie shows from possible initial conditions how a possible process could generate an actual idealized phenomenon. While Grüne-Yanoff discusses Forber's account, he does not describe this case as one of Forber's local or global HPE. Indeed, this example does not seem to readily fit within the global vs local distinction. It is not an inquiry into the strictly formal constraints of psychological possibility (global) nor does it postulate a known process or concrete population. Yet, I think Grüne-Yanoff is right to view this case as a HPE. At any rate, the problem is not so much that the example could not fit within Forber's categories, but that as useful as they are in describing particular cases of HPEs, they do not seem to naturally account for the variety of epistemic contexts.[6] The characterization I propose provides a general characterization that situates differences in the varying interpretations of the possibility operator and in the actuality or possibility of the constituent propositions (e.g., explanandum or initial conditions) of HPEs.

Forber's discussion of Brandon (1990) gives rise to a similar problem. Brandon views HAEs and HPEs as located on a same continuum of evidential support, HPEs having less of it than HAEs. As we have seen, Forber holds that Brandon's HPEs are not really HPEs, but tentative HAEs. Since Forber distinguishes HPEs from HAEs by their aim, not their result, evidence that supports a HAE can't carry to HPEs. But I do not see why a 'tentative HAE' could never, out of principle, count as a HPE.[7] If evidence for a HAE is incomplete and partial, why could it not support a HPE? The evidence we have might be insufficient for a HAE, yet be sufficient for a HPE. One's evidence may be lacking for establishing that an explanans actually caused an explanandum, though it may license a corresponding HPE.

In sum, I do not see any principled reason to restrict some questions and explanatory endeavours from qualifying as HPEs from the outset. Dray-type proponents have rightfully pointed

6 Bokulich (2014) makes a similar claim using a case study on the 'stripes' of the tiger bush. She claims that global and local constraints can be mixed depending on the level of abstraction.

7 Grüne-Yanoff (2013a, p. 859) presents a similar case that he considers to be a HPE. See also Resnik (1991) and Bokulich (2014).

out some specific instances of how-possible questions, but consideration of the diversity of practices and questions scientists address militate against setting strict constraints. As we have seen, sometimes what a model does is to establish the possibility of the process, not of the explanandum. More formally, one would have prior good reasons to believe that 'p', but not that 'q'. A model may exhibit that it is in fact possible that '$\Diamond q$' and therefore could increase confidence in '$\Diamond$(p because q)'. We could also have cases where we have independent reasons to believe that 'p' and 'q', but not that '$\Diamond$(p because q)'. In this situation, a model may connect the dots between 'p' and 'q' and establish that '$\Diamond$(p because q)'. The characterization of HPEs that I propose encompasses all forms of how-possible questions since it does not involve prior beliefs or explanatory aims. It also does not exclude that a HAE can provide a satisfactory answer to these question.

If there are substantial differences between the account I propose and the Dray-type, how is it different from the Hempel-type? One straightforward possibility is simply to reconstruct it as a false or unknown HAE: 'p because q', where 'q'—and therefore 'p because q'—is either false or not known to be true. For Hempel, what differentiated a potential explanation—the classic Hempel-type HPE—from a true explanation—a HAE—is that the latter fulfilled the external conditions whereas the former did not have to. Put differently, HPEs had the same form as HAEs, except for possibly being false or not known to be true. Understood as a potential explanation, a Hempel-type HPE thus does not tell whether or not 'p because q' is the case since it can be unknown. What it tells is how 'p', 'q', and 'p because q' stand in logical relation to one another. This characterization emphasizes that what separates HPEs from HAEs is whether the former satisfy the external conditions. Would they be met, i.e., were 'p' and 'q' be true, then 'p because q' would be true. This also accords with more recent accounts of HPEs close to the Hempel-type. For instance, according to Brandon (1990) the main difference between HAEs and HPEs is one of evidential support, not of form. In other words, HPEs and HAEs have the same internal conditions. What demarcates them lies in the satisfaction of the external conditions.

The preceding analysis shows that there is a significant difference in terms of modal information between the account of HPEs I propose and the Hempel-type. The Hempel-type shows that a given set of propositions meets the internal conditions of adequacy, e.g., that the explanandum logically follows from the explanans. The main information it provides is one of logical explanatory entailment. This information, albeit necessary and valuable, is limited. Crucially, fulfilling the internal conditions of adequacy does not provide any information beyond the mere logical possibility of the explanation. It does not tell whether the explanation or its constituents—i.e., '*p*' and '*q*'—are possible according to the typically interesting sorts of modalities like causal possibility. However, as testified by the quotes at the beginning of this section, modal information that goes beyond logical possibility appears essential to HPEs.

In fact, even accounts that have been traditionally classified in the Hempel-type put stronger modal demands on what should count as a HPE. For instance, both Salmon (1989, p. 137) and Brandon (1990, pp. 178-179) consider consistency with known facts to be a hallmark of HPEs. On the Dray-type side, Forber (2010, p. 34) appeals to a "causal principle of possibility". Furthermore, HPEs usually considered in the literature (e.g. Grüne-Yanoff 2013a) do not simply show that there is logical entailment between the explanans and the explanandum. This is too easy. Instead, they also appear to provide information about what is empirically possible in ways that go beyond mere logical possibility. For instance, a HPE can show how a causally possible mechanism may explain an actual phenomenon, like the checkerboard model does (Schelling 1978; Ylikoski and Aydinonat 2014).

Hence, perhaps a second way of characterizing the form of the Hempel-type HPE could be that it also has the form '$\Diamond$(*p* because *q*)', but the possibility operator should only receive a formal—e.g., logical—interpretation. This sort of possibility then leaves open whether the explanation is, for instance, causally possible. This characterization has the merit of highlighting the main issue with the Hempel-type HPE: it is too lax. As a matter of fact, a common worry in the HPEs literature (e.g. Brandon 1990) concerns so-called 'just-so' stories, a reference to Rudyard Kipling's ([1902] 1912) fabulous origin stories (e.g., *How the Camel*

Got His Hump) for children. We want to be able to demarcate between just-so stories that only have the form of explanations and genuine HPEs that tell us something, perhaps limited, about the world. More generally, we want to be able to demarcate between true and false HPEs. There is a difference between 'explaining' the camel's hump by citing its grumpiness expressed through repeated 'Humph!' and a HPE of the hump based on natural selection. While both may have the form of a HPE, only the evolutionary HPE is acceptable.

One challenge for accounts of HPEs is thus to allow for a demarcation criterion between genuine HPEs and 'just-so' stories. Just-so stories do not provide any substantial information about the world whereas HPEs inform us about the modal status of the explanation. This all suggests the need for stronger interpretations of the possibility operator. It may then be objected that my account does little in terms of a positive proposal to respond to worries raised by 'just-so' stories. While I indeed do not propose any substantive criterion, characterizing HPEs as propositions of the form '$\Diamond$(p because q)' is an important step. First, it allows to accommodate various standards through the possibility operator. What sort of possibility scientists are interested in depends on the particular research setting. My goal here is not to a priori specify what possibilities are important across all contexts.

Second, contrary to some strands of the Hempel-type, my account in principle allows for standards that go beyond meeting the internal conditions for explanation or beyond consistency with facts. Perhaps we need a more neutral label than 'just-so stories' for those set of propositions that only satisfy the internal conditions of adequacy. My point is simply that a good account of HPEs should also have the resources to go beyond these. By allowing further modalities, especially causal, my account broadens the scope of the Hempel-type. Moreover, as we will see in the next section, it does so by making explicit that (true) HPEs also need to satisfy some external conditions of adequacy.

5.4.2 External conditions

Typically, the external conditions for HAE require that 'p', 'q', and 'p because q' are all true. More precisely, a realist account of explanation will require that 'p' and 'q' are true whereas an antirealist may be content if only 'p because q' is true under certain conditions. For an antirealist, the explanatory relationship may be true in the absence of true explanantia and explananda (see Khalifa 2011).

Of course, if 'p because q' is true, then '$\Diamond$(p because q)' is trivially true. But the reverse is not true. A proposition '$\Diamond$(p because q)' may be true while the corresponding HAE 'p because q' is false. A HPE provides information about possibility whereas a HAE provides information about actuality. HPEs thus have more minimal empirical requirements. For '$\Diamond$(p because q)' to be true, none of 'p', 'q', or 'p because q' need to actually be the case.

HPEs are propositions that express that a given explanation is possible. HPEs may be true or false. Naturally, my goal is not to develop a full-fledged account of modality.[8] However, let me state some relevant considerations. Here I assume, along with the others in the literature on HPEs that take modality as one key feature, that modal claims have a truth value and that it can be established. As I said earlier, what is possible or not depends on the interpretation of the possibility operator. What is logically possible may not be nomically possible and thus the 'same' proposition may be true under one modality and false under another.

For the realist, the position I will assume going forward, an explanation can only be possible if all of its constituents are also possible *under the same modal interpretation*. If either 'p' or 'q' are not only false, but also impossible, then it can't be possible that 'p because q'. For if a given explanans 'q' is causally impossible, then the corresponding causal HPE '$\Diamond$(p because q)' can't be true. To use an example from the previous chapter, since it is causally impossible that the young earth creationists' flood caused the Grand Canyon, then the HPE is false. An impossible cause can't possibly provide a causal explanation of a given explanandum.[9]

8 For discussions on this topic, see, e.g., Gendler and Hawthorne (2002).

9 One may prefer to say that causal possibility, perhaps unlike other modalities (e.g., metaphysical) is a property of the explanatory link 'because' rather than of the explanans. Nothing substantial hinges on adopting that particular interpretation, although I believe my formulation is more general. Ultimately, this depends on how we want to decompose 'because q' in 'p because q'. If we follow Reutlinger (2016), as I did in the previous chapter, we might say

It could be logically possible, but not causally, hence the requisite to fix the interpretation of the possibility operator.

Conversely, for '$\Diamond$(p because q)' to be true, both '$\Diamond p$' and '$\Diamond q$' need to be true. Again, the possibility of a causal explanation requires the causal possibility of its constituent explanans.[10] This also means that '$\Diamond$(p because q)' does not imply that 'p' and 'q' are not the case. The possibility operator may only specify the 'because' relation. In fact, it may be the case that in HPEs either 'p' or 'q' are true, or both. For instance, certain generalizations, initial conditions, and explanandum may be actual, but the possibility of the explanation may not have been assessed.

How does this view differ from existing accounts of HPEs? The account I propose has different external conditions than the Hempel and Dray-types. The contrast is starkest with the Hempel-type where there are simply no requirements on the possibility of either 'p' or 'q'. But by not requiring them to be possible, as I already indicated, a Hempel-type HPE can in principle contain all sorts of quirky explanantia and explananda. In this sense, the external conditions of the Hempel-type are looser than the ones I propose. In fact, it is a HPE precisely because it does not satisfy the external conditions. As long as it is not a HAE, but has the same form, we have a Hempel-type HPE. Crucially, as we have seen, a Hempel-type HPE may have the right form, yet this says nothing about the possibility of either 'q' or of 'p because q'. The possibility of the statement 'p because q' is irrelevant. To know that '$\Diamond$(p because q)', however, requires that one makes a sort of modal assessment since 'p' or 'q' can't simply be false, they must be possible. To make the judgment over the HPE, one thus has to bring in background knowledge.

Should we then just call HPEs, following Hempel, a 'more or less strongly supported by the evidence' explanation? In a sense, HPEs are more or less strongly supported by the evidence. But for Hempel, what is more or less supported by the evidence

---

that 'because' more precisely corresponds to the implication condition (see sec. 4.6) and that 'q' is a causal generalization. We could then reconstruct causal explanations as propositions of the form 'p because (q causes p)', where 'q causes p' is a causal generalization that explains—e.g., implies and supports counterfactuals—the occurrence of the phenomenon p.

10 As I will discuss in the next section, idealizing models with putatively false explanantia only prima facie pose a challenge to my account of HPEs.

is the HAE, not the HPE. This is a crucial difference because the evidence we have may support judgments of possibility, but not of actuality. For instance, we may regard the checkerboard model as providing a HPE of residential segregation in the sense that it is, in fact, possible that the preferences for not living in a minority status cause segregation. Yet, we can also consider that we have no evidence that it is the HAE. In fact, we may even have evidence that some other cause is the HAE. In short, we may have zero evidence that the HPE is the actual explanation, yet still consider that it is possible. A HPE is therefore fully supported by the evidence for the specific claim it is making and this evidence may be independent of the evidence for the HAE.

The difference in external conditions with the Dray-type is more subtle as we can find various proposals for external conditions. However, some Dray-type HPEs put actuality constraints on either the explanandum or the explanans. On one side of the spectrum, as we have seen, for Dray (1968, pp. 399-401) a HPE identifies an *actual* necessary condition for the explanandum. To show that a given phenomenon was not impossible, one could not only appeal to a possible explanans. The necessary condition identified had to be actually instantiated. This necessary condition then rebutted the presumption of impossibility. According to Dray, if only a possible necessary condition was identified, then it had to be called a 'possible explanation how-possibly'. In other words, it would have the form of a HPE, yet would not satisfy the external conditions. Under my account, a HPE may include actual elements, but does not require any.

On the other side of the spectrum, as stated earlier, Forber's global HPEs, in particular, have minimal external conditions. He holds that mere consistency with formal constraints is sufficient and that no *additional* empirical support is necessary. On the contrary, his local HPEs have some actuality requirements. Only the initial conditions '*c*' may be speculative, the explanandum and the generalizations being actual. He holds that local HPEs "*are* just-so stories that speculate about the adaptive (or non-adaptive) evolutionary history of a lineage" (Forber 2010, 36, emphasis in original). For him, this is not per se a problem as long as they are recognized as such.

Contrary to some versions of the Hempel-type and Dray-type, the external conditions I propose are either more or less demanding. They can be more demanding insofar as the mere logical form of an explanation is often not sufficient. A stronger modal appraisal is called for. It compels to evaluate whether it is true that the explanation is possible according to a relevant modality, for instance causal possibility. One does so by looking at the possibility of the explanans, the explanandum, or the explanatory link. Contrary to some versions of the Dray-type, they can be less demanding as my account does not require any actual proposition. HPEs can include claims of actuality, but actuality is not required to establish claims of possibility.

5.4.3 Applying the conditions: a recap

I would like to end this discussion of the internal and external conditions of HPEs by a brief synthesis of how it helps to organize our thinking about the existing literature. Its principal virtue is to show where the differences between versions of HPEs lie. One conclusion I reach is that contemporary accounts (e.g. Brandon 1990; Forber 2010) are closer in terms of form than they seem to be. Indeed, they all emphasize that the end product of a HPE is a possibility claim. In this respect, I would say that Brandon's and Forber's account are similar to mine in that they take HPEs to be propositions of the form '$\Diamond(p$ because $q)$'.

Reydon's (2012) analysis of Forber's (2010) account supports this claim. Indeed, Reydon argues that Forber's global HPEs should be seen as "genuine explanations in need of explananda", thus implying that they have the same form as HAEs. However, they do not meet the external conditions of HAEs.[11] He also makes the case that Forber's local HPEs are in fact similar to Brandon (1990) and Resnik (1991), that they all have the same "logical structure" (Reydon 2012, p. 309). Finally, Reydon denies that local HPEs are a species of explanation—i.e., that they actually explain—because of a lack of (complete) empirical support.

11 This view has similarities to Sugden's (2011) 'explanations in search of observations', which I discuss in section 5.

		External conditions	
Account	Internal cond.	Possibility	Actuality
Dray (1957)	$\Diamond p$ because c	Epistemic (?) p	Necessary c
Hempel (1965)	p because q	Logical	—
Brandon (1990)	$\Diamond(p$ because $q)$	Epistemic c, 'because'	q
Forber (2010)	$\Diamond(p$ because $q)$	Local causal c, 'because'	p, q
Forber (2010)	$\Diamond(p$ because $q)$	Global causal p, q, c	—

Table 5.2: An application of the conditions

We can find more variance in what external conditions the different accounts put forward. Brandon (1990) appears to say that HPEs should include actual generalizations 'q' and that they should be epistemically possible. Forber's (2010) global HPEs should be causally possible—note the different modality–relative to the global information set, but imposes no actuality requirement. His local HPEs use the same modality, but relative to the local information set. And the explanandum and the generalizations should be actual.

Table 5.2 summarizes the preceding discussion, which suggests two things. First, that the internal conditions intersect in significant ways. The differences are less deep than we may think. Second, that most of the disagreements are located within the external conditions. What kind of modality matters and whether there are requirements about actuality are two important sources of contention.

5.5 HPES, MODELS, AND EXPLANATION

Since many accounts of scientific modelling rely on HPEs to appraise models, we should have a good idea of whether they are a genuine species of explanation, what sort of epistemic benefits we may expect from them, and whether scientific models can indeed provide HPEs. To answer these questions, let us first look at the last issue, namely, the relationship between models and explanations.

Models *are not* explanations. Explanations are simply sets of propositions satisfying the internal and external conditions stated by one's favourite theory (Strevens 2013). Whereas explanations are linguistic entities, models are widely viewed as being non-linguistic. This is clearly true of physical models such as the MONIAC (aka Phillips Hydraulic Computer), but also of the mathematical models usually discussed in the literature such as the Lotka-Volterra model (Weisberg 2007b). So, when we say that a model explains a given phenomenon, what this means is that there is a given model proposition—propositions that are true *within* the model—according to which '*p* because *q*' and that this proposition is epistemically related to the real-world proposition '*p* because *q*'. When we say 'Model *M* explains why *p* with *q*', we must be careful to not conclude that the model *is* the explanation. Instead, what it means is that the model provides sufficient reasons to believe that '*p* because *q*' is true. It is in this sense that models explain.[12]

Indeed, what many models do is to enable or justify the beliefs scientists have towards claims of this sort. Claveau and Vergara Fernández (2015) argue that models play an evidential role when model propositions make a difference to one's beliefs or justification for real-world propositions. When using a model, the model propositions one knows enter her evidential network for real-world propositions. Sometimes, these model propositions make a difference by enabling belief in the real-world proposition or by increasing justification. For instance, they show, as illustrated by the Diamond-Mortensen-Pissarides (DMP) model of the labor market, that entertaining certain model propositions can for instance increase the justification one has in the real-world proposition that higher unemployment benefits lead to higher unemployment. But besides this specific example, the fact that modelling is such a widespread—and successful—epistemic activity would be rather mysterious if model propositions would not enter and make a difference to scientists' evidential networks. That models can and do provide evidence for propositions of the

12 Explanations are always given on the background of prior beliefs and competing explanations and this does not imply a pragmatic or psychological account of explanation (see Woodward 2003, sec 5.12).

form '$\Diamond$(*p* because *q*)' or '*p* because *q*' should hardly be controversial.

That said, one could ask in virtue of what *exactly* can models provide such evidence. For my current purposes, I do not want to take a particular stance on what constitutes these reasons to believe. It might be by virtue of the model standing in some suitable representation relationship with the world, e.g., of isomorphism (e.g. French and Ladyman 1999) or similarity (e.g. Giere 2010). The model might be a fiction that licenses inferences to the world (e.g. Suárez 2009). Solving the (difficult) problem of how model propositions can correspond or license inferences to world propositions is naturally out of the scope of this paper. Nothing hinges on the specifics of how models provide these reasons, we only need to grant that models do play this evidential role.

Discussing one recent and influential account of models may help to see the evidential role of models and its relationship to possibility. Sugden (2000, 2009, 2011, 2013) argues, especially in his more recent work, that posited similarities between models and the world may license inductive inferences from the model to the world. For instance, Sugden describes the inductive schema of explanation where the facts that:

1. the explanandum '*p*' is caused by the explanans '*q*' in the model world, and that;

2. both '*p*' and '*q*' occur in the world;

3. provide "*reason to believe*" that '*p*' is caused by '*q*' (Sugden 2000, 19, emphasis in original, 2013, p. 240).

Here the model propositions clearly enter into the modeller's evidential network. What happens in the model serves to justify the inference to the world. While similarity is according to Sugden the key notion to license inductive inferences, he argues that one important dimension along which to judge it is credibility.[13]

13 Sugden recently made this point explicit: "The fundamental explanatory concept in my account of models is not credibility but similarity" (Sugden 2013, p. 240). This puts into question to what extent his account is then different from 'isolationists' accounts like Mäki (2009a) , but this is out of the scope of the current chapter.

In this context, credibility means that the confidence we have in our inferences is "greater the extent to which we can understand the relevant model as a description of how the world *could be*" (Sugden 2000, 24, emphasis in original). Credibility in that sense is not about considering that the model *is* real, but about judging that it is compatible with one's knowledge and beliefs about the world (Sugden 2000, p. 25, 2009, p. 18, 2011, p. 718).

Using the account of HPEs developed here helps to clarify Sugden's views and the debate about them. In a discussion of Schelling's (1971, 1978) model of residential segregation, Sugden (2011) argues that the 1971 model is an "explanation in search of an observation" (2011, p. 722) whereas the 1978 model was really trying to tell us something about the world. In other words, the 1978 model is an explanation and the 1971 model is a potential explanation (e.g. Sugden 2011, p. 734).[14] While Sugden does not really define what he means by 'potential explanation', for him the 1971 model is less credible than the 1978 one. It *could* be real, but not in the same manner. Or consider how Sugden argues against the 'conceptual exploration' (Hausman 1992b) view according to which theoretical modelling is about exploring the internal formal properties of models. Sugden believes that the checkerboard model, especially the 1978 one, and Akerlof's (1970) 'market for lemons' go beyond mere conceptual exploration. According to him:

> they are sketches of processes which, according to their creators, might explain phenomena we can observe in the real world. *But the sense of 'might explain' here is not just the kind of logical possibility that could be discovered by conceptual exploration* (Sugden 2000, 11, my emphasis; see also Sugden 2009, p. 23).

My account provides a ready-made clarification of what distinguishes the 1971 and 1978 models and of what demarcates conceptual exploration from other more ambitious modelling exercises. It locates the disagreement in the divergent interpretations of the possibility operator. In the terms I propose, Sugden

14 Sugden appears to use a pragmatic or instrumentalist account of explanation (see Sugden 2013). Whether this is the right account and whether the checkerboard provides a HAE or not is disputed (cf. Aydinonat 2007).

argues that the checkerboard model goes beyond providing evidence for claims of logical possibility. While he does not frame the issue in terms of different modalities, it is clearly the underlying argument. As explained earlier, his notion of credibility appears to be closer to a causal or epistemic interpretation of modality. Likewise, we could interpret the difference between the 1971 and 1978 models as one of modal appraisal. The 1971 model is not considered to be possible in the same sense as the 1978 one.

Furthermore, this may also help understand the source of the debate between Sugden and rival accounts (e.g. Aydinonat 2007; Grüne-Yanoff 2009; Mäki 2009a). Without clearly specifying what sort of possibility commentators consider to be relevant for the case at hand, they run the risk of talking past each other. My account of HPEs is therefore not in disagreement with current views of models. It rather provides a framework within which we can think more clearly about the potential sources of disagreements.

One potential major objection to my account comes from van Riel (2015). He argues that models can't provide HPEs since many successful models do not describe possibilities, but rather *impossibilities*. His argument is that since many of the entities or processes postulated by models are conceptually impossible (e.g., models of water as particles or as continuous medium are inconsistent), then these models can't be considered to be HPEs. Indeed, many idealized entities and processes that models contain could not, sensu stricto, possibly be the case. Sugden, for instance, recognizes this state of affairs: "Economic models often contain idealisations which, if interpreted literally, *cannot* be true" (2009, 18, emphasis in original). But why is this not an obstacle?

The misunderstanding lies in taking model propositions at their face value. Of course, if the possibility operator is applied literally to them, then many, if not the vast majority, of models will depict impossibilities.[15] However, the objects of knowledge, HAEs and HPEs, are not propositions about the model, but about

15 Van Riel (2015, p. 3845) is not concerned with logical possibility since it is not the possibility typically relevant for scientific usage. His argument concerns metaphysical, conceptual, epistemic, and nomological interpretations of modality.

the world. A model does not provide a HPE in virtue of its model propositions being literally true of reality. Instead, it provides evidence for corresponding world propositions (Claveau and Vergara Fernández 2015). Accordingly, we should not interpret the possibility operator as directly bearing on the model propositions. Possibility judgments are rather on the appropriate world propositions. For a given HPE of the form '$\Diamond$(*p* because *q*)' to hold, it is thus not necessary that the model propositions of '*p*' or '*q*' are possible. It is simply required, as we have seen above, that these model propositions be part of one's evidential network for the HPE.

In fact, as Van Riel's (2015) own account of the content of model-based information suggests, it is unclear to what extent he would disagree with my proposal. Van Riel's project is to provide an account of the information supplied by models that deviate—e.g., depict impossibilities—from reality. He suggests that using a hyper-intensional operator of the form 'according to model *M* in context *C*, (*p* because *q*)' allows propositions based on false models to be factive and cancels ontological commitments. It is hyper-intensional "in the sense that substitution of co-intensional expressions may turn a truth into a falsehood" (van Riel 2015, p. 3851). To use his own example, we can't *know* that the difference between day and night is due to the sun circulating around the earth since this is not actually the case. However, we can know that *according to the geocentric model*, this proposition is true.

So it seems that van Riel and I partly have different projects. He offers an account of model-based knowledge and holds that models *are* not HPEs, i.e., that models do not literally provide information about possibilities. On this we both agree. It seems the potential confusion stems from not distinguishing properly model propositions from real-world ones. But I see no reason in his account and elsewhere to deny that model-based propositions may provide evidence for real-world claims of the form '$\Diamond$(*p* because *q*)'. If anything, his account lays the ground for a translation manual between model and real-world propositions—which I here assume the feasibility—and thus concurs with my views. I therefore do not believe that my account of HPEs is a target for his arguments. Some false models may provide reasons to believe that a given explanation is, or not, possible.

That said, there might be a connection between the degree of idealization of model propositions and their capacity to serve as evidence for world propositions. That models misrepresent and idealize is almost a platitude. Perhaps the 'more impossible' the model propositions, the less they can serve as evidence for world propositions. However, to what extent certain idealizations facilitate or hinder our capacity to learn about the world is an important, but separate, question. There is no a priori reason to doubt that some idealized models contribute to our knowledge.

Again, we here only need to grant that some idealized models can indeed serve as evidence for world propositions. Bokulich (2014), for instance, argues that some highly idealized models may be closer to providing HAEs than others providing more concrete causal details. Nonetheless, it might also be the case that sometimes scientists do not restrict adequately their possibility judgments. They may incorrectly consider some world propositions to be possible. One benefit of the proposed account of HPEs is that it renders explicit the fact that possibility judgements are made. This can allow to probe the interpretation of the modal operator—e.g., causal possibility or epistemic possibility—as well as the evidence provided for the judgements. If one considers that a model provides evidence for a possibility claim, one needs to justify why she thinks so. Knowing what sort of claim is put forward is a first and significant step in its assessment.

Generalizing van Riel's argument about impossibilities to cases of modelling impossible targets (Weisberg 2013, sec. 7.2.2) may however raise another challenge for my account. For instance, Weisberg gives the example of a model of a perpetual motion machine. Perpetual motion machines are nomically impossible. For the machine to work, the physical laws would need to be different. These models can't of course provide HAEs, but they also can't provide HPEs, or so it seems, since the target is nomically impossible.

Yet, as I mentioned earlier in section 4.6, it appears that modelling impossible targets is both a common practice and one that is epistemically valuable. According to Weisberg (2013, pp. 128-129), modelling impossibilities allows to learn about the contingency of actual states of affairs by investigating alternative possible histories. We can also learn their necessity by exploring what laws of nature would need to be different for a nomically impossible fact to be merely contingent. Suppose one thinks that a perpetual motion machine would be possible. Or simply suppose that one does not know why perpetual motion machines are impossible. By modelling one using actual physical laws, we can learn why, in fact, it is not possible. The laws nomically prevent the machine from working perpetually. We cannot have the laws of nature we have and a perpetual machine.

Knowing impossibility claims of the form '$\neg\Diamond$(*p* because *q*)' is equally valuable as knowing claims of possibility. If the proposition '*q*' is nomically impossible, '$\Diamond$(*p* because *q*)' can't be true under a nomic interpretation. Some models thus give us true propositions of the form '$\neg\Diamond$(*p* because *q*)'. Knowing possibilities is especially important on the background of beliefs of impossibility, whereas knowing impossibilities is significant on the background of beliefs of possibility (cf. Grüne-Yanoff 2009). Both should be regarded as HPEs.

When we say that a model provides a HPE, what it means is that it provides reasons to believe that '$\Diamond$(*p* because *q*)'. One important question we may then ask is whether HPEs thus defined constitute a genuine species of explanation, assuming that propositions such as '*p* because *q*' are explanations. More generally, in what relation do HPEs stand with respect to HAEs? Put differently, HPEs do not explain actuality, but do they actually explain? I think one major reason why we want to settle this question is not so much for taxonomical reasons, but rather to know whether HPEs are epistemically similar to HAEs. This is especially important in the context of the puzzle of model-based explanation.

One reason we may want to reserve the genus 'explanation' for what we have so far called HAEs is because HPEs are unsatisfactory answers to explanation-seeking why-questions. If I ask why the financial crisis of 2007-8 occurred and one answers me in return that it was possibly caused by the greed of bankers, I will not be fully satisfied with the answer. I still will not know what was *actually* responsible for the crisis, which was the information my question asked for. Denying the explanation status to HPEs thus sends the clear signal that HAEs are what we ultimately care about. HPEs, all worthwhile they are and perhaps the best we can achieve in certain circumstances, do not actually explain phenomena.

Van Riel (2015, p. 3846) makes a similar point when he argues that since HPEs lack the actual 'because' explanatory entailment usually associated with explanations, then they should not be considered as such. Furthermore, it may also be philosophically simpler since we do not have to provide an account of what it is to 'explain non-actuality'. Indeed, if HPEs are a genuine species of explanation, additional analysis would be needed to determine what is the explanans and explanandum of these explanations. However, even if we accept that HPEs are not HAEs—an uncontroversial point—, another crucial question that arises is whether HPEs make epistemic contributions comparable to HAEs. And if they do not, is it then appropriate to call these both 'explanations'?

A reason we may want to say that HPEs are genuine explanations is because they may, like HAEs, afford understanding. An important epistemic goal of science is understanding. Science affords understanding of reality. One way it does so is by explaining phenomena. As van Riel notes, a prime candidate argument for the view that HPEs afford understanding is to consider they are genuine explanations. For if they explain, then surely they also afford understanding.

But, as I argued in chapter 4, it is also possible to make an argument to the effect that HPEs, while not actually explaining, afford understanding. Having a scientific explanation is a sufficient condition for scientific understanding. But is it necessary? And if propositions afford understanding, does this imply they should be qualified as an explanation? Put differently, that propositions explain is a cue they afford understanding. However, if propositions afford understanding, is it equally a cue that they constitute an explanation? That HPEs may contribute to learning (Grüne-Yanoff 2009, 2013a) or understanding (Rohwer and Rice 2013; Ylikoski and Aydinonat 2014; Rice 2016) has some support in the literature. However, granting that HPEs afford understanding does not imply they should necessarily be viewed as a species of explanation. For one, Lipton (2009) argues that HPEs are a prime case of understanding *without* having an explanation. If

Lipton is right, then HPEs may afford understanding without being explanations.

The previous analysis unfortunately does not provide a straight answer about whether or not HPEs should be viewed as a genuine species of explanation. What it does, however, is spell out the commitments that come with using the notion of HPEs. On the one hand, if HPEs are explanations, then it accounts for the fact that they explain or are explanatory. This would provide a straightforward answer to what sort of epistemic contribution models providing HPEs make. However, we would need to motivate further why they should share the genus 'explanation' with HAEs whose contribution is in many respects superior. On the other hand, if HPEs are not a species of explanation, this does not necessarily imply they can't afford, for instance, learning or understanding. Nevertheless, more would need to be said concerning the value of knowledge of possibility. In particular, we need to understand better the conditions under which it can inform us about the world.

5.6 CONCLUSION

Although prima facie obvious, that what demarcates HPEs from HAEs is the modal information they provide has been obscured in previous discussions over HPEs. This single criterion is both simple and clear while also acknowledging the idea that HPEs make a genuine epistemic contribution in the form of knowledge of possibility.

When all is said and done, both the Dray and Hempel-types proponents were on the right track concerning many features of HPEs. Supporters of the Dray-type were right to draw attention to the fact that HPEs may make a genuine epistemic contribution. The crucial difference between HAEs and HPEs is the end product, i.e., their particular propositional content, viz. claims of possible or actual explanations. Hempel-type advocates were right to downplay the difference between HPEs and HAEs insofar as HAEs also provide claims of possibility. In some cases, we may lack the empirical support to establish a claim of actuality, but may have enough for a possibility claim. In terms of form, they are not as radically different as the Dray-type would have liked us to believe.

This characterization of HPEs makes plain what it is for a model to provide a HPE: it provides evidence for propositions of the form '$\Diamond$(*p* because *q*)'. Knowing the general form of HPEs and how models relate to them allows to assess with more precision the contribution of particular models. Some models provide reasons to believe the explanans is possible, others the explanandum, or they may even provide evidence about their impossibility.

However, some questions remain open. Whether HPEs are a genuine species of explanation is one of them. At any rate, if we want the term 'explanation' to be reserved to HAEs—and there might be good reasons for this—, then we should better find a less confusing and more appropriate one for HPEs. While it appears safe to claim that HPEs may afford epistemic benefits in the form of learning or understanding, more work needs to be done in order to clarify how exactly knowledge of possibility can do it. But at least we now know better where to look and what to look for.

6 NON-CAUSAL UNDERSTANDING WITH ECONOMIC MODELS: THE CASE OF GENERAL EQUILIBRIUM

6.1 INTRODUCTION

In this chapter I am concerned with economic models in particular. Whether economic models explain economic phenomena and thus provide understanding is a contentious debate (e.g. Hausman 1992b; Alexandrova 2008; Grüne-Yanoff 2009; Mäki 2009a; Sugden 2009; Rice and Smart 2011; Reiss 2012b; Alexandrova and Northcott 2013; Hindriks 2013; Kuorikoski and Ylikoski 2015; Ylikoski and Aydinonat 2014). As we have seen in the previous chapter, some proposals suggest that economic models afford understanding by way of providing HPEs (e.g. Aydinonat 2007; Grüne-Yanoff 2009, 2013b; Rohwer and Rice 2013; Ylikoski and Aydinonat 2014).

One underlying idea behind these accounts is that HPEs, while not actually explaining anything, still provide causal knowledge that affords understanding. For instance, a model may show a possible cause of a phenomenon. It appears such accounts are successful in solving the puzzle of model-based understanding. However, models that appear to neither provide actual nor possible *causal* knowledge have not been discussed. As the preceding two chapters showed, there is a growing literature on non-causal HAEs (e.g. A. Baker 2012; Baron et al. 2017; Batterman 2002; Batterman and Rice 2014; Lange 2013, 2017; Pincock 2015; Reutlinger and Saatsi 2018a**Saatsi:2016aa**) and on causal HPEs (e.g. Craver 2006; Forber 2010; Reydon 2012; Grüne-Yanoff 2013a; Bokulich 2014; Ylikoski and Aydinonat 2014; Rice 2016), but not on *non-causal* HPEs.

Kenneth Arrow and Gérard Debreu's (1954) seminal model of general equilibrium therefore poses an interesting challenge to philosophical and methodological accounts of models.[1] Since the model's main contribution is a purely mathematical result, viz. two existence theorems, it seems we should conclude that the model can't afford understanding of economic phenomena. And yet, it does provide understanding according to economists, despite not providing a causal account of the economy:

> Not only have Debreu's works contributed to mathematical economics; they have contributed to the science of economics as a whole. [...] Even today, the conceptual framework offered by general equilibrium analysis is continuing to add to our understanding of economic phenomena–and sometimes showing us how little we really understand (Varian 1984, pp. 4-5).

> Who can read them [Arrow-Debreu] without finding their understanding of market power and the role of markets deepened (Hahn 1985, p. 20)?

> Therefore, whether you are a student, an economic writer, a journalist, a policymaker, or an interested citizen, if you want to understand modern macroeconomics, you have to have passing familiarity with the basic structure and properties of the ADM [Arrow-Debreu-McKenzie model[2]] model [...] (Athreya 2013, p. 33).

> But it was not until long afterward that this system of equations [Walras's] was scrutinized to ascertain

1 That Arrow and Debreu made a significant contribution to economics is rarely disputed from within the discipline. The public recognition and the amount of followers their work gathered are momentous. They are both laureates of the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel. Arrow and Debreu (1954) is considered the "centerpiece" of the formalist revolution of the 1950s in economics (Blaug 2003).

2 McKenzie (1954) also provided a proof of the existence of a general equilibrium. He too used Kakutani's (1941) theorem, but his result was less general because assumptions were made on demand functions instead of preferences (Geanakoplos 2008).

> whether it had an economically meaningful solution, i.e., whether this theoretical structure of vital importance for understanding the market system was logically consistent (Nobelprize.org 2014).

From the perspective of economics, that the Arrow and Debreu (1954) model made a major contribution to understanding has clear support. However, their work was also the target of very harsh criticisms from economic methodologists, economists, and philosophers alike. Commenting on the type of economic research Arrow and Debreu contributed to establish, Blaug said that "[t]he result of all this is that we now understand almost less about how actual markets work, than did Adam Smith or even Léon Walras" (2002, pp. 38-39).[3] There is, to say the least, an enormous discrepancy between considering the Arrow-Debreu model worthy of receiving two Nobel prizes and considering that we now understand *less*. The de facto significance of their work is at odds with some of its critical appraisal. Following the Arrow-Debreu model, do we understand less, or more?

I aim to answer this question. Relying on insights from the previous chapters, especially chapters 2, 4, and 5, I offer an example of theoretical modelling that yields knowledge of *non-causal* dependence. More precisely, I argue that the Arrow-Debreu model provides knowledge of mathematical dependence and affords understanding. Basically, the model shows that, provided certain assumptions are made concerning a competitive economy, there exists a solution to the set of equations, thereby proving the existence of a general equilibrium.

As I will argue, the model, while not exhibiting any actual nor possible causal factors, nevertheless establishes mathematical results that solved a long-standing puzzle of economic theory. These results indicate how the solution mathematically depends on the assumptions made. In short, the model provides a how-possibly *mathematical* explanation of the general equilibrium. This HPE, I argue, affords understanding not only of the model itself, but also of the world. The account developed reveals how mathematical knowledge can inform claims about the world,

3 Hands (2001, p. 294) also reports that for economic methodologists with Lakotosian leanings, general equilibrium theory perfectly exemplifies the problems of contemporary economics.

allow w-inferences, and thus improve our understanding. In what follows, I trace the origins of the general problem Arrow-Debreu addressed in section 2. In section 3 I briefly present the model. Section 4 introduces an account of model-based understanding. In sections 5 and 6, respectively, I discuss why we should consider the Arrow-Debreu model to be a mathematical how-possibly explanation and how it affords understanding of the world.

6.2 GENERAL EQUILIBRIUM, FROM SMITH TO ARROW-DEBREU

The idea behind the concept of general equilibrium can be traced back to Adam Smith's ([1776] 1904) famous metaphor of the invisible hand (Arrow and Hahn 1971; Arrow 1974; Debreu 1984, 2008; Geanakoplos 2008).[4] Smith's crucial insight was to see that the unintended consequence of everyone's pursuit of their self-interest can coincide with the promotion of the collective good. Individual participation in the economy, even though motivated by self-interest, is "led by an invisible hand to promote an end which was no part of his intention" (Smith [1776] 1904, p. IV.2.9). Explaining social phenomena as being the unintended consequences of individuals' actions has since Smith been a central theme of economics (Ullmann-Margalit 1978; Aydinonat 2008). As a contemporary economics textbook introduces its subject to its readers, "[o]ne of our goals in this book is to understand how this invisible hand works its magic" (Mankiw 2012, p. 11).

At the time of *The Wealth of Nations*, Smith's insight was more of a conjecture than anything else (Stiglitz 1991). Indeed, he did not provide any kind of mathematical argument for why we should believe a mechanism or process such as the invisible hand would work in the first place. His argument was very informal and not rigorous by modern standards (Backhouse 1998). As Arrow and Hahn (1971) note, common sense suggests it is impossible to derive social benefit from purely self-interested motives. Yet, the

4 The origins of Smith's idea can be found in earlier works, most famously Mandeville's ([1724] 1988) *Fable of the Bees*. Smith's formulation has been the most influential.

invisible-hand hypothesis was sufficiently convincing as to foster further investigations.[5]

The invisible-hand hypothesis can be understood as stating that the interaction of individuals in a market economy will result in a situation that is beneficial for all. The modern interpretation of Smith's invisible hand is the first fundamental theorem of welfare economics (see Arrow 1951; Debreu 1959), which says that "*any competitive equilibrium is necessarily Pareto optimal*" (Mas-Collell et al. 1995, 308, emphasis in original).[6] The market for a commodity is in equilibrium when its supply equals its demand. A partial equilibrium is one in which the supply equals the demand in one market. But since prices in one market ultimately depend on prices in other markets, that raises the question whether there is an equilibrium in all markets. The theorem says that a competitive equilibrium, a certain interpretation of a competitive economy where there is an equilibrium in all markets, would lead to a Smith-like situation where private interests mutually coincide. Crucially, the theorem says that *if* there is such a thing as a competitive equilibrium, then it will be Pareto optimal.

However, whether such an equilibrium can actually exist is a deeply related but separate question. For if the equilibrium could not exist, then it would be a significant blow to the invisible hand hypothesis. If the economy can't reach equilibrium, it is more difficult to maintain that exchange in a market economy can come to a point where both producers and consumers are content. A major issue stemming from Smith's idea therefore concerns the existence of equilibrium insofar as the first fundamental

5 The oldest usage I could find of the 'invisible-hand hypothesis' label is due to West (1970). It is now a more common way of talking of Smith's invisible hand (e.g. Reiss 2013b).

6 Some (e.g. Blaug 1996, 60, chap. 1, sec. 31; Backhouse 1997, pp. 128-130, 1998, p. 1853; Reiss 2013b, chap. 12) have pointed out—rightly so—that the precise formulation of the problem that preoccupied Smith underwent a significant transformation under and since Arrow-Debreu. However, the fact of the matter remains that practitioners view the general equilibrium model as a significant contribution to the research agenda that Smith set. Smith was much more concerned with the process that leads to market outcomes than with the outcomes themselves. In contrast, as Hahn puts it, "GE [general equilibrium theory] is strong on equilibrium and very weak on how it comes about" (1984, p. 140). See also Blaug (1992, p. 163).

theorem of welfare economics suggests that there is indeed such an equilibrium.

An important first step into making the equilibrium argument more rigorous was made by Léon Walras (1954) in his *Elements of Pure Economics*.[7] Whereas Smith's argument was essentially informal, Walras constructed a mathematical model. According to Debreu (2008), only a mathematical model such as Walras's could fully account for the interdependence of all the variables a general equilibrium presupposes. As Debreu explains, Walras's model still did not answer a fundamental question concerning the general equilibrium, i.e., whether such an equilibrium existed or not. The model rather assumed that there was a set of prices and quantities where supply would equal demand. The main argument supporting that belief was to count the equations and the unknowns in the model and to observe that they were in equal number (Debreu 2008).[8] But as Debreu noted, that argument was in no way definitive. In other words, it was not a proof. Yet, it was common belief that the equilibrium—the set of prices and quantities where markets clear—could exist despite the absence of a definitive proof.

Arrow and Hahn (1971) explicitly linked investigations of general equilibrium to the need to analyze logically the idea that uncoordinated economic action can result in an equilibrium. From Smith onwards there was a widespread belief that something akin to a general equilibrium could in principle exist, but that belief was not fully warranted. The question of the existence of the equilibrium was therefore by and large unanswered. The Arrow-Debreu model was a seminal contribution to answering this question. The existence problem only found a definitive solution with Arrow-Debreu's proof. Arrow and Debreu (1954) "appeared to bring closure to an argument that was at least two centuries old" (Weintraub and Gayer 2001, p. 421).

7 The book was originally published in French in 1874.

8 Blaug (1996, chap. 13) notes that Walras also tried to show by which mechanism the equilibrium could actually be achieved—the *tâtonnement* process—which in a sense can also be an argument in favour of the existence of general equilibrium.

6.3 THE ARROW–DEBREU MODEL

Arrow and Debreu (1954) is a paradigmatic example of a mathematical model in economics. It is highly mathematical and its connection to real-world phenomena appears to be at best indirect. Its main results "are two theorems stating very general conditions under which a competitive equilibrium will exist" (Arrow and Debreu 1954, p. 266). To establish these results, Arrow and Debreu used the axiomatic method. Essentially, the axiomatic method consists in building a mathematical system from basic propositions or postulates, namely the axioms. The axioms are taken as the building blocks of the system and they are then used to deductively derive the theorems, i.e., propositions that are not as basic as the axioms but are logically entailed by them. The system thus built has a "mathematical form that is completely separated from its economic content" (Debreu 1986, p. 1265).[9] The logical consistency of the mathematical structure is independent of its economic interpretation.

Using the axioms, some additional assumptions, and mathematical techniques—e.g. rules of inference—they made deductions about the logical properties of the system, deductions that are, let me repeat it, in principle independent of any empirical content. Arrow-Debreu defined a competitive equilibrium by four conditions:

1. y_j^* maximizes $p^* \cdot y_j$ over the set Y_j, for each j (Arrow and Debreu 1954, p. 268)
2. x_i^* maximizes $u_i(x_i)$ over the set $\{x_i \mid x_i \,\epsilon\, X_i,\ p^* \cdot x_i \leqq p^* \cdot \zeta_i + \sum_{j=1}^{n} \alpha_{ij} p^* \cdot y_j^*\}$ (p. 271)
3. $p^* \,\epsilon\, P = \{p \mid p \,\epsilon\, R^l, p \geqq 0, \sum_{h=1}^{l} p_h = 1\}$ (ibid.)
4. $z^* \leqq 0$, $p^* \cdot z^* = 0$ (ibid.)

where R^l is the set of all vectors with l components, y_j denotes a production plan for the production unit j in the set Y_j of production plans, x_i the consumption x of a consumption unit

9 Debreu (1984, p. 405), in his Nobel lecture, similarly explains what is axiomatization and how it was linked to the coordination problem Adam Smith apprehended.

i in the set X_i of consumption vectors, $u_i(x_i)$ the utility derived for unit i from consuming x, z is a vector whose components are the excess demand over supply for the commodities, and the asterisks (*) equilibrium values, e.g., p^* is an equilibrium price vector.[10]

To establish the existence of the equilibrium, Arrow-Debreu first needed to provide precise assumptions about a competitive economy. Conditions 1 and 2 basically express the idea that producers (1) and consumers (2) maximize their profits and utility, respectively. Condition 3 says that prices are non negative and not all zero and condition 4 normalizes prices. Conditions 1 and 2 are the ones most clearly connected to the Smithian idea of individuals pursuing their self-interest. The competitive model formalizes this pursuit since one important motivation behind the model is to see whether it is compatible with a general equilibrium. They then made the following assumptions about the production possibilities and consumers' preferences. Assuming $Y = \sum_{j=1}^{n} Y_j$, then Y represents all the possible production plans Y and $(-Y)$ contains the additive inverse components of Y. Ω is the non-negative orthant of R^l (Weintraub 1983, p. 35), i.e., a set for which all components are part of R^l and are greater or equal than 0. u_i indicates the utility of the ith consumption unit. ζ_i denotes the vector of initial endowment of various commodities held by the ith consumption unit and α_{ij} is a contractual claim to each production unit j's share of profit.

I.A. Y_j is a closed convex subset of R^l containing 0 $(j = 1, \ldots, n)$. (Arrow and Debreu 1954, p. 267)

I.B. $Y \cap \Omega = 0$ (ibid.)

I.C. $Y \cap (-Y) = 0$ (ibid.)

II. The set of consumption vectors X_i available to individual i $(= 1, \ldots, m)$ is a closed convex subset of R^1 which is

10 For a more exhaustive description or reconstruction, readers are invited to consult the Arrow and Debreu (1954) paper itself, Weintraub (1983) or Geanakoplos (2008). I am here first and foremost concerned with their methodology and the main results they achieved. The exposition of the model will thus be minimal.

bounded from below; i.e., there is a vector ξ_i such that $\xi_i \leqq x_i$ *for all* $x_i \epsilon X_i$. (p. 268)

III.A. $u_i(x_i)$ is a continuous function on X_i. (p. 269)

III.B. For any $x_i \epsilon X_i$, there is an $x_i' \epsilon X_i$ such that $u_i(x_i') > u_i(x_i)$. (ibid.)

III.C. If $u_i(x_i) > u_i(x_i')$ and $0 < t < 1$, *then* $u_i[tx_i + (1-t)x_i'] > u_i(x_i')$. (ibid.)

IV.A. $\zeta_i \epsilon R^l$; for some $x_i \epsilon X_i, x_i < \zeta_i$; (p. 270)

IV.B. for all $i, j, \alpha_{ij} \geqq 0$; for all $j, \sum_{i=1}^{m} \alpha_{ij} = 1$. (ibid.)

To establish whether there could be a general equilibrium, additional assumptions on the production possibilities and consumers preferences needed to be done. Assumptions I.a–c define the production plans and its constraints, namely that (I.a) production plans are convex, which rules out increasing returns to scale, i.e., productivity gains of adding inputs, that (I.b) it is impossible to get an aggregate production plan without using any input, and that (I.c) two production possibility vectors can't exactly cancel each other, i.e., one's outputs can't equal another's inputs since labour is also necessary and can't be produced. Assumptions II–III.c concern consumption behaviour. Assumption II defines the consumption plans from which agents can choose. Assumptions III.a–c posit how agents choose over the consumptions plans: III.a continuity, III.b non-satiation, and III.c convexity of agents' preferences. A continuous preference function ensures that there are no big jumps in demand from small changes in prices. Convexity means that agents prefer an average of commodities rather than an extreme of one or the other. Non-satiation indicates that there is no consumption bundle an agent would prefer to all others, another bundle could be preferred. Assumptions IV.a and IV.b concern the initial endowments. Based on the preceding conditions of a competitive equilibrium and assumptions, they then propose a first theorem, the proposition they need to prove.

THEOREM I. For any economic system satisfying Assumptions I-IV, there is a competitive equilibrium (Arrow and Debreu 1954, p. 272).

What the theorem says is that if an (economic) system satisfies assumptions I–IV, then there *exists* an equilibrium such that it satisfies conditions 1–4 of how they define a competitive equilibrium. Having submitted the theorem, they next proceed to prove it. The proof is purely mathematical. They first show that if there is an equilibrium, then it necessarily satisfies the conditions for a competitive equilibrium. They then demonstrate that the plans the agents choose from satisfy assumptions I-IV and that an equilibrium could be inferred from it, which necessarily is a competitive one. The proof thus demonstrates that assumptions I–IV are sufficient conditions for the existence, as defined, of a competitive equilibrium.

In the fourth part of the article, Arrow-Debreu proposed to establish the existence of the equilibrium, but by relaxing the assumption (IV.a) that every agent has a positive amount of every good for trading. They considered this assumption too strong—i.e., unrealistic—and thus want to see if they can derive the same result—the equilibrium—without it. Relaxing the assumption, however, comes "at the cost of making certain additional assumptions in different directions and complicating the proof" (Arrow and Debreu 1954, p. 280). Basically, they modified assumption IV.a so that IV.a′ and IV.b are incorporated in IV′ and formulated other ones (V, VI, and VII) to make the relaxation of IV.a possible.

I will not enter into the details of these new assumptions because what is ultimately important to note is not the assumptions themselves, but the fact that Arrow-Debreu tested the robustness of the existence of the equilibrium under changes of assumptions. Suffice it to say that in effect they only supposed that agents can supply at least one type of productive labour instead of owning a positive amount of everything.

THEOREM II For an economic system satisfying Assumptions I-III, IV′, and V-VII, there is a competitive equilibrium (Arrow and Debreu 1954, p. 281).

Theorem II is thus essentially the same as Theorem I (the equilibrium exists), but derived from slightly different assumptions. This shows that assumption IV.a is not necessary for the existence of the equilibrium, at least as it was initially formulated in the proof of the first theorem. In the fifth part they proved that Theorem II is true.

I started this section by saying the model was highly mathematical. However, it is not mathematical in the sense that it is idealized and represents causal patterns using an abstract mathematical formalism. It is mathematical, I contend, in the sense that the relevant facts and claims it establishes are mathematical, not causal. Yet, the model also appears to be populated by producers and consumers who act on commodities, production plans, prices, etc. Are not the relationships between these fundamentally causal? As Hahn emphasizes, "[t]he first important point to understand about this construction [Arrow-Debreu equilibrium] is that it makes no formal or explicit causal claims at all" (Hahn 1984, p. 47). It does not claim that actual economies will end up in an equilibrium nor does it describe mechanisms or processes that could generate it. However, it is motivated by what Hahn calls a "weak causal proposition" (ibid.), namely that if a sequence of economic states reach an end state, then it will necessarily be an equilibrium. This is because of the assumed economizing action of agents who will engage in exchanges until no one would prefer a different economic state. Conditions 1 and 2 above convey this weak causal proposition. This simply implies that the axioms and assumptions are not arbitrarily chosen, they are not "plucked out of thin air" (Hahn 1985, p. 12). That said, as we will see in more details in section 6, it does not imply that the main contribution of the model is causal. The motivation for the model may have been partly causal, but the axioms and results are not.

6.4 A MATHEMATICAL HPE

Chapter 5, in particular, made the case that knowledge of relations of dependence obtained through HPEs may afford understanding. Does the Arrow-Debreu model make a similar contribution to our understanding of the world? Inasmuch as typical cases of HPEs—e.g., the checkerboard model—afford understanding via a HPE that provides causal information about phenomena of interest—e.g., residential segregation—, I submit the Arrow-

Debreu model affords understanding via a HPE that provides knowledge of mathematical dependence about the general equilibrium.

6.4.1 Mathematical explanations

To say that the Arrow-Debreu model is a mathematical HPE raises an immediate issue: What is it to explain empirical phenomena *mathematically*? In section 4.4 , I argued that there are good reasons to believe causal knowledge is not necessary for explanation and discussed one specific example, the Königsberg bridges cases (Lange 2013; Pincock 2007; Reutlinger 2016). While there are still many unresolved questions concerning the exact role of mathematics in scientific explanations, there is wide support to the idea that some scientific explanations explain in virtue of the mathematical facts they cite (Lange 2017; Reutlinger 2017b).[11]

At first sight, it may appear that Sober's (1983) account of equilibrium explanation would readily fit the general equilibrium case. However, whether so-called equilibrium explanations are in fact causal is contentious (see e.g. Woodward 2003, pp. 6-7; Kuorikoski 2007; Strevens 2008, 267ff.). Since what I ultimately want to stress is that the HPE involved is essentially non-causal, I prefer to remain agnostic towards the exact status of equilibrium explanations. What is important for the point I am making is that non-causal facts, mathematical ones I hold, are responsible for affording understanding of the world.

To clarify how we could apply this idea of mathematical explanation to the Arrow-Debreu case, let us discuss a simple and intuitive example recounted by Lange (2013, p. 488):

> That Mother has three children and twenty-three strawberries, and that twenty-three cannot be divided evenly by three, explains why Mother failed when she tried a moment ago to distribute her strawberries evenly among her children without cutting any.

In this case, it is ultimately not the physical facts about the mother or the strawberries that explain why the mother failed

11 One dissenting view is Strevens (2018).

to distribute them evenly. Out of mathematical necessity, she possibly can't divide evenly twenty-three units of anything by three. One would not explain this impossibility by citing physical facts about strawberries, the children, or what she exactly did when she attempted to divide them. Of course, if the mother had had twenty-four strawberries, then she would have been able to divide them evenly between her three children. In a sense, an explanation of the mother's failure has to cite these facts.

Mathematical facts can only provide an actual explanation of phenomena if there is a mapping or correspondence between the mathematical facts and the physical ones (Pincock 2007; Bueno and Colyvan 2011; Bueno and French 2012). But this only shows that mathematical structures and properties can map on physical facts, not that the latter are essentially responsible for the failure. In this example, the relevant facts are rather mathematical, namely that there is no integer that multiplied by three equals twenty-three. The mother failed this time as she and others would also fail every single other time because of this mathematical impossibility. This simple example illustrates quite clearly that mathematical facts can explain empirical phenomena.

One feature of mathematical explanations is thus that they establish necessary conditions for their explanandum (Lange 2013; Pincock 2015). According to Lange, mathematical explanations show why a given explanandum is *more* than causally necessary. They appeal to facts that are modally stronger than causal information and constrain the causal space of possibility. They do not explain by virtue of "describing the world's network of causal relations in particular, but rather by describing the framework inhabited by any possible causal relation" (Lange 2013, p. 509). This makes plain why the mathematical facts explain the mother's failure above. Even though her failure depends in a sense on the specific physical facts of the situation, it depends *more* on the mathematical impossibility. This sort of necessitation encompasses the physical facts that could obtain. This does not imply that a given phenomenon can't also have a causal explanation, but simply that the mathematical facts necessitate it in a way that causal ones can't.

A notable difference between causal and mathematical explanations is that the latter, instead of exhibiting causal dependence between explanans and explanandum, exhibit *mathematical* dependence. There is nothing mysterious about this latter sort of dependence. Claims of causal dependence can be established by evaluating the truth of counterfactual conditionals of the form 'Had antecedent A been, consequent C would have been' on the background of a set of relevant causal facts.

Similarly, to say that C mathematically depends on A simply means that the truth of this counterfactual is evaluated on the background of the relevant mathematical facts.[12] It is in this sense that the mother's failure depends on the mathematical impossibility of dividing equally twenty-three by three. Had the mother had *twenty-four* strawberries, then she would have been able to divide them equally by three.

It is also in this sense that the existence of the general equilibrium depends on the assumptions of the model. Had the aggregate production set not been convex, thus implying increasing returns to scale (assumption I.a in sec. 3), then the equilibrium would not exist.[13]

We can see how this is connected to the account of understanding presented in the previous section. Knowing on what the failure or the equilibrium depend, we understand why it is so. Knowing the mathematical dependence allows to make w-inferences about it. We are able to answer a range of w-questions on the circumstances under which the mother would fail, or be able, to divide the strawberries, or about the existence of the equilibrium. In other words, we understand.

However, if the mathematical dependence in the strawberries case explains the failure, the same is not true for the Arrow-Debreu model. Indeed, it does not appear to provide an actual mathematical explanation of general equilibrium. It fails to do so for two reasons. Firstly, it seems it not to identify mathematical facts on which the equilibrium would necessarily depend. When mathematical conditions are only sufficient, we can't know only by appealing to them whether a phenomenon depends on them

12 Some accounts go further and reduce all modal notions of necessity and possibility to the epistemology of counterfactual conditionals (Stalnaker 1968; Lange 2005; Williamson 2007a; Kroedel 2012).

13 "Convexity really does perform a crucial role in the proof of existence; it is also a very restrictive assumption and empirically very vulnerable" (Feiwel 1987, p. 52). See also Debreu (2008).

or not. This is because it may depend on another set of sufficient conditions. Knowing that some conditions are sufficient for bringing about a general equilibrium may therefore not be very illuminating for there may be a number of other such conditions.

Secondly, it does not identify actual mathematical structures or properties that map onto the world. An actual mathematical explanation of general equilibrium would require identifying that there is a mapping between the mathematical structure and the phenomenon (Bueno and Colyvan 2011; Pincock 2007, 2012) and showing how the phenomenon depends on some necessary mathematical properties. In the case of the Arrow-Debreu model, it is rather dubious whether there is such a phenomenon as the general equilibrium in the first place and to what extent the conditions for a competitive economy are satisfied.

6.4.2 A mathematical HPE

While the Arrow-Debreu model fails to be an actual explanation of general equilibrium, we should view its contribution as being a mathematical HPE. Causal HPEs, such as the checkerboard model, establish claims of *causal* possibility. Mathematical HPEs establish claims of *mathematical* possibility. These claims, I contend, may afford understanding of the world.

To make sense of the Arrow-Debreu model's contribution to understanding, it is first important to keep in mind the nature of the problem that they were trying to solve (see sec. 6.2). As Rosenberg (1992, p. 202) observes, it was far from obvious that prices and quantities in all markets could be determined simultaneously. Yet, the invisible-hand hypothesis was nevertheless sufficiently compelling for economists to believe it. We can interpret the Walrasian tradition, Rosenberg says, as solving a 'how-possible' problem about the existence of the equilibrium. As Arrow and Hahn put it, "[t]he proposition [the invisible-hand hypothesis] having been put forward and very seriously entertained, it is important to know not only whether it *is* true, but also whether it *could be* true" (1971, vii, emphasis in original). Prior to Arrow-Debreu, economists not only did not know whether it was true, but also did not know if it could be true. There was no satisfactory answer to this how-possible question. The very possibility of the equilibrium was not demonstrated. Thus, what the Arrow-Debreu model did was not to establish that there is a general equilibrium in actual economies. What it did was establishing its mathematical possibility. It could, under the conditions specified in the model, be true. It is in this sense that the Arrow-Debreu model is a HPE.

However, even if we grant that the Arrow-Debreu model is a HPE, we could say that the relevant modality is not mathematical. The model would indicate how it could *causally* be true, not *mathematically*. For instance, one could say that the model showed how interdependent processes of economizing action by consumers and producers could lead to a general equilibrium. While I agree that the Arrow-Debreu model can ultimately bear on causal claims of this sort, it misinterprets the actual problem they were trying to solve and their contribution. The 'how-possible' problem—i.e., the existence of a general equilibrium—was inherently mathematical. Economists simply did not know whether the equilibrium characterization of the economy was formally consistent and admitted of a solution. Economists had already a causal interpretation. They were lacking the mathematical form. The novelty of the Arrow-Debreu model was not to describe unthought-of causal factors or how they interact, but to prove that within a given mathematical framework the equilibrium necessarily exists. The 'how-possible' problem they were aiming to solve was essentially a consistency check: Is there a set of consistent assumptions for which an equilibrium exists? Answering this question was just a very difficult mathematical problem which did not require describing and assessing causal factors. Koopmans makes this point very explicitly:[14]

> The test of mathematical existence of an object of analysis postulated in a model is in the first instance a

---

14 This is also consistent with how Arrow-Debreu motivate the empirical implication of the model: "The investigation of the existence of solutions is of interest both for descriptive and for normative economics. Descriptively, the view that the competitive model is a reasonably accurate description of reality, at least for certain purposes, presupposes that the equations describing the model are consistent with each other. Hence, one check on the empirical usefulness of the model is the prescription of the conditions under which the equations of competitive equilibrium have a solution" (Arrow and Debreu 1954, p. 265).

> check on the absence of contradictions among the assumptions made. If we assume that not all members of a body of contradictory statements can have empirical relevance, this logical test has to be passed before any question about the relation of a model to some aspect of reality can seriously be raised (Koopmans 1957, p. 55).

The problem of existence is a mathematical problem, not a causal one. Crucially, the existence theorems, and thus the equilibrium, do not depend on a particular specification or interpretation of causal facts, but on mathematical ones. As Debreu emphasized, there is a "divorce of form and content" (Debreu 1986, p. 1265) (see also sec. 6.3) between the mathematical structure and its economic interpretation.

Following Ylikoski and Aydınonat's (2014) terminology, we can characterize Arrow-Debreu as providing a 'mathematical scheme'. Like the checkerboard model , which does not have a specific target and can be applied to the general phenomenon of segregation, the Arrow-Debreu model lays out a mathematical framework, or template, which can be used to structure and model relationships between variables. It is not about a particular phenomenon, e.g., general equilibrium in Canada, but about general equilibrium *in general*. It is meant as a template that can be applied to study the properties of equilibrium, or the absence thereof. The various axioms, definitions, assumptions, and theorems of an axiomatic theory help us to represent economic phenomena by providing "*an empty schema of 'possible realities'*" (Ingrao and Israel 1990, 285, emphasis in original). That schema provides a mathematical framework in principle divorced from its empirical economic interpretation, but a schema that also constrains the economic (causal) space of possibility. For instance, the framework is incompatible with increasing returns to scale on the production side or with the possibility of a Pareto improvement while at equilibrium (Hahn 1984, p. 6). So while the existence of the equilibrium does not depend on a specific economic interpretation, the mathematical framework also constrains the possible interpretations.

6.5 ARROW-DEBREU AND UNDERSTANDING THE WORLD

6.5.1 Understanding the model vs the world

If we accept that the Arrow-Debreu model is indeed a mathematical HPE, one could still object that it does not provide understanding of the world, but merely of the model. A common distinction made when assessing the epistemic contribution of models is between understanding the model and understanding the world. One understands the model when one gets a better grasp of the model's implications, how some results are derived, whether some assumptions are robust, etc. Put differently, one understands the model when one knows how its results depend on its assumptions. I take it to be uncontroversial that the Arrow-Debreu model provides such understanding of the model.[15] But if it is *only* about the model, then the precedence of the existence question, as expressed by Koopmans (1957, p. 55) in the previous section, would be rather puzzling. It is indeed a question about the model, but whose answer can change the justification we have for certain beliefs about the world. Therefore, it is also, at least indirectly, a question about the world.

The contentious issue is thus how the model also provides understanding of the world. As we have seen earlier in chapter 2, one understands the world when one can make correct *w*-inferences about it. Making correct *w*-inferences requires to have knowledge of relations of dependence. To actually explain a phenomenon implies showing on what it actually depends. The analogy with causal HPEs is once again enlightening. The checkerboard model, for example, does not establish on what residential segregation actually depends. Yet, it is widely claimed that the checkerboard model improved our understanding of the world. Before the checkerboard model, we did not have reasons

15 In fact, one can find views along these lines from practitioners themselves, for instance: "The question of the existence of a competitive equilibrium is of course a question not about the world but about the model" (Koopmans and Bausch 1959, p. 120). In a sense, this is obviously correct. When economists prove the existence of a mathematical object within a set of equations, they are at least answering a question about the model.

to believe residential segregation could causally depend on the preferences for not being in a minority status. The model showed it was possible. It is a HPE.

Likewise, the Arrow-Debreu model does not show on what the equilibrium actually depends, but rather that it can possibly exist given certain conditions are met. In both cases—the checkerboard and Arrow-Debreu models—knowledge of dependence plays a crucial role in warranting beliefs about the world. Asked in an interview about the relevance of the existence problem, Arrow answered the following: "Therefore, I would like at least to be sure that it is a coherent theory [general equilibrium theory]. That does not prove that the real world is like that, but at least it gives me a chance" (1987, p. 197). The thought is that if it is not possible to provide a consistent mathematical formalization of the general equilibrium, then it is unlikely that it is an actual empirical phenomenon.

This is because, as we have seen, claims of mathematical possibility are modally stronger than those about causal possibility. Of course, mathematical possibility does not entail causal possibility. The set of possible mathematical relations of dependence is greater than the set of possible causal relations of dependence. For the same reason, causal impossibility does not entail mathematical impossibility. Some relationships may be causally impossible, yet can be mathematically represented. However, if a phenomenon is causally possible, should it not also be mathematically possible? And if it is mathematically impossible, does it imply that it is also causally impossible? As the various practitioners quoted above pointed out in slightly different terms, if one wants to know whether the invisible-hand hypothesis *is* true, an important question one may ask is whether it *could* be true. They considered that establishing a claim of mathematical possibility would inform them about the world. Thus, Arrow and Debreu (1954) provided a model that improves our understanding of the real world conditions under which we can expect, or not, a general equilibrium.

Backhouse (1997, p. 133) is right to say that our inability to rigorously prove something does not imply it is not true. As Arrow acknowledged above, economists' inability to prove the existence of the general equilibrium up until the first existence proofs (e.g. Wald [1936] 1951) did not imply that existence was impossible. The existence proof relied on mathematical tools that were developed in the decades before by, among others, Von Neumann (1945) and Kakutani (1941). The theorem by Kakutani was then used by Nash (1950) in order to establish the existence of an equilibrium in a finite game (Debreu 2008). Basically, Arrow and Debreu (1954) used the same mathematical machinery to extend that result to a *n*th-agent game, the 'economy'.

However, economists' inability to prove existence can, and was considered to be, evidence against it. As we have seen in section 5.5, models serve that evidential role when model propositions make a difference to the justification we have about world propositions (Claveau and Vergara Fernández 2015). They do so by strengthening the evidential network one has about these propositions.[16] For instance, Theorem I in Arrow and Debreu (1954) can contribute to justifying the belief that a general equilibrium is causally possible, a belief many economists had, but which was in need of further justification. Conversely, the absence of a proof, or worse, an impossibility proof, could also serve as evidence against its causal possibility. To those who perhaps believed it was impossible for the general equilibrium to exist on the ground of mathematical impossibility, Arrow-Debreu showed it was possible.[17] They had to revise their beliefs. And to those who already believed that it was possible, Arrow-Debreu strengthened their existing beliefs. The immediate goal of the model is not to actually explain why there is a general equilibrium in the world, but rather to investigate the conditions under which it *could* exist. The model establishes mathematical dependence

16 There is a difference between the checkerboard model as accounted by Grüne-Yanoff (2009) and the Arrow-Debreu model. For Grüne-Yanoff, a change of confidence can only happen when a model 1) presents a credible model that 2) contradicts an impossibility thesis. The Arrow-Debreu model is usually not considered to be credible and it does not contradict an impossibility thesis since many economists already believed that it was possible.

17 Of course, the mathematical impossibility of an object depends on which assumptions are used. Arrow's (1963) famous impossibility theorem is a good example of such an impossibility result. It is true that no rank-order voting system can aggregate individual preferences and also can satisfy some set of minimal conditions, but that does not mean that preferences can't be aggregated in general. Likewise, one could hold that the general equilibrium is impossible on the ground of assumptions that are considered to be plausible.

between a mathematical object, the competitive equilibrium, and some assumptions one can make—or has to make—to derive the object. While not actually explaining anything, its particular contribution is the determination of various mathematical conditions, sufficient and/or necessary, for the existence of a general equilibrium. Arrow-Debreu provided decisive evidence against the claim that it is mathematically impossible for the equilibrium to exist under a relevant set of assumptions.

6.5.2 Arrow-Debreu and *w*-questions

However, even if we accept that the Arrow-Debreu model is a mathematical HPE, one could say it remains unclear what *w*-questions about the world it contributes to answer. Since it is a central feature of the account of model-based understanding presented earlier, this could put pressure on the claim that the Arrow-Debreu model affords understanding of the world. The model established a set of mathematical conditions on which the equilibrium depends. These conditions constrain the space of possibility within which the general equilibrium may receive an empirical interpretation.

To give a more concrete example, in Arrow and Debreu (1954) the assumptions are only jointly sufficient for equilibrium, not necessary. However, they are not all equally dispensable. We have already seen the critical role the assumption of convexity has. To give another example, in proving Theorem II, Arrow-Debreu illustrate what kind of *w*-questions their model can allow to answer. Basically, they want to see what assumptions are needed to derive the existence of the equilibrium. They do it by relaxing assumption IV.a, which states that every individual has a positive amount of each commodity. They consider this assumption as "clearly unrealistic" (Arrow and Debreu 1954, p. 270). The proof of Theorem II, which relies on slightly different assumptions needed to relax IV.a, is an exercise the goal of which is to exhibit how the results may change under modifications of the assumptions. It shows that IV.a, at least in its initial form, is not necessary for the existence of the equilibrium. However, whereas having a positive amount of each commodity is not necessary, they assert that "to have equilibrium, it is necessary that each individual possess some asset or be capable of supplying some labor service which commands a positive price at equilibrium" (Arrow and Debreu 1954, p. 270). Changing the assumptions of the proof reveals how the theorem depends on them and allows one to answer a range of *w*-questions.

There is therefore a mathematical dependence relation between the assumptions and the truth of the theorem. But this dependence has implications for empirical interpretations. It allows, for instance, to answer the following *w*-question: 'What if agents had no assets and no possibility to sell their labour at a positive equilibrium price?' One *w*-inference we can make is that there would be no general equilibrium. This is because we know that such an economic interpretation does not instantiate mathematical properties considered necessary for the equilibrium. Answering this *w*-question is only possible because we have knowledge of mathematical dependence. This allows one to answer various *w*-questions about the world. For instance, we would know that if the world did not instantiate some conditions deemed necessary, then a general equilibrium would not, out of necessity, ensue. Conversely, we would know what features of the world would be sufficient to obtain an equilibrium. Uncovering sufficient and necessary conditions for equilibrium thus has the potential to allow us to answer a range of *w*-questions about why the equilibrium exists or not.

Looking at the cluster of models (Rodrik 2015; Ylikoski and Aydinonat 2014) using similar assumptions can also provide a much richer account of what relations of dependence are at play and of which assumptions may be necessary. The checkerboard model, for instance, has been subject to various robustness analyses (see Aydinonat 2008; Muldoon et al. 2012) to establish under what conditions segregation patterns can be expected.

In similar fashion, the Arrow-Debreu model has now been subject for decades to similar robustness tests (e.g. Arrow and Hahn 1971; Arrow and Intriligator 1982). In particular, under what conditions the equilibrium is unique (are there multiple sets of prices and quantities?) or stable (will an economy out of equilibrium come back to the equilibrium value?) have been topics of intense scrutiny. It was, for instance, a central motivation of Arrow and Hahn to study those abstract mathematical relations of dependence and see "how robust this result [the existence of general equilibrium] is" (1971, p. vii).

One goal of their book was to analyse the various conditions under which general equilibrium would still hold. Later on the same page they justify this practice the following way: "In attempting to answer the question 'Could it be true?', we learn a good deal about why it might not be true" (Arrow and Hahn 1971, p. vii). Hands (2016) claims that seeing these attempts as derivational robustness analysis (see Woodward 2006; Kuorikoski, Lehtinen, and Marchionni 2010) is an accurate description of the practice. Derivational robustness analysis involves testing the sensitivity of a model's derivations—e.g. existence of the equilibrium—to changes of its assumptions. It can allow one to see what role different assumptions play in driving the model's results. Crucially, derivational robustness analysis amounts to asking various *w*-questions about changes of assumptions (Ylikoski 2014; Kuorikoski and Ylikoski 2015). Hands also argues that preliminary successes in robustness analysis increased the credibility of the general equilibrium model. But, Hands notes, failures of robustness in some respects, especially concerning stability, have also reduced confidence in the model as the best characterization of a competitive economy.

One could then object that since the project failed to uncover robust conditions, then the Arrow-Debreu model can't afford understanding. However, as Arrow and Hahn noted above, we can also understand "why it might not be true". As it happens, for economists, the model still has some important benefits in that it "provides us with a clear benchmark against which to measure the dysfunction of the real world [. . .]" (Athreya 2013, p. 33).[18] That the model can serve as a benchmark[19] supposes that economists are fairly confident that the conditions they have for equilibrium are relevant for studying the actual world, perhaps that they are quasi-necessary or sufficiently good approximations of what would be required in the world.[20] It supposes that the

18 See also Hahn (1973) for a similar justification.
19 A role that Debreu (1986, p. 1268) recognized.
20 One way of understanding the notion of 'quasi-necessity' is in terms of necessity relative to a particular context or modality. Some conditions may be necessary in a context, but not in another. Or some 'necessary' conditions may themselves be necessitated by modally stronger modal facts (see Lange 2009).

knowledge they have concerning the conditions under which the equilibrium exists can be used, as a matter of fact, to answer *w*-questions. For instance, it gives a framework in which actual deviations like informational asymmetry can be studied (e.g. Akerlof 1970). Instead of affording understanding why there *is* a general equilibrium, the model may afford understanding why there is *not*.

Stiglitz assesses in similar terms Arrow-Debreu's contribution. According to him, their achievement was not to prove that the equilibrium generally exists, but instead it "was to find those special and limiting conditions under which the Invisible Hand theorems hold" (Stiglitz 1991, p. 18). Since we know that the world does not instantiate certain conditions specified by the model (e.g., some markets have increasing returns to scale), we may understand why there is no equilibrium. What may make the (broadly construed) Arrow-Debreu model suitable as a benchmark is thus not its causal similarity to the actual world, but that it states conditions that economists consider quasi-necessary for the equilibrium. The mathematical relations of dependence constrain sufficiently the economic space of possibility. Despite the lack of similarity to the actual world, the model may therefore allow to make *w*-inferences about real world phenomena. As Hahn aptly puts it, the aim of the model "is not realism but aid to understanding" (1985, pp. 20-21).

Finally, one might—rightfully—worry that *any* mathematical model could afford understanding of the world. This would indeed be an undesirable consequence. A good account of model-based understanding should be able to discriminate between mathematical models that afford understanding and those that do not. I addressed similar worries in section 4.6 and in chapter 5. According to my account of HPEs, they have the following form '$\Diamond$(p because q)'. Hence, it first crucial to specify the modality of the possibility operator. Then, we need to assess the truth—i.e., possibility—of the HPE. Moreover, to provide understanding a model has to allow one to answer *w*-questions about the world, as the checkerboard model does, for instance.

However, not all models can allow to answer such questions. For instance, a mathematical relation of dependence that could not receive any relevant empirical interpretation could not be used to answer w-questions about the world. Moreover, mathematical relations of dependence that would not constrain at all the space of empirical possibility would simply not exhibit on what a phenomenon depends. Even though the account is in principle liberal regarding what can afford understanding, not anything goes.

One compelling reason why the Arrow-Debreu model passes this test—that of offering a true HPE—is that economists by and large regard it as establishing a (true) claim of mathematical possibility. Not all models pass it though. For instance, Walras's (1954) model came close, but ultimately fell short of offering the mathematical HPE economists sought. One might object that practitioners can be wrong. This can indeed be the case, but as I argued in section 1.1.2, it is also to some extent undesirable for a philosophical account to conflict with practitioners' judgements. That the Arrow-Debreu model actually affords understanding of the world is a good explanation of why practitioners consider it does. In turn, that the model provides a mathematical HPE is, I contend, a good explanation of why the model affords said understanding.

6.6 CONCLUSION

There is now a vast literature on non-causal HAEs as well as on causal HPEs. My goal with this chapter was to fill a gap in the literature by discussing a case of what I take to be a non-causal HPE, namely the Arrow-Debreu (1954) model of general equilibrium.

The model, in the eye of economists, affords understanding. I argued that it affords understanding understanding in virtue of being a *mathematical* HPE. Being so, the model provides knowledge of mathematical dependence that can be used to answer various questions about the real world. Viewing the Arrow-Debreu model as providing knowledge of mathematical dependence allows to interpret it as making a genuine contribution to our understanding despite the fact that it does not provide us with knowledge of causal dependence.

This interpretation contributes to elucidate partly the discrepancy between the different appraisals of the model's contribution to understanding. The model does not provide causal knowledge, yet its contribution to our knowledge, mathematical in this case, is difficult to dispute. That knowledge, in turn, also affords understanding of the world we live in.

7 CODA

So, to respond to one of my supervisor's worry, did I give economics a free pass to absolution? I have tried to provide an epistemological framework that clarifies how we should think and reason about the epistemic contribution of theoretical modelling. As far as I am concerned, the absolution—or condemnation—is still up for debate.

As much as I endeavoured to come up with definitive answers about the 'explanation paradox', it might also give the impression that I raised more questions than I have answered. As long as we can think more clearly about what are the truly significant questions to answer, I will consider that an important desideratum has been fulfilled.

There are fours main areas where I think further research would be particularly important and helpful.

7.1 MEASURING UNDERSTANDING

One first area in need of additional investigation concerns the evaluation, or we could say *measurement*, of understanding. In chapter 2, I raised various doubts about InfBUn viewed as a substantive account of understanding. I argued that the evaluative interpretation was preferable.

However, although that second interpretation does not face the same crucial objections, it is still wanting in two main respects. First, one feature of knowledge or understanding that the epistemological literature emphasizes is that they are a cognitive achievement *because of* ability. However, that relation 'because of' relation has proven to be very elusive. Shall we construe it in causal-explanatory terms (e.g. Greco 2010) or in strictly dispositional ones (e.g. Sosa 2007)? And what are the relevant *abilities* we should deem responsible for understanding? What are the cognitive processes that underlie these abilities?

Furthermore, it seems this is not only a philosophical question, but also an empirical one. For instance, many suggest that the notion of 'grasping' captures an important feature of genuine understanding (e.g. Grimm 2014; Strevens 2013). But, as Trout notes, "[i]f we had a psychologically informed and empirically rigorous account of 'grasping', that would go a long way toward characterizing understanding. But we don't" (2018, p. 238). Empirical work could definitely inform our philosophical theories of understanding. In the meantime, it is not clear how philosophy can contribute to clarify the 'because of' relation without more empirical input.[1]

Second, we simply lack explicit, accepted, and exhaustive metrics of evaluation. Actually performing *w*-inferences may be very good evidence for understanding across scientific contexts. But it can't be the whole story. Shall we only look at behaviour? Probably not. And is making inferences the only behaviour that matters? Here, the existing work on explanatory power may be a fruitful avenue of inquiry (e.g. Cohen 2016; Lipton 2004; Schupbach and Sprenger 2011; Ylikoski and Kuorikoski 2010). Since explanation and understanding are intimately linked, we may expect measures of explanatory goodness and of understanding to be related. However, if understanding and explanation also come apart, as I hold they do, then we should also expect the metrics to not be fully similar.

This also raises the question as to *who* should decide what makes a good explanation or understanding. Philosophers have traditionally monopolized that discussion using good old conceptual analysis. But there is also experimental work that studies folk attributions of understanding and explanation (e.g. Wilkenfeld, Plunkett, et al. 2016; Wilkenfeld and Lombrozo 2018; Wilkenfeld, Plunkett, et al. 2018). What is the relevance of these results for our philosophical theories of understanding? Shall we trust philosophers' intuitions about how best to evaluate understanding? Growing work in experimental philosophy suggests we should not (see Machery 2017). In any case, there is room for a thorough investigation of these questions.

1 Trout (2018), for one, is sceptical.

7.2 BROAD KNUN, HPES, AND MODALITY

Broad KnUn, the epistemology of understanding I argued for in chapter 4, relies on the notion of 'possibility'. The account of HPEs I presented in chapter 5 also uses the modal notions of possibility and actuality to demarcate HPEs from HAEs. According to that account, HPEs are simply propositions of the form '$\Diamond$(p because q)'. These propositions can also be true of false. This supposes that:

1. We can specify the relevant interpretation of the modality operator.
2. We can indeed evaluate the truth of possibility claims.

As I acknowledged throughout the thesis, my proposal deliberately left open these issues. My goal was to convince the reader that a key distinction we needed to make when assessing the epistemic contribution of models lay in the distinction between actuality and possibility.

However, a truly complete and general account of broad KnUn and of HPEs would need to answer these questions. Accordingly, I believe that one fruitful avenue of inquiry would consist in unpacking the various interpretations and truth conditions of the possibility operator. For instance, are some modal interpretations prior or stronger than others (Kment 2014; Lange 2009)? What are the most frequent or adequate interpretations of modality in the sciences? Can all kinds of modality afford understanding, or only a subset of them (e.g. causal possibility)?

Provided we can map out the different sorts of modality, we still need to know how to assess their truth. Whether or under what conditions we have epistemic access to modal truths is contentious (Machery 2017; Williamson 2007b). Williamson provides one potential line of defence. For him, modal judgments are just of the same type as ordinary instances of counterfactual judgments. Since we have good reasons to think we can evaluate the truths of some counterfactuals, then that we can assess the truth of some HPEs is not that far-fetched. However, we are nevertheless still far from having a comprehensive account of that works across all modalities. For example, what sort of evidence do we need to judge that something is causally possible?

Answering all these questions would have required delving into the epistemology of modality and other deep metaphilosophical issues. While this was beyond the scope of this thesis, it would certainly be important to address them.

7.3 MODELS AS EVIDENCE

Another area concerns the evidential role of models (see Claveau and Vergara Fernández 2015; Grüne-Yanoff 2009). Although I found the evidential view a very fruitful way of approaching the epistemic contribution of models, I was surprised by the resistance I faced when discussing with fellow colleagues. Indeed, there is this widespread conception that the models themselves somehow *are* similar to their target or that *they* are the explanation.[2]

As I explained, I rather consider that HAEs or HPEs are world propositions and that building and manipulating models—and their model propositions—change the justification we have in the world propositions. This moves the focus away from talking about, for instance, idealizations or scientific representation. While we may still need, for instance, to talk about idealizations to assess the evidence, it nonetheless changes the focus.

All that said, we do not currently have a fully-fledged account, pace Claveau and Vergara Fernández (2015), of how models can serve that evidential role. Under what conditions models may serve as evidence for world propositions like explanations? How should we assess contextual features? A piece of evidence is always evidence on a given background.

Also, models usually do not come alone, but in families or clusters: what is thus the proper unit of analysis (Aydinonat 2018; Vergara Fernández 2018; Ylikoski and Aydinonat 2014)? Likewise, how is the evidence socially conceived and processed? Science is a social institution and thus it seems social epistemology could inform our account of models as evidence.

2 This position is different than the so-called ontic conception of explanation (see Wright 2015). The confusion lies in taking model propositions at face value and thus carrying the explanatory work.

Many of my arguments relied on the descriptive accuracy of the premise that models actually do serve an evidential role. While I think this was a reasonable and modest claim, there is still work to be done on the descriptive and normative fronts.

7.4 NON-CAUSAL GENERALIZATIONS

A final area where I believe we could, no, *need* to make significant progress concerns the distinction between causal and non-causal explanation. As I showed in chapter 4, inquiries about non-causal forms of explanation are thriving (e.g. Lange 2017; Reutlinger and Saatsi 2018a). I also argued in chapter 6 that the Arrow-Debreu (1954) model of general equilibrium provided a *mathematical* HPE. I made that argument on the basis of current accounts of non-causal explanation.

However, we lack a thorough understanding of what demarcates non-causal generalizations (or facts) from causal ones (Reutlinger 2016). Causal explanations explain and afford understanding in virtue of the causal generalizations they use. In the same way, non-causal explanations explain and afford understanding on the basis of non-causal generalizations. Then, what are their respective features? Do we need an account that distinguishes between different layers and strengths of necessities (Lange 2017), or shall we instead refine the counterfactual framework to accommodate mathematics (Baron et al. 2017)?

In any case, that understanding is crucial if we want to adequately delineate the different types of explanation. Moreover, this can only enrich our grasp of what role mathematics play in our understanding of phenomena.

We may not have all the answers about understanding with models. At least, I hope we now know better what next questions to ask.

SAMENVATTING (SUMMARY)

Het gebruik van theoretische modellen brengt veel epistemologische uitdagingen met zich mee. Modellen lijken economische fenomenen te verklaren, maar er lijken ook cruciale kenmerken aan te ontbreken die theorieën over wetenschappelijke verklaring gewoonlijk vereisen. Ze lijken bijvoorbeeld geen getrouwe weergave te geven van de relevante causale factoren.

Mijn proefschrift beoogt een antwoord te geven op de volgende centrale en algemene vraag: wat kunnen we leren van theoretische modellen? Het doet dit door aan te tonen dat modellen epistemische voordelen kunnen bieden in de vorm van begrip, zelfs wanneer ze niet echt verklaren of wanneer ze geen causale kennis opleveren.

Het tweede hoofdstuk van mijn proefschrift kijkt naar de inferentialistisch-behavioralistische theorie van begrip en stelt deze theorie geen onderscheid kan maken tussen het gevoel van begrip en echt begrip. Ik stel dat we deze theorie kunnen zien als een evaluatieve theorie van begrip in plaats van een inhoudelijke. Het derde hoofdstuk analyseert een mogelijke oplossing voor het probleem van de representatie, namelijk een oplossing die gevonden kan worden in de feitelijke versie van inferentialisme. Ik beargumenteer echter dat deze theorie geen onderscheid kan maken tussen verklarende representatie en slechts fenomenologische representatie. Ik concludeer dat het inferentialisme dus voor een dilemma staat. Het vierde hoofdstuk schetst wat ik de 'smalle kennistheorie van begrip' noem en toont vervolgens aan dat de twee centrale claims die deze maakt, namelijk dat men alleen iets begrijpt als 1) men kennis heeft van oorzaken en 2) dat deze kennis gevormd wordt door een verklaring, onhoudbaar zijn. In het vijfde hoofdstuk wordt een nieuwe theorie van 'how-possibly' verklaringen ontwikkeld, waarin wordt beweerd dat wat deze verklaringen onderscheidt van 'how-actually' verklaringen is dat de eerste kennis van de mogelijkheid verschaft, terwijl de laatste kennis van de actualiteit verschaft. Het zesde en laatste hoofdstuk is een gedetailleerde case study van de algemene evenwichtstheorie in de economie. Ik beargumenteer dat het model een niet-causaal begrip biedt via een wiskundige how-possibly verklaring.

SUMMARY

Theoretical modelling raises many epistemological challenges. Models appear to explain economic phenomena, yet they also seem to lack crucial features theories of scientific explanation usually require, for instance faithfully representing causal factors of interest.

My doctoral thesis aims at answering the following central and general question: what, if anything, can we learn from theoretical models? It does so by showing that models may provide epistemic benefits in the form of understanding even when they do not actually explain or when they do not provide causal knowledge.

The second chapter of my thesis examines the inferentialist-behavioural account of understanding and argues that it does not allow to distinguish between the sense of understanding and genuine understanding. I submit that we can view it as offering an evaluative account of understanding instead than a substantive one. The third chapter analyses one purported solution to the problem of representation, namely the factive brand of inferentialism. I argue the account can't distinguish merely phenomenological from explanatory representation and conclude by presenting a dilemma inferentialism faces. The fourth chapter characterizes what I call the narrow knowledge account of understanding and then shows that its two tenets, i.e., that one understands only if 1) one has knowledge of causes and 2) that knowledge is provided by an explanation, are untenable. The fifth chapter develops a novel account of how-possibly explanations which claims that what demarcates these explanations from how-actually explanations is that the former provide knowledge of possibility whereas the latter provide knowledge of actuality. The sixth and final chapter is a detailed case study of the general equilibrium case in economics. I argue that the model offers non-causal understanding via a how-possibly mathematical explanation.

BIBLIOGRAPHY

Achinstein, Peter

1983 *The Nature of Explanation*, Oxford University Press, New York. (Cit. on pp. 3, 80.)

Ainslie, George

2001 *Breakdown of Will*, Cambridge University Press, Cambridge. (Cit. on p. 122.)

Akerlof, George A.

1970 "The Market for "Lemons": Quality Uncertainty and the Market Mechanism", *Quarterly Journal of Economics*, 84, 3, pp. 488-500. (Cit. on pp. 134, 166.)

Alexandrova, Anna

2008 "Making Models Count", *Philosophy of Science*, 75, 3, pp. 383-404. (Cit. on pp. 9, 12, 143.)

Alexandrova, Anna and Robert Northcott

2009 "Progress in Economics: Lessons From the Spectrum Auctions", in *The Oxford Handbook of Philosophy of Economics*, ed. by Harold Kincaid and Don Ross, Oxford University Press, New York, pp. 306-336. (Cit. on p. 36.)

2013 "It's Just a Feeling: Why Economic Models Do Not Explain", *Journal of Economic Methodology*, 20, 3, pp. 262-267. (Cit. on pp. 9, 12-13, 52, 143.)

Arnott, Gareth and Robert W. Elwood

2008 "Information Gathering and Decision Making about Resource Value in Animal Contests", *Animal Behaviour*, 76, 3, pp. 529-542. (Cit. on p. 93.)

Arrow, Kenneth J.

1951 "An Extension of the Basic Theorems of Classical Welfare Economics", in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, ed. by Jerzy Neyman, University of California Press, Berkeley, pp. 507-532. (Cit. on p. 147.)

1963 *Social Choice and Individual Values*, Second, Cowles Foundation for Research in Economics, Yale University. (Cit. on p. 162.)

1974 "General Economic Equilibrium: Purpose, Analytic Techniques, Collective Choice", *The American Economic Review*, 64, 3, pp. 253-272. (Cit. on p. 146.)

1987 "Oral History I: An Interview", in *Arrow and the Ascent of Modern Economic Theory*, ed. by George R. Feiwel, Macmillan Press, Basingstoke, pp. 191-242. (Cit. on p. 161.)

Arrow, Kenneth J. and Gerard Debreu

1954 "Existence of an Equilibrium for a Competitive Economy", *Econometrica*, 22, 3, pp. 265-290. (Cit. on pp. 17, 24, 144-145, 148-152, 158, 161-164, 167, 173.)

Arrow, Kenneth J. and Frank Hahn

1971 *General Competitive Analysis*, Elsevier, Amsterdam. (Cit. on pp. 146, 148, 157, 164-165.)

Arrow, Kenneth J. and Michael D. Intriligator

1982 (eds.), *Handbook of Mathematical Economics*, North-Holland, Amsterdam, vol. 2. (Cit. on p. 164.)

Athreya, Kartik B.

2013 *Big Ideas in Macroeconomics: A Nontechnical View*, MIT Press, Cambridge, MA. (Cit. on pp. 144, 165.)

Aydinonat, N. Emrah

2007 "Models, Conjectures and Exploration: An Analysis of Schelling's Checkerboard Model of Residential Segregation", *Journal of Economic Methodology*, 14, 4, pp. 429-454. (Cit. on pp. 91, 115-117, 134-135, 143.)

2008 *The Invisible Hand in Economics: How Economists Explain Unintented Social Consequences*, Routledge, London. (Cit. on pp. 146, 164.)

2018 "The Diversity of Models as a Means to Better Explanations in Economics", *Journal of Economic Methodology*, 25, 3, pp. 237-251. (Cit. on p. 172.)

Backhouse, Roger

1997 *Truth and Progress in Economic Knowledge*, Edward Elgar, Cheltenham. (Cit. on pp. 147, 161.)

1998 "If Mathematics Is Informal, Then Perhaps We Should Accept That Economics Must Be Informal Too", *The Economic Journal*, 108, 451, pp. 1848-1858. (Cit. on pp. 146-147.)

2009 "Friedman's 1953 Essay and the Marginalist Controversy", in *The Methodology of Positive Economics: Reflections on the Milton Friedman Legacy*, ed. by Uskali Mäki, Cambridge University Press, Cambridge, pp. 217-240. (Cit. on p. 5.)

2010 *The Puzzle of Modern Economics: Science or Ideology?*, Cambridge University Press. (Cit. on p. 3.)

Baker, Alan

2012 "Science-Driven Mathematical Explanation", *Mind*, 121, 482, pp. 243-267. (Cit. on pp. 85, 143.)

Baker, Victor R. and Russell C. Bunker

1985 "Cataclysmic Late Pleistocene Flooding from Glacial Lake Missoula: A Review", *Quaternary Science Reviews*, 4, 1, pp. 1-41. (Cit. on p. 101.)

Banerjee, Abhijit V.

1992 "A Simple Model of Herd Behavior", *The Quarterly Journal of Economics*, 107, 3, pp. 797-817. (Cit. on pp. 7, 24.)

Baron, Sam

2016 "Explaining Mathematical Explanation", *The Philosophical Quarterly*, 66, 264, pp. 458-480. (Cit. on p. 86.)

Baron, Sam, Mark Colyvan, and David Ripley

2017 "How Mathematics Can Make a Difference", *Philosophers' Imprint*, 17, 3, pp. 1-19. (Cit. on pp. 85-86, 120, 143, 173.)

Batterman, Robert W.

2002 *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*, Oxford University Press, New York. (Cit. on pp. 80, 85, 143.)

2009 "Idealization and Modeling", *Synthese*, 169, 3, pp. 427-446. (Cit. on p. 16.)

2010 "On the Explanatory Role of Mathematics in Empirical Science", *British Journal for the Philosophy of Science*, 61, 1, pp. 1-25. (Cit. on p. 86.)

Batterman, Robert W. and Collin Rice

2014 "Minimal Model Explanations", *Philosophy of Science*, 81, 3, pp. 349-376. (Cit. on pp. 16, 85, 143.)

Benjamin, Oliver

2016 *The Dude De Ching: Annotated Edition*, Abide University Press. (Cit. on p. ix.)

Bertinet, Richard

2005 *Dough: Simple Contemporary Bread*, Kyle Cathie, London. (Cit. on p. vii.)

Blaug, Mark

1992 *The Methodology of Economics: Or How Economists Explain*, 2nd ed., Cambridge University Press, Cambridge. (Cit. on pp. 5, 147.)

1996 *Economic Theory in Retrospect*, 5th ed., Cambridge University Press, Cambridge. (Cit. on pp. 147-148.)

2002 "Ugly Currents in Modern Economics", in *Fact and Fiction in Economics. Models, Realism, and Social Construction*, ed. by Uskali Mäki, Cambridge University Press, Cambridge, pp. 35-56. (Cit. on pp. 10, 24, 145.)

2003 "The Formalist Revolution of the 1950s", *Journal of the History of Economic Thought*, 25, 2, pp. 145-156. (Cit. on p. 144.)

2009 "The Debate over F53 after Fifty Years", in *The Methodology of Positive Economics: Reflections on the Milton Friedman Legacy*, ed. by Uskali Mäki, Cambridge University Press, Cambridge, pp. 349-354. (Cit. on p. 5.)

Bokulich, Alisa

2008 "Can Classical Structures Explain Quantum Phenomena?", *The British Journal for the Philosophy of Science*, 59, 2, pp. 217-235. (Cit. on p. 66.)

2011 "How Scientific Models Can Explain", *Synthese*, 180, 1, pp. 33-45. (Cit. on pp. 16, 66.)

2014 "How the Tiger Bush Got Its Stripes: 'How Possibly' vs. 'How Actually' Model Explanations", *The Monist*, 97, 3, pp. 321-338. (Cit. on pp. 13, 89, 107, 109, 113-115, 123, 137, 143.)

2016 "Fiction As a Vehicle for Truth: Moving Beyond the Ontic Conception", *The Monist*, 99, 3, pp. 260-279. (Cit. on pp. 66-67.)

Boland, Lawrence A.

1979 "A Critique of Friedman's Critics", *Journal of Economic Literature*, 17, 2, pp. 503-522. (Cit. on p. 5.)

1989 *The Methodology of Economic Model Building. Methodology after Samuelson*, Routledge, London. (Cit. on p. 5.)

Bolinska, Agnes

2013 "Epistemic Representation, Informativeness and the Aim of Faithful Representation", *Synthese*, 190, 2, pp. 219-234. (Cit. on p. 59.)

Brandon, Robert N.

1990 *Adaptation and Environment*, Princeton University Press, Princeton. (Cit. on pp. 107, 109, 111, 113, 115, 118, 123-125, 130-131.)

Bueno, Otávio and Mark Colyvan

2011 "An Inferential Conception of the Application of Mathematics", *Noûs*, 45, 2, pp. 345-374. (Cit. on pp. 86, 155, 157.)

Bueno, Otávio and Steven French

2012 "Can Mathematics Explain Physical Phenomena?", *The British Journal for the Philosophy of Science*, 63, 1, pp. 85-113. (Cit. on p. 155.)

Caldwell, Bruce J.

1994 *Beyond Positivism: Economic Methodology in the Twentieth Century*, Revised, Routledge, London. (Cit. on p. 5.)

Carman, Christián and José Díez

2015 "Did Ptolemy Make Novel Predictions? Launching Ptolemaic Astronomy into the Scientific Realism Debate", *Studies in History and Philosophy of Science Part A*, 52, pp. 20-34. (Cit. on p. 66.)

Carter, J. Adam and Bolesław Czarnecki

2016 "Extended Knowledge-How", *Erkenntnis*, 81, 2, pp. 259-273. (Cit. on p. 45.)

Carter, J. Adam, Jesper Kallestrup, Spyridon Orestis Palermos, and Duncan Pritchard

2014 "Varieties of Externalism", *Philosophical Issues*, 24, 1, pp. 63-109. (Cit. on pp. 45-46.)

Carter, J. Adam and Spyridon Orestis Palermos

2014 "Active Externalism and Epistemic Internalism", *Erkenntnis*, 80, 4, pp. 753-772. (Cit. on pp. 46-47.)

Carter, J. Adam and Duncan Pritchard

2015 "Knowledge-How and Cognitive Achievement", *Philosophy and Phenomenological Research*, 91, 1, pp. 181-199. (Cit. on p. 35.)

Cartwright, Nancy

1989 *Nature's Capacities and Their Measurement*, Oxford University Press, Oxford. (Cit. on p. 8.)

2009 "If No Capacities Then No Credible Worlds. But Can Models Reveal Capacities?", *Erkenntnis*, 70, 1, pp. 45-58. (Cit. on p. 8.)

Chakravartty, Anjan

2010 "Informational versus Functional Theories of Scientific Representation", *Synthese*, 172, 2, pp. 197-213. (Cit. on pp. 58, 71.)

Chick, Victoria

1998 "On Knowing One's Place: The Role of Formalism in Economics", *The Economic Journal*, 108, 451, pp. 1859-1869. (Cit. on p. 10.)

Clark, Andy

2008 *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*, Oxford University Press, New York. (Cit. on pp. 33, 45.)

Clark, Andy and David Chalmers

1998 "The Extended Mind", *Analysis*, 58, 1, pp. 7-19. (Cit. on pp. 33, 45-46.)

Clark, William A. V. and Mark Fossett

2008 "Understanding the Social Context of the Schelling Segregation Model", *Proceedings of the National Academy of Sciences*, 105, 11, pp. 4109-4114. (Cit. on pp. 91, 117.)

Claveau, François and Melissa Vergara Fernández

2015 "Epistemic Contributions of Models: Conditions for Propositional Learning", *Perspectives on Science*, 23, 4, pp. 405-423. (Cit. on pp. 132, 136, 162, 172.)

Coase, Ronald H.

1992 "The Institutional Structure of Production", *The American Economic Review*, 82, 4, pp. 713-719. (Cit. on p. 10.)

Cohen, Michael P.

2016 "On Three Measures of Explanatory Power with Axiomatic Representations", *The British Journal for the Philosophy of Science*, 67, 4, pp. 1077-1089. (Cit. on p. 170.)

Colander, David

2007 *The Making of an Economist, Redux*, Princeton University Press, Princeton. (Cit. on p. 3.)

Colander, David and Arjo Klamer

1987 "The Making of an Economist", *The Journal of Economic Perspectives*, 1, 2, pp. 95-111. (Cit. on p. 3.)

Contessa, Gabriele

2007 "Scientific Representation, Interpretation, and Surrogative Reasoning", *Philosophy of Science*, 75, 1, pp. 48-68. (Cit. on pp. 58-59, 71.)

Cooper, Gregory

1996 "Theoretical Modeling and Biological Laws", *Philosophy of Science*, 63, S28-S35. (Cit. on p. 115.)

Craver, Carl F.

2006 "When Mechanistic Models Explain", *Synthese*, 153, 3, pp. 355-376. (Cit. on pp. 13, 99, 107, 113-115, 118, 143.)

Currie, Adrian

2015 "Philosophy of Science and the Curse of the Case Study", in *The Palgrave Handbook of Philosophical Methods*, ed. by Chris Daly, Palgrave Macmillan, Houndsmills, pp. 553-572. (Cit. on p. 15.)

Da Costa, Newton C. A. and Steven French

2003 *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*, Oxford University Press, Oxford. (Cit. on p. 56.)

de Regt, Henk W.

2004 "Discussion Note: Making Sense of Understanding", *Philosophy of Science*, 71, 1, pp. 98-109. (Cit. on pp. 3, 10.)

2009 "The Epistemic Value of Understanding", *Philosophy of Science*, 76, 5, pp. 585-597. (Cit. on pp. 3, 35, 42, 73, 110.)

2017 *Understanding Scientific Understanding*, Oxford University Press, New York. (Cit. on pp. 29, 34.)

de Regt, Henk W. and Dennis Dieks

2005 "A Contextual Approach to Scientific Understanding", *Synthese*, 144, 1, pp. 137-170. (Cit. on pp. 3, 10, 29, 42.)

de Regt, Henk W., Sabina Leonelli, and Kai Eigner

2009 (eds.), *Scientific Understanding. Philosophical Perspectives*, University of Pittsburgh Press, Pittsburgh. (Cit. on p. 3.)

Debreu, Gerard

1959 *Theory of Value. An Axiomatic Analysis of Economic Equilibrium*, Yale Universisity Press, New Haven and London. (Cit. on p. 147.)

1984 "Economic Theory in the Mathematical Mode", *The Scandinavian Journal of Economics*, 86, 4, pp. 393-410. (Cit. on pp. 146, 149.)

1986 "Theoretic Models: Mathematical Form and Economic Content", *Econometrica*, 54, 6, pp. 1259-1270. (Cit. on pp. 149, 159, 165.)

2008 "Existence of General Equilibrium", in *The New Palgrave Dictionary of Economics*, ed. by Steven N. Durlauf and Lawrence E. Blume, 2nd ed., Nature Publishing Group, Basingstoke, pp. 112-116. (Cit. on pp. 146, 148, 156, 162.)

Dray, William H.

1954 "Explanatory Narrative in History", *The Philosophical Quarterly*, 4, 14, pp. 15-27. (Cit. on p. 108.)

1957 *Laws and Explanation in History*, Clarendon Press, Oxford. (Cit. on pp. 108, 110-111, 118, 122.)

1968 "On Explaining How-Possibly", *The Monist*, 52, 3, pp. 390-407. (Cit. on pp. 13, 107-109, 111, 117, 121, 129.)

Elgin, Catherine Z.

2007 "Understanding and the Facts", *Philosophical Studies*, 132, 1, pp. 33-42. (Cit. on pp. 27, 29, 38, 105.)

Elster, Jon

2007 *Explaining Social Behavior. More Nuts and Bolts for the Social Sciences*, Cambridge University Press, Cambridge, UK. (Cit. on p. 64.)

Euler, Leonhard

1741 "Solutio Problematis Ad Geometriam Situs Pertinentis", *Commentarii academiae scientiarum Petropolitanae*, 8, pp. 128-140. (Cit. on p. 86.)

Feiwel, George R.

1987 "The Potential and Limits of Economic Analysis: The Contributions of Kenneth J. Arrow", in *Arrow and the Ascent of Modern Economic Theory*, ed. by George R. Feiwel, Macmillan Press, Basingstoke, pp. 1-188. (Cit. on p. 156.)

Flowers, R. M. and K. A. Farley

2012 "Apatite 4He/3He and (U-Th)/He Evidence for an Ancient Grand Canyon", *Science*, 338, 6114, pp. 1616-1619. (Cit. on p. 100.)

Forber, Patrick

2010 "Confirmation and Explaining How Possible", *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 41, 1, pp. 32-40. (Cit. on pp. 13, 89, 93, 107-109, 112, 115, 118, 122, 125, 129-131, 143.)

2012 "Conjecture and Explanation: A Reply to Reydon", *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, 1, pp. 298-301. (Cit. on pp. 94, 112.)

Fourcade, Marion, Etienne Ollion, and Yann Algan

2015 "The Superiority of Economists", *The Journal of Economic Perspectives*, 29, 1, pp. 89-113. (Cit. on p. 1.)

French, Steven and James Ladyman

1999 "Reinflating the Semantic Approach", *International Studies in the Philosophy of Science*, 13, 2, pp. 103-121. (Cit. on pp. 56, 133.)

Friedman, Michael

1974 "Explanation and Scientific Understanding", *Journal of Philosophy*, 71, 1, pp. 5-19. (Cit. on pp. 3, 10, 79.)

Friedman, Milton

1953 "The Methodology of Positive Economics", in *Essays in Positive Economics*, University of Chicago Press, Chicago, pp. 3-43. (Cit. on p. 5.)

Frigg, Roman

2006 "Scientific Representation and the Semantic View of Theories", *Theoria*, 55, pp. 49-65. (Cit. on pp. 55, 71.)

Frigg, Roman and James Nguyen

2017 "Models and Representation", in *Springer Handbook of Model-Based Science*, ed. by Lorenzo Magnani and Tommaso Bertolotti, Springer Handbooks, Springer, Cham, pp. 49-102. (Cit. on pp. 57-58, 68-69.)

Fumagalli, Roberto

2016 "Why We Cannot Learn from Minimal Models", *Erkenntnis*, 81, 3, pp. 433-455. (Cit. on p. 94.)

Geanakoplos, John

2008 "Arrow–Debreu Model of General Equilibrium", in *The New Palgrave Dictionary of Economics*, ed. by Steven N. Durlauf and Lawrence E. Blume, 2nd ed., Nature Publishing Group, Basingstoke, pp. 222-232. (Cit. on pp. 144, 146, 150.)

Gendler, Tamar Szabó and John Hawthorne

2002 (eds.), *Conceivability and Possibility*, Oxford University Press, Oxford. (Cit. on p. 127.)

Giere, Ronald N.

1988 *Explaining Science. A Cognitive Approach*, University of Chicago Press, Chicago. (Cit. on p. 56.)

2002a "Models as Parts of Distributed Cognitive Systems", in *Model Based Reasoning: Science, Technology, Values*, ed. by Lorenzo Magnani and Nancy J. Nersessian, Kluwer/-Plenum, New York, pp. 227-241. (Cit. on p. 46.)

2002b "Scientific Cognition as Distributed Cognition", in *The Cognitive Basis of Science*, ed. by Peter Carruthers, Stephen P. Stich, and Michael Siegal, Cambridge University Press, p. 285. (Cit. on p. 46.)

2004 "How Models Are Used to Represent Reality", *Philosophy of Science*, 71, 5, pp. 742-752. (Cit. on p. 56.)

2007 "Distributed Cognition without Distributed Knowing", *Social Epistemology*, 21, 3, pp. 313-320. (Cit. on p. 46.)

Giere, Ronald N.

2010 "An Agent-Based Conception of Models and Scientific Representation", *Synthese*, 172, 2, pp. 269-281. (Cit. on p. 133.)

Gijsbers, Victor

2013 "Understanding, Explanation, and Unification", *Studies in History and Philosophy of Science Part A*, 44, 3, pp. 516-522. (Cit. on p. 73.)

Goldman, Alvin I.

1976 "Discrimination and Perceptual Knowledge", *Journal of Philosophy*, 73, 20, pp. 771-791. (Cit. on p. 39.)

Graham Kennedy, Ashley

2012 "A Non Representationalist View of Model Explanation", *Studies in History and Philosophy of Science Part A*, 43, 2, pp. 326-332. (Cit. on p. 16.)

Greco, John

2010 *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*, Cambridge University Press. (Cit. on pp. 35-36, 169.)

2012 "A (Different) Virtue Epistemology", *Philosophy and Phenomenological Research*, 85, 1, pp. 1-26. (Cit. on p. 36.)

2014 "*Episteme*: Knowledge and Understanding", in *Virtues & Their Vices*, ed. by Kevin Timpe and Craig A. Boyd, Oxford University Press, Oxford, pp. 285-302. (Cit. on pp. 29, 34, 40.)

Grimm, Stephen R.

2006 "Is Understanding A Species Of Knowledge?", *British Journal for the Philosophy of Science*, 57, 3, pp. 515-535. (Cit. on pp. 3, 29, 40.)

2008 "Explanatory Inquiry and the Need for Explanation", *British Journal for the Philosophy of Science*, 59, 3, pp. 481-497. (Cit. on p. 3.)

2009 "Reliability and the Sense of Understanding", in *Scientific Understanding. Philosophical Perspectives*, ed. by Henk W. de Regt, Sabina Leonelli, and Kai Eigner, University of Pittsburgh Press, Pittsburgh, pp. 83-99. (Cit. on pp. 12, 32.)

2010 "The Goal of Explanation", *Studies In History and Philosophy of Science Part A*, 41, 4, pp. 337-344. (Cit. on pp. 3, 34-35, 85, 104.)

2014 "Understanding as Knowledge of Causes", in *Virtue Epistemology Naturalized*, ed. by Abrol Fairweather, Springer, Dordrecht, pp. 329-345. (Cit. on pp. 35, 40, 43, 170.)

Grüne-Yanoff, Till

2009 "Learning from Minimal Economic Models", *Erkenntnis*, 70, 1, pp. 81-99. (Cit. on pp. 9, 91-92, 94, 117, 135, 138-139, 143, 162, 172.)

2011 "Isolation Is Not Characteristic of Models", *International Studies in the Philosophy of Science*, 25, 2, pp. 1-19. (Cit. on p. 8.)

2013a "Appraising Models Nonrepresentationally", *Philosophy of Science*, 80, 5, pp. 850-861. (Cit. on pp. 9, 13, 89, 91, 107, 115-116, 122-123, 125, 139, 143.)

2013b "Genuineness Resolved: A Reply to Reiss' Purported Paradox", *Journal of Economic Methodology*, 20, 3, pp. 255-261. (Cit. on pp. 12-14, 27, 52, 89, 115, 143.)

Guajardo, Jaime, Daniel Leigh, and Andrea Pescatori

2014 "Expansionary Austerity? International Evidence", *Journal of the European Economic Association*, 12, 4, pp. 949-968. (Cit. on p. 16.)

Guala, Francesco

2005 *The Methodology of Experimental Economics*, Cambridge University Press, Cambridge. (Cit. on p. 36.)

Hahn, Frank

1973 *On the Notion of Equilibrium in Economics: An Inaugural Lecture*, Cambridge University Press, Cambridge. (Cit. on p. 165.)

Hahn, Frank

1984 *Equilibrium and Macroeconomics*, Blackwell, Oxford. (Cit. on pp. 147, 153, 159.)

1985 *Money, Growth and Stability*, Basil Blackwell, Oxford. (Cit. on pp. 144, 153, 166.)

Hake, Richard R.

1998 "Interactive-Engagement versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses", *American Journal of Physics*, 66, 1, pp. 64-74. (Cit. on p. 41.)

Hall, R. L. and C. J. Hitch

1939 "Price Theory and Business Behaviour", *Oxford Economic Papers*, 2, pp. 12-45. (Cit. on p. 5.)

Hamermesh, Daniel S.

2013 "Six Decades of Top Economics Publishing: Who and How?", *Journal of Economic Literature*, 51, 1, pp. 162-172. (Cit. on p. 27.)

Hands, D. Wade

2001 *Reflection without Rules: Economic Methodology and Contemporary Science Theory*, Cambridge University Press, Cambridge. (Cit. on p. 145.)

2016 "Derivational Robustness, Credible Substitute Systems and Mathematical Economic Models: The Case of Stability Analysis in Walrasian General Equilibrium Theory", *European Journal for Philosophy of Science*, 6, 1, pp. 31-53. (Cit. on p. 165.)

Hausman, Daniel M.

1992a *Essays on Philosophy and Economic Methodology*, Cambridge University Press, Cambridge. (Cit. on p. 6.)

1992b *The Inexact and Separate Science of Economics*, Cambridge University Press, Cambridge. (Cit. on pp. 9, 12, 134, 143.)

1998 "Problems with Realism in Economics", *Economics and Philosophy*, 14, 02, pp. 185-213. (Cit. on p. 6.)

2009 "Laws, Causation, and Economic Methodology", in *The Oxford Handbook of Philosophy of Economics*, ed. by Harold Kincaid and Don Ross, Oxford University Press, New York, pp. 35-54. (Cit. on pp. 11, 15.)

2018 "Philosophy of Economics", in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Fall 2018, Metaphysics Research Lab, Stanford University. (Cit. on p. 6.)

Hempel, Carl G.

1965a "Aspects of Scientific Explanation", in *Aspects of Scientific Explanation: And Other Essays in the Philosophy of Science*, Free Press, New York, pp. 331-497. (Cit. on pp. 3, 22, 79, 107-110, 113, 118, 122.)

1965b "The Logic of Functional Analysis", in *Aspects of Scientific Explanation: And Other Essays in the Philosophy of Science*, Free Press, New York, pp. 297-330. (Cit. on p. v.)

Hempel, Carl G. and Paul Oppenheim

1948 "Studies in the Logic of Explanation", *Philosophy of Science*, 15, 2, pp. 135-175. (Cit. on pp. 3, 108, 110.)

Hestenes, David, Malcolm Wells, and Gregg Swackhamer

1992 "Force Concept Inventory", *The Physics Teacher*, 30, 3, pp. 141-158. (Cit. on p. 41.)

Hills, Alison

2016 "Understanding Why", *Noûs*, 50, 4, pp. 661-688. (Cit. on pp. 18, 29, 34-35, 38-39, 43, 79.)

Hindriks, Frank

2013 "Explanation, Understanding, and Unrealistic Models", *Studies in History and Philosophy of Science Part A*, 44, 3, pp. 523-531. (Cit. on pp. 108, 143.)

Hotelling, Harold

1929 "Stability in Competition", *Economic Journal*, 39, 153, pp. 41-57. (Cit. on p. 59.)

Hutchins, Edwin

1995 *Cognition in the Wild*, MIT Press. (Cit. on p. 45.)

Illari, Phyllis and Federica Russo

2014 *Causality: Philosophical Theory Meets Scientific Practice*, Oxford University Press, Oxford. (Cit. on p. 72.)

Ingrao, Bruna and Giorgio Israel

1990 *The Invisible Hand: Economic Equilibrium in the History of Science*, MIT Press, Cambridge, MA. (Cit. on p. 159.)

Jordà, Òscar and Alan M. Taylor

2016 "The Time for Austerity: Estimating the Average Treatment Effect of Fiscal Policy", *The Economic Journal*, 126, 590, pp. 219-255. (Cit. on p. 16.)

Kakutani, Shizuo

1941 "A Generalization of Brouwer's Fixed Point Theorem", *Duke Mathematical Journal*, 8, 3, pp. 457-459. (Cit. on pp. 144, 162.)

Kelp, Christoph

2013 "Extended Cognition and Robust Virtue Epistemology", *Erkenntnis*, 78, 2, pp. 245-252. (Cit. on p. 45.)

2015 "Understanding Phenomena", *Synthese*, 192, 12, pp. 3799-3816. (Cit. on pp. 29, 40.)

Khalifa, Kareem

2011 "Understanding, Knowledge, and Scientific Antirealism", *Grazer Philosophische Studien*, 83, 1, pp. 93-112. (Cit. on pp. 3, 90, 127.)

2012 "Inaugurating Understanding or Repackaging Explanation?", *Philosophy of Science*, 79, 1, pp. 15-37. (Cit. on pp. 10, 29, 73, 78, 81.)

2013a "Is Understanding Explanatory or Objectual?", *Synthese*, 190, 6, pp. 1153-1171. (Cit. on p. 90.)

2013b "The Role of Explanation in Understanding", *The British Journal for the Philosophy of Science*, 64, 1, pp. 161-187. (Cit. on pp. 28, 73, 75, 84, 89.)

2013c "Understanding, Grasping and Luck", *Episteme*, 10, 01, pp. 1-17. (Cit. on pp. 40, 79.)

2017 *Understanding, Explanation, and Scientific Knowledge*, Cambridge University Press, Cambridge. (Cit. on pp. 29-30, 34-35, 38, 50.)

Khalifa, Kareem and Michael Gadomski

2013 "Understanding as Explanatory Knowledge: The Case of Bjorken Scaling", *Studies in History and Philosophy of Science Part A*, 44, 3, pp. 384-392. (Cit. on p. 79.)

Kim, Jaegwon

1994 "Explanatory Knowledge and Metaphysical Dependence", *Philosophical Issues*, 5, pp. 51-69. (Cit. on p. 73.)

Kipling, Rudyard

[1902] 1912 *Just So Stories*, Doubleday Page & Company, New York; trans. 1902. (Cit. on p. 125.)

Kirchhoff, Michael David and Will Newsome

2012 "Distributed Cognitive Agency in Virtue Epistemology", *Philosophical Explorations*, 15, 2, pp. 165-180. (Cit. on p. 45.)

Kitcher, Philip

1981 "Explanatory Unification", *Philosophy of Science*, 48, 4, pp. 507-531. (Cit. on pp. 3, 8, 79-80.)

1989 "Explanatory Unification and the Causal Structure of the World", in *Scientific Explanation*, ed. by Philip Kitcher and Wesley C. Salmon, University of Minnesota Press, Minneapolis, pp. 410-505. (Cit. on p. 31.)

Kment, Boris

2012 "Varieties of Modality", in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Winter 2012. (Cit. on p. 99.)

2014 *Modality and Explanatory Reasoning*, Oxford University Press, Oxford. (Cit. on p. 171.)

2017 "Varieties of Modality", in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Spring 2017, Metaphysics Research Lab, Stanford University. (Cit. on pp. 73, 120.)

Knuuttila, Tarja

2011 "Modelling and Representing: An Artefactual Approach to Model-Based Representation", *Studies in History and Philosophy of Science Part A*, 42, 2, pp. 262-271. (Cit. on pp. 69, 71.)

Koopmans, Tjalling C.

1957 *Three Essays on the State of Economic Science*, McGraw-Hill, New York. (Cit. on pp. 159-160.)

Koopmans, Tjalling C. and Augustus F. Bausch

1959 "Selected Topics in Economics Involving Mathematical Reasoning", *SIAM Review*, 1, 2, pp. 79-148. (Cit. on p. 160.)

Koslicki, Kathrin

2012 "Varieties of Ontological Dependence", in *Metaphysical Grounding: Understanding the Structure of Reality*, ed. by Fabrice Correia and Benjamin Schnieder, Cambridge University Press, p. 186. (Cit. on p. 73.)

Kroedel, Thomas

2012 "Counterfactuals and the Epistemology of Modality", *Philosophers' Imprint*, 12, 12, pp. 1-14. (Cit. on p. 156.)

Kropotkin, Peter

[1897] 2002 "Anarchist Morality", in *Anarchism. A Collection of Revolutionary Writings*, ed. by Roger N. Baldwin, Dover Publications, Mineola, NY, pp. 79-113. (Cit. on p. 1.)

Krugman, Paul

1991 "Increasing Returns and Economic Geography", *Journal of Political Economy*, 99, 3, pp. 483-499. (Cit. on p. 7.)

2009 "How Did Economists Get It So Wrong?", *The New York Times*. (Cit. on p. 3.)

Kuhn, Thomas S.

[1962] 1996 *The Structure of Scientific Revolutions*, Third, University of Chicago Press, Chicago. (Cit. on p. 31.)

Kuorikoski, Jaakko

2007 "Explaining with Equilibria", in *Rethinking Explanation*, ed. by Johannes Persson and Petri Ylikoski, Springer, Dordrecht, pp. 149-162. (Cit. on p. 154.)

2011 "Simulation and the Sense of Understanding", in *Models, Simulations, and Representations*, ed. by Paul Humphreys and Cyrille Imbert, Routledge, New York, pp. 168-187. (Cit. on pp. 18, 29, 32.)

Kuorikoski, Jaakko and Aki Lehtinen

2009 "Incredible Worlds, Credible Results", *Erkenntnis*, 70, 1, pp. 119-131. (Cit. on pp. 17, 29, 45-46, 56, 60-61.)

Kuorikoski, Jaakko, Aki Lehtinen, and Caterina Marchionni

2010 "Economic Modelling as Robustness Analysis", *British Journal for the Philosophy of Science*, 61, 3, pp. 541-567. (Cit. on p. 165.)

Kuorikoski, Jaakko and Petri Ylikoski

2015 "External Representations and Scientific Understanding", *Synthese*, 192, 12, pp. 3817-3837. (Cit. on pp. 18, 29-32, 41, 45, 48, 56, 60-65, 67-70, 72, 74, 92, 99, 143, 165.)

Kvanvig, Jonathan L.

2003 *The Value of Knowledge and the Pursuit of Understanding*, Cambridge University Press, Cambridge. (Cit. on pp. 29, 38, 90, 105.)

Kydland, Finn E. and Edward C. Prescott

1982 "Time to Build and Aggregate Fluctuations", *Econometrica*, 50, 6, pp. 1345-1370. (Cit. on p. 7.)

Ladyman, James

2000 "What's Really Wrong with Constructive Empiricism? Van Fraassen and the Metaphysics of Modality", *The British Journal for the Philosophy of Science*, 51, 4, pp. 837-856. (Cit. on p. 64.)

Lange, Marc

2005 "A Counterfactual Analysis of the Concepts of Logical Truth and Necessity", *Philosophical Studies*, 125, 3, pp. 277-303. (Cit. on p. 156.)

2009 *Laws and Lawmakers: Science, Metaphysics, and the Laws of Nature*, Oxford University Press. (Cit. on pp. 99, 165, 171.)

2013 "What Makes a Scientific Explanation Distinctively Mathematical?", *The British Journal for the Philosophy of Science*, 64, 3, pp. 485-511. (Cit. on pp. 85-87, 120, 143, 154-155.)

Lange, Marc

2017 *Because Without Cause: Non-Causal Explanations in Science and Mathematics*, Oxford University Press, New York. (Cit. on pp. 10, 72, 80, 85, 120, 143, 154, 173.)

Laudan, Larry

1981 "A Confutation of Convergent Realism", *Philosophy of Science*, 48, 1, pp. 19-49. (Cit. on p. 66.)

Lawler, Insa

2018 "Understanding Why, Knowing Why, and Cognitive Achievements", *Synthese*, pp. 1-21. (Cit. on p. 29.)

Lawson, Tony

2009 "The Current Economic Crisis: Its Nature and the Course of Academic Economics", *Cambridge Journal of Economics*, 33, 4, pp. 759-777. (Cit. on p. 10.)

Leuridan, Bert, Erik Weber, and Maarten Van Dyck

2008 "The Practical Value of Spurious Correlations: Selective versus Manipulative Policy", *Analysis*, 68, 4, pp. 298-303. (Cit. on p. 64.)

Lewis, David

1986 "Causal Explanation", in *Philosophical Papers*, ed. by David Lewis, Oxford University Press, New York, vol. II, pp. 214-240. (Cit. on p. 80.)

Lipton, Peter

2004 *Inference to the Best Explanation*, 2nd, Routledge, London. (Cit. on pp. 80, 170.)

2009 "Understanding Without Explanation", in *Scientific Understanding. Philosophical Perspectives*, ed. by Henk W. de Regt, Sabina Leonelli, and Kai Eigner, University of Pittsburgh Press, Pittsburgh, pp. 43-63. (Cit. on pp. 73, 89, 97, 139.)

Machamer, Peter, Lindley Darden, and Carl F. Craver

2000 "Thinking about Mechanisms", *Philosophy of Science*, 67, 1, pp. 1-25. (Cit. on p. 108.)

Machery, Edouard

2017 *Philosophy Within Its Proper Bounds*, Oxford University Press, Oxford. (Cit. on pp. 170-171.)

Machlup, Fritz

1955 "The Problem of Verification in Economics", *Southern Economic Journal*, 22, 1, pp. 1-21. (Cit. on p. 5.)

Mäki, Uskali

1992 "On the Method of Isolation in Economics", *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 26, pp. 19-54. (Cit. on p. 8.)

2009a "MISSing the World. Models as Isolations and Credible Surrogate Systems", *Erkenntnis*, 70, 1, pp. 29-43. (Cit. on pp. 8, 71, 133, 135, 143.)

2009b "Unrealistic Assumptions and Unnecessary Confusions: Rereading and Rewriting F53 as a Realist Statement", in *The Methodology of Positive Economics: Reflections on the Milton Friedman Legacy*, ed. by Uskali Mäki, Cambridge University Press, Cambridge, pp. 90-116. (Cit. on p. 5.)

2011 "Models and the Locus of Their Truth", *Synthese*, 180, 1, pp. 47-63. (Cit. on p. 71.)

Mandeville, Bernard

[1724] 1988 *The Fable of the Bees or Private Vices, Publick Benefits*, ed. by Frederick Benjamin Kaye, Liberty Fund, Indianapolis, vol. 2 vols. (Cit. on p. 146.)

Mankiw, N. Gregory

2012 *Principles of Macroeconomics*, 6th, South-Western, Cengage Learning, Mason, OH. (Cit. on p. 146.)

2015 *Principles of Economics*, Seventh, Cengage Learning, Stamford, CT. (Cit. on p. 6.)

Mas-Collell, Andreu, Michael Whinston, and Jerry Green

1995 *Microeconomic Theory*, Oxford University Press, Oxford. (Cit. on p. 147.)

Mayer, Thomas

2009 "The Influence of Friedman's Methodological Essay", in *The Methodology of Positive Economics: Reflections on the Milton Friedman Legacy*, ed. by Uskali Mäki, Cambridge University Press, Cambridge, pp. 119-142. (Cit. on p. 5.)

Maynard Smith, John

1978 "Optimization Theory in Evolution", *Annual Review of Ecology and Systematics*, 9, pp. 31-56. (Cit. on p. 93.)

1982 *Evolution and the Theory of Games*, Cambridge University Press, Cambridge. (Cit. on p. 115.)

Maynard Smith, John and Geoffrey A. Parker

1976 "The Logic of Asymmetric Contests", *Animal Behaviour*, 24, 1, pp. 159-175. (Cit. on p. 92.)

Maynard Smith, John and George R. Price

1973 "The Logic of Animal Conflict", *Nature*, 246, 5427, pp. 15-18. (Cit. on pp. 92-93, 115.)

McCoy, C. D. and Michela Massimi

2018 "Simplified Models: A Different Perspective on Models as Mediators", *European Journal for Philosophy of Science*, 8, 1, pp. 99-123. (Cit. on p. 16.)

McKenzie, Lionel

1954 "On Equilibrium in Graham's Model of World Trade and Other Competitive Systems", *Econometrica*, 22, 2, pp. 147-161. (Cit. on p. 144.)

Menger, Carl

1985 "Economics as a Theoretical Science and Its Relationship to the Historical and Practical Economic Sciences", in *Investigations into the Method of the Social Sciences with Special Reference to Economics (Problems of Economics and Sociology)*, ed. by Louis Schneider, New York University Press, New York and London, pp. 35-94. (Cit. on p. 6.)

Mill, John Stuart

1844 *Essays on Some Unsettled Questions of Political Economy*, John W. Parker, West Strand, London. (Cit. on p. 4.)

Mongin, Philippe

1997 "The Marginalist Controversy", in *Handbook of Economic Methodology*, ed. by John B. Davis, D. Wade Wades, and Uskali Mäki, Edward Elgar, Cheltenham, UK ; Northampton, MA, pp. 277-281. (Cit. on p. 5.)

Monton, Bradley and Bas C. van Fraassen

2003 "Constructive Empiricism and Modal Nominalism", *The British Journal for the Philosophy of Science*, 54, 3, pp. 405-422. (Cit. on p. 64.)

Morris, Kevin

2012 "A Defense of Lucky Understanding", *The British Journal for the Philosophy of Science*, 63, 2, pp. 357-371. (Cit. on pp. 35, 79.)

Morrison, Margaret

2015 *Reconstructing Reality: Models, Mathematics, and Simulations*, OUP Usa. (Cit. on p. 16.)

Muldoon, Ryan, Tony Smith, and Michael Weisberg

2012 "Segregation That No One Seeks", *Philosophy of Science*, 79, 1, pp. 38-62. (Cit. on pp. 91, 164.)

Nash, John F.

1950 "Equilibrium Points in N-Person Games", *Proceedings of the National Academy of Sciences*, 36, 1, pp. 48-49. (Cit. on p. 162.)

Newman, Mark

2012 "An Inferential Model of Scientific Understanding", *International Studies in the Philosophy of Science*, 26, 1, pp. 1-26. (Cit. on p. 47.)

2013 "Refining the Inferential Model of Scientific Understanding", *International Studies in the Philosophy of Science*, 27, 2, pp. 173-197. (Cit. on p. 47.)

2017 "Theoretical Understanding in Science", *British Journal for the Philosophy of Science*, 68, 2, pp. 571-595. (Cit. on p. 42.)

Nobelprize.org

2014 "The Prize in Economics 1983 - Presentation Speech". (Cit. on p. 145.)

Noë, Alva

2005 "Against Intellectualism", *Analysis*, 65, 288, pp. 278-290. (Cit. on p. 35.)

Northcott, Robert and Anna Alexandrova

2014 "Armchair Science", *British Society for the Philosophy of Science Annual Conference 2014 (Cambridge; 10-11 July 2014)*. (Cit. on p. 26.)

Nounou, Antigone and F.A. Muller

2015 "Scientific Understanding with and without Scientific Explanation", unpublished typoscript. (Cit. on p. 10.)

Ostry, Jonathan D., Prakash Loungani, and Davide Furceri

2016 "Neoliberalism: Oversold?", *Finance & Development*, 53, 2. (Cit. on p. 16.)

Palermos, Spyridon Orestis

2014 "Knowledge and Cognitive Integration", *Synthese*, 191, 8, pp. 1931-1951. (Cit. on p. 45.)

2015 "Active Externalism, Virtue Reliabilism and Scientific Knowledge", *Synthese*, pp. 1-32. (Cit. on pp. 37, 45-46.)

Parker, Geoffrey A.

1974 "Assessment Strategy and the Evolution of Fighting Behaviour", *Journal of Theoretical Biology*, 47, 1, pp. 223-243. (Cit. on p. 93.)

Parker, Wendy S.

2015 "Getting (Even More) Serious about Similarity", *Biology & Philosophy*, 30, 2, pp. 267-276. (Cit. on p. 71.)

Pincock, Christopher

2007 "A Role for Mathematics in the Physical Sciences", *Noûs*, 41, 2, pp. 253-275. (Cit. on pp. 86, 96, 154-155, 157.)

2012 *Mathematics and Scientific Representation*, Oxford University Press, New York. (Cit. on p. 157.)

2015 "Abstract Explanations in Science", *The British Journal for the Philosophy of Science*, 66, 4, pp. 857-882. (Cit. on pp. 73, 85-87, 120, 143, 155.)

2018 "Accommodating Explanatory Pluralism", in *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, ed. by Alexander Reutlinger and Juha Saatsi, Oxford University Press, Oxford, pp. 39-56. (Cit. on pp. 62, 73.)

Pitt, Jack

1959 "Generalizations in Historical Explanation", *The Journal of Philosophy*, 56, 13, pp. 578-586. (Cit. on p. 122.)

Potochnik, Angela

2015 "Causal Patterns and Adequate Explanations", *Philosophical Studies*, 172, 5, pp. 1163-1182. (Cit. on p. 80.)

2017 *Idealization and the Aims of Science*, University of Chicago Press. (Cit. on p. 16.)

Poznic, Michael

2016 "Representation and Similarity: Suárez on Necessary and Sufficient Conditions of Scientific Representation", *Journal for General Philosophy of Science*, 47, 2, pp. 331-347. (Cit. on p. 71.)

Pritchard, Duncan

2008 "Knowing the Answer, Understanding and Epistemic Value", *Grazer Philosophische Studien*, 77, 1, pp. 325-339. (Cit. on p. 39.)

2009a "Apt Performance and Epistemic Value", *Philosophical Studies*, 143, 3, pp. 407-416. (Cit. on p. 29.)

2009b "Knowledge, Understanding and Epistemic Value", *Royal Institute of Philosophy Supplements*, 64, pp. 19-43. (Cit. on p. 35.)

2010 "Cognitive Ability and the Extended Cognition Thesis", *Synthese*, 175, 1, pp. 133-151. (Cit. on pp. 45-46.)

2014 "Knowledge and Understanding", in *Virtue Epistemology Naturalized*, ed. by Abrol Fairweather, Springer, Dordrecht, pp. 315-327. (Cit. on pp. 20, 38, 80.)

Putnam, Hilary

1975 *Mathematics, Matter and Method*, Cambridge University Press, Cambridge. (Cit. on p. 51.)

Reiner, Richard

1993 "Necessary Conditions and Explaining How-Possibly", *The Philosophical Quarterly*, 43, 170, pp. 58-69. (Cit. on p. 122.)

Reiss, Julian

2008 *Error in Economics: Towards a More Evidence-Based Methodology*, Routledge, Milton Park. (Cit. on p. 92.)

2012a "Idealization and the Aims of Economics: Three Cheers for Instrumentalism", *Economics and Philosophy*, 28, 3, pp. 363-383. (Cit. on pp. 5, 65.)

2012b "The Explanation Paradox", *Journal of Economic Methodology*, 19, 1, pp. 43-62. (Cit. on pp. 2, 7, 9, 12, 59, 117, 143.)

2013a "Models, Representation, and Economic Practice", in *Models, Simulations and the Reduction of Complexity*, ed. by Jörg-Henning Wolf and Ulrich Gähde, DeGruyter, Hamburg, pp. 107-116. (Cit. on pp. 13, 75.)

2013b *Philosophy of Economics: A Contemporary Introduction*, Routledge, New York, NY. (Cit. on p. 147.)

2013c "The Explanation Paradox Redux", *Journal of Economic Methodology*, 20, 3, pp. 280-292. (Cit. on pp. 3, 12-13.)

Resnik, David B.

1991 "How-Possibly Explanations in Biology", *Acta Biotheoretica*, 39, 2, pp. 141-149. (Cit. on pp. 95, 108, 114-115, 123, 130.)

Reutlinger, Alexander

2013 *A Theory of Causation in the Social and Biological Sciences*, Palgrave Macmillan, Basingstoke. (Cit. on p. 72.)

2016 "Is There A Monist Theory of Causal and Noncausal Explanations? The Counterfactual Theory of Scientific Explanation", *Philosophy of Science*, 83, 5, pp. 733-745. (Cit. on pp. 21, 85-87, 94, 96, 104-105, 127, 154, 173.)

2017a "Does the Counterfactual Theory of Explanation Apply to Non-Causal Explanations in Metaphysics?", *European Journal for Philosophy of Science*, 7, 2, pp. 239-256. (Cit. on p. 98.)

2017b "Explanation beyond Causation? New Directions in the Philosophy of Scientific Explanation", *Philosophy Compass*, pp. 1-11. (Cit. on pp. 8, 72, 85, 154.)

Reutlinger, Alexander, Dominik Hangleiter, and Stephan Hartmann

2018 "Understanding (with) Toy Models", *The British Journal for the Philosophy of Science*, 69, 4, pp. 1069-1099. (Cit. on p. 98.)

Reutlinger, Alexander and Juha Saatsi

2018a (eds.), *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, Oxford University Press, Oxford. (Cit. on pp. 10, 85, 143, 173.)

2018b "Introduction", in *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, ed. by Alexander Reutlinger and Juha Saatsi, Oxford University Press, Oxford, pp. 1-11. (Cit. on p. 80.)

Reydon, Thomas A.C.

2012 "How-Possibly Explanations as Genuine Explanations and Helpful Heuristics: A Comment on Forber", *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, 1, pp. 302-310. (Cit. on pp. 13, 93, 95, 107-109, 112, 114-115, 130, 143.)

Rice, Collin

2015 "Moving Beyond Causes: Optimality Models and Scientific Explanation", *Noûs*, 49, 3, pp. 589-615. (Cit. on p. 16.)

2016 "Factive Scientific Understanding without Accurate Representation", *Biology & Philosophy*, 31, 1, pp. 81-102. (Cit. on pp. 13, 73, 92, 107, 115, 118, 139, 143.)

Rice, Collin and Joshua Smart

2011 "Interdisciplinary Modeling: A Case Study of Evolutionary Economics", *Biology & Philosophy*, 26, 5, pp. 655-675. (Cit. on p. 143.)

Rodrik, Dani

2015 *Economics Rules: The Rights and Wrongs of the Dismal Science*, W. W. Norton & Company, New York. (Cit. on pp. 3, 25, 164.)

Rohwer, Yasha

2014 "Lucky Understanding without Knowledge", *Synthese*, 191, 5, pp. 945-959. (Cit. on p. 79.)

Rohwer, Yasha and Collin Rice

2013 "Hypothetical Pattern Idealization and Explanatory Models", *Philosophy of Science*, 80, 3, pp. 334-355. (Cit. on pp. 89, 92-93, 115, 139, 143.)

Rosenberg, Alexander

1992 *Economics : Mathematical Politics or Science of Diminishing Returns?*, University of Chicago Press, Chicago. (Cit. on pp. 10, 157.)

Ruben, David-Hillel

1990 *Explaining Explanation*, Routledge, London. (Cit. on p. 80.)

Rueger, Alexander

2005 "Perspectival Models and Theory Unification", *The British Journal for the Philosophy of Science*, 56, 3, pp. 579-594. (Cit. on p. 60.)

Saatsi, Juha

2011 "The Enhanced Indispensability Argument: Representational versus Explanatory Role of Mathematics in Science", *The British Journal for the Philosophy of Science*, 62, 1, pp. 143-154. (Cit. on p. 86.)

Salmon, Wesley C.

1984 *Scientific Explanation and the Causal Structure of the World*, Princeton University Press, Princeton. (Cit. on pp. 3, 13, 79.)

1989 *Four Decades of Scientific Explanation*, University of Pittsburgh Press, Pittsburgh. (Cit. on pp. 113, 121, 125.)

1998 *Causality and Explanation*, Oxford University Press, Oxford. (Cit. on p. 80.)

Schelling, Thomas C.

1971 "Dynamic Models of Segregation", *Journal of Mathematical Sociology*, 1, pp. 143-186. (Cit. on pp. 91, 116, 134.)

1978 *Micromotives and Macrobehavior*, W. W. Norton & Company, New York. (Cit. on pp. 24, 91, 116, 125, 134.)

Schupbach, Jonah N. and Jan Sprenger

2011 "The Logic of Explanatory Power", *Philosophy of Science*, 78, 1, pp. 105-127. (Cit. on p. 170.)

Sliwa, Paulina

2015 "IV—Understanding and Knowing", *Proceedings of the Aristotelian Society (Hardback)*, 115, 1pt1, pp. 57-74. (Cit. on pp. 29, 34, 38, 40.)

Smith, Adam

[1776] 1904 *An Inquiry into the Nature and Causes of the Wealth of Nations*, ed. by Edwin Cannan, 5th, Methuen & Co., London; trans. 1776. (Cit. on pp. 24, 146.)

Sober, Elliott

1983 "Equilibrium Explanation", *Philosophical Studies*, 43, 2, pp. 201-210. (Cit. on p. 154.)

2003 "Two Uses of Unification", in *Vienna Circle Institute Yearbook*, ed. by Friedrich Stadler, Kluwer Academic Publishers, New York, vol. 10, pp. 205-216. (Cit. on p. 109.)

Solow, Robert M.

1997 "How Did Economics Get That Way and What Did It Get?", *Daedalus*, 126, 1, pp. 39-58. (Cit. on p. 2.)

Sosa, Ernest

2007 *A Virtue Epistemology: Apt Belief and Reflective Knowledge*, Clarendon Press, Oxford, vol. I. (Cit. on pp. 35, 169.)

Stalnaker, Robert C.

1968 "A Theory of Conditionals", in *IFS*, ed. by William L. Harper, Robert Stalnaker, and Glenn Pearce, The University of Western Ontario Series in Philosophy of Science, 15, Springer Netherlands, pp. 41-55. (Cit. on p. 156.)

Stanley, Jason and Timothy Williamson

2001 "Knowing How", *The Journal of Philosophy*, 98, 8, pp. 411-444. (Cit. on p. 35.)

Stewart, Heather

2009 "This Is How We Let the Credit Crunch Happen, Ma'am ..." *the Guardian*. (Cit. on p. 16.)

Stiglitz, Joseph E.

1991 *The Invisible Hand and Modern Welfare Economics*, Working Paper 3641, National Bureau of Economic Research. (Cit. on pp. 146, 166.)

Strevens, Michael

2008 *Depth: An Account of Scientific Explanation*, Harvard University Press, Cambridge, MA. (Cit. on pp. 3, 59, 72, 78, 85, 154.)

2013 "No Understanding without Explanation", *Studies in History and Philosophy of Science Part A*, 44, 3, pp. 510-515. (Cit. on pp. 22, 29, 34, 59, 73, 78-79, 89-90, 94, 100, 109, 119, 132, 170.)

2018 "The Mathematical Route to Causal Understanding", in *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*, ed. by Alexander Reutlinger and Juha Saatsi, Oxford University Press, Oxford, pp. 96-116. (Cit. on pp. 87, 154.)

Suárez, Mauricio

2003 "Scientific Representation: Against Similarity and Isomorphism", *International Studies in the Philosophy of Science*, 17, 3, pp. 225-244. (Cit. on pp. 57, 71.)

2004 "An Inferential Conception of Scientific Representation", *Philosophy of Science*, 71, 5, pp. 767-779. (Cit. on pp. 19, 55-58, 68, 75.)

2008 "Experimental Realism Reconsidered. How Inference to the Most Likely Cause Might Be Sound", in *Nancy Cartwright's Philosophy of Science*, ed. by Stephan Hartmann, Carl Hoefer, and Luc Bovens, Routledge, New York, pp. 137-163. (Cit. on p. 70.)

2009 "Scientific Fictions as Rules of Inference", in *Fictions in Science: Philosophical Essays on Modeling and Idealization*, ed. by Mauricio Suárez, Routledge, pp. 158-178. (Cit. on pp. 60, 133.)

2010 "Scientific Representation", *Philosophy Compass*, 5, 1, pp. 91-101. (Cit. on p. 58.)

2015a "Deflationary Representation, Inference, and Practice", *Studies in History and Philosophy of Science Part A*, 49, pp. 36-47. (Cit. on pp. 56-58.)

2015b "Scientific Representation, Denotation, and Fictional Entities", in *Recent Developments in the Philosophy of Science: EPSA13 Helsinki*, ed. by Uskali Mäki, Ioannis Votsis, Stéphanie Ruphy, and Gerhard Schurz, European Studies in Philosophy of Science, 1, Springer International Publishing, pp. 331-341. (Cit. on p. 60.)

Suárez, Mauricio and Albert Solé

2006 "On the Analogy between Cognitive Representation and Truth", *Theoria*, 55, pp. 39-48. (Cit. on pp. 56-57.)

Sugden, Robert

2000 "Credible Worlds: The Status of Theoretical Models in Economics", *Journal of Economic Methodology*, 7, 1, pp. 1-31. (Cit. on pp. 91, 116, 133-134.)

2009 "Credible Worlds, Capacities and Mechanisms", *Erkenntnis*, 70, 1, pp. 3-27. (Cit. on pp. 8, 95, 133-135, 143.)

2011 "Explanations in Search of Observations", *Biology and Philosophy*, 26, 5, pp. 717-736. (Cit. on pp. 92, 130, 133-134.)

2013 "How Fictional Accounts Can Explain", *Journal of Economic Methodology*, 20, 3, pp. 237-243. (Cit. on pp. 8, 52, 71, 133-134.)

Sullivan, Emily

2018 "Understanding: Not Know-How", *Philosophical Studies*, 175, 1, pp. 221-240. (Cit. on pp. 29, 34-36, 43.)

Suppe, Frederick

1977 *The Structure of Scientific Theories*, University of Illinois Press, Urbana. (Cit. on p. 56.)

Suppes, Patrick

2002 *Representation and Invariance of Scientific Structures*, CSLI Publications, Stanford. (Cit. on p. 56.)

Swoyer, Chris

1991 "Structural Representation and Surrogative Reasoning", *Synthese*, 87, 3, pp. 449-508. (Cit. on p. 56.)

Teller, Paul

2001 "Twilight of the Perfect Model Model", *Erkenntnis*, 5, 3, pp. 393-415. (Cit. on pp. 56, 71.)

Toon, Adam

2015 "Where Is the Understanding?", *Synthese*, 192, 12, pp. 3859-3875. (Cit. on pp. 43-44.)

Trout, J. D.

2002 "Scientific Explanation And The Sense Of Understanding", *Philosophy of Science*, 69, 2, pp. 212-233. (Cit. on pp. 10, 12, 32, 73, 83.)

2007 "The Psychology of Scientific Explanation", *Philosophy Compass*, 2, 3, pp. 564-591. (Cit. on pp. 32, 78-79, 90.)

2018 "Understanding and Fluency", in *Making Sense of the World. New Essays on the Philosophy of Understanding*, ed. by Stephen R. Grimm, Oxford University Press, New York, pp. 232-253. (Cit. on pp. 32, 170.)

Ullmann-Margalit, Edna

1978 "Invisible-Hand Explanations", *Synthese*, 39, 2, pp. 263-291. (Cit. on p. 146.)

Vaesen, Krist

2011 "Knowledge without Credit, Exhibit 4: Extended Cognition", *Synthese*, 181, 3, pp. 515-529. (Cit. on p. 45.)

van Fraassen, Bas C.

1980 *The Scientific Image*, Oxford University Press, Oxford. (Cit. on pp. 8, 56, 80.)

1989 *Laws And Symmetry*, Oxford University Press, Oxford. (Cit. on p. 120.)

van Riel, Raphael

2015 "The Content of Model-Based Information", *Synthese*, 192, 12, pp. 3839-3858. (Cit. on pp. 107, 119-120, 135-136, 139.)

2016 "If You Understand, You Won't Be Lucky", *Grazer Philosophische Studien*, 93, 2, pp. 196-211. (Cit. on p. 79.)

VanLehn, Kurt and Brett van de Sande

2009 "Acquiring Conceptual Expertise from Modeling: The Case of Elementary Physics", in *Development of Professional Expertise: Toward Measurement of Expert Performance and Design of Optimal Learning Environments*, ed. by K. Anders Ericsson, Cambridge University Press, Cambridge, pp. 356-378. (Cit. on pp. 41-42.)

Varian, Hal R.

1984 "Gerard Debreu's Contributions to Economics", *The Scandinavian Journal of Economics*, 86, 1, pp. 4-14. (Cit. on p. 144.)

Vergara Fernández, Melissa

2018 *The Use of Models in Economics*, PhD thesis, Erasmus University Rotterdam, Rotterdam. (Cit. on p. 172.)

Verreault-Julien, Philippe

2010 "La Théorie de l'évolution Culturelle d'Hayek : Reconstruction Rationnelle", *Phares*, 10, pp. 147-167. (Cit. on p. 2.)

Vickers, Peter

2013 "A Confrontation of Convergent Realism", *Philosophy of Science*, 80, 2, pp. 189-211. (Cit. on p. 66.)

Von Neumann, John

1945 "A Model of General Economic Equilibrium", *The Review of Economic Studies*, 13, 1, pp. 1-9. (Cit. on p. 162.)

Waitt, Richard B.

1980 "About Forty Last-Glacial Lake Missoula Jökulhlaups through Southern Washington", *The Journal of Geology*, 88, 6, pp. 653-679. (Cit. on p. 101.)

Wald, Abraham

[1936] 1951 "On Some Systems of Equations of Mathematical Economics", *Econometrica*, 19, 4, pp. 368-403. (Cit. on p. 162.)

Walras, Léon

1954 *Elements of Pure Economics: Or the Theory of Social Wealth*, Translated by Willam Jaffé, Allen & Unwin, London; trans. 1926. (Cit. on pp. 24, 148, 167.)

Wayne, Andrew

2011 "Expanding the Scope of Explanatory Idealization", *Philosophy of Science*, 78, 5, pp. 830-841. (Cit. on p. 16.)

Weintraub, E. Roy

1983 "On the Existence of a Competitive Equilibrium: 1930-1954", *Journal of Economic Literature*, 21, 1, pp. 1-39. (Cit. on p. 150.)

Weintraub, E. Roy and Ted Gayer

2001 "Equilibrium Proofmaking", *Journal of the History of Economic Thought*, 23, 4, pp. 421-442. (Cit. on p. 148.)

Weisberg, Michael

2007a "Three Kinds of Idealization", *Journal of Philosophy*, 104, 12, pp. 639-659. (Cit. on p. 64.)

2007b "Who Is a Modeler?", *British Journal for the Philosophy of Science*, 58, pp. 207-233. (Cit. on p. 132.)

2012 "Getting Serious about Similarity", *Philosophy of Science*, 79, 5, pp. 785-794. (Cit. on pp. 56, 71.)

2013 *Simulation and Similarity. Using Models to Understand the World*, Oxford University Press, Oxford. (Cit. on pp. 16, 71, 91-92, 102, 116, 137.)

West, E. G.

1970 "Review of The Individual in Society: Papers on Adam Smith", *Economica*, 37, 145, pp. 93-95. (Cit. on p. 147.)

Wilkenfeld, Daniel A.

2017 "MUDdy Understanding", *Synthese*, 194, 4, pp. 1273-1293. (Cit. on pp. 19, 49, 52.)

Wilkenfeld, Daniel A. and Jennifer K. Hellmann

2014 "Understanding beyond Grasping Propositions: A Discussion of Chess and Fish", *Studies in History and Philosophy of Science Part A*, 48, pp. 46-51. (Cit. on p. 35.)

Wilkenfeld, Daniel A. and Tania Lombrozo

2018 "Explanation Classification Depends on Understanding: Extending the Epistemic Side-Effect Effect", *Synthese*. (Cit. on p. 170.)

Wilkenfeld, Daniel A., Dillon Plunkett, and Tania Lombrozo

2016 "Depth and Deference: When and Why We Attribute Understanding", *Philosophical Studies*, 173, 2, pp. 373-393. (Cit. on p. 170.)

2018 "Folk Attributions of Understanding: Is There a Role for Epistemic Luck?", *Episteme*, 15, 1, pp. 24-49. (Cit. on p. 170.)

Williamson, Timothy

2007a "Philosophical Knowledge and Knowledge of Counterfactuals", *Grazer Philosophische Studien*, 74, 1, pp. 89-123. (Cit. on p. 156.)

2007b *The Philosophy of Philosophy*, Blackwell, Malden, MA. (Cit. on p. 171.)

Woodward, James

2003 *Making Things Happen. A Theory of Causal Explanation*, Oxford University Press, New York. (Cit. on pp. 3, 20, 30, 44, 59, 62, 64, 71-72, 79, 85, 90, 96-97, 104, 132, 154.)

2006 "Some Varieties of Robustness", *Journal of Economic Methodology*, 13, 2, pp. 219-240. (Cit. on p. 165.)

Wright, Cory

2015 "The Ontic Conception of Scientific Explanation", *Studies in History and Philosophy of Science Part A*, 54, pp. 20-30. (Cit. on p. 172.)

Ylikoski, Petri

2009 "The Illusion of Depth of Understanding in Science", in *Scientific Understanding. Philosophical Perspectives*, ed. by Henk W. de Regt, Sabina Leonelli, and Kai Eigner, University of Pittsburgh Press, Pittsburgh, pp. 100-119. (Cit. on pp. 17, 29, 31-32, 36, 41, 51, 63, 83.)

2013 "Causal and Constitutive Explanation Compared", *Erkenntnis*, 78, 2, pp. 277-297. (Cit. on pp. 18, 29-30, 72-73.)

2014 "Agent-Based Simulation and Sociological Understanding", *Perspectives on Science*, 22, 3, pp. 318-335. (Cit. on pp. 29, 63, 165.)

Ylikoski, Petri and N. Emrah Aydinonat

2014 "Understanding with Theoretical Models", *Journal of Economic Methodology*, 21, 1, pp. 19-36. (Cit. on pp. 13, 18, 25, 29, 89, 91-92, 95, 99, 107, 115-116, 125, 139, 143, 159, 164, 172.)

Ylikoski, Petri and Jaakko Kuorikoski

2010 "Dissecting Explanatory Power", *Philosophical Studies*, 148, 2, pp. 201-219. (Cit. on pp. 17, 29-31, 41, 72, 170.)

Zangwill, Nick

2017 "Epistemic/Non-Epistemic Dependence", *Noûs*, pp. 1-22. (Cit. on p. 73.)

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić. The style was inspired by Robert Bringhurst's seminal book on typography *"The Elements of Typographic Style"*. `classicthesis` is available for both LaTeX and LyX:

`https://bitbucket.org/amiede/classicthesis/`