The Use of Models in Economics

Melissa Vergara Fernández

The Use of Models in Economics
Melissa Vergara Fernández
ISBN 978-94-6299-947-3
2018

Photos: Arthur van der Laan
Cover Design and inside work: Arthur van der Laan and Melissa Vergara Fernández
Print: Ridderprint BV, The Netherlands

THE USE OF MODELS IN ECONOMICS

Het gebruik van modellen in economie

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

Friday, 29 June 2018 at 09:30 hrs

by

Melissa Vergara Fernández
born in Bogotá, Colombia

Erasmus University Rotterdam

Doctoral committee

Promotors:	Prof. Dr. J. Reiss
	Prof. Dr. J.J. Vromen
Other members:	Prof. Dr. M. Morgan
	Prof. Dr. R. Backhouse
	Dr. M.J. Boumans

To my father, who despite having been born in the thirties in a small conservative town by the Colombian Caribbean coast, instilled in me that I am capable of accomplishing anything I intend.

A mi padre, quien a pesar de haber nacido en los años treinta en un pequeño pueblo conservador de la costa de Colombia, siempre me inculcó que soy capaz de lograr aquello que me proponga.

Contents

Preface and Acknowledgements

I started this project with the hope that I'd be able to bring clarity into how it is that models, with their patent falsities, help scientists, and economists in particular, to learn about the world. Specifically, I was interested in recovering the aspects of the philosophical literature that could be relevant for economists to be aware of in their use of models. The philosophical literature was huge; philosophers had been concerned with models for a few decades and the explosion of research on this area in the last two had been remarkable. I remember that, when I told Marcel Boumans that I was considering writing my dissertation on models, he suggested that I reconsider: the literature on models was already so vast that there was barely anything original that could be said. I pursued it nonetheless because I thought I could write a succint philosophical account of both the power and limitations of models addressing economists. The project would be more a matter of translation into a language that appealed to economists than anything else. Philosophers would then, finally, be able to tell economists the extent to which they could trust their models; they could not be trusted blindly, as had allegedly been the case with the financial crisis. In a sense, I thought that the hardest part of the job would be to get the attention of economists; after all, they are known for their arrogance and for the outright dismissal of methodology and the history of economics, as the scrapping of these subjects in economic departments around the world attests.

Even though I think the project of writing about models for economists and other social scientists—instead of for our philosopher peers—is an important one that still needs to be undertaken, at the time it was misguided for at least two reasons. First, because I didn't have enough knowledge of the economics discipline. The more I read about economics the more I realise how little I knew then and how little I still know. Obviously, addressing economists involves to be able to speak in a language with which they are familiar and to have common understanding of the discipline. In this respect, I've come to understand better why economists have simply opted for not paying attention to outsiders: plenty of criticism has drawn a radically distorted picture of the discipline with which they obviously do not identify with.

Second, because despite the burgeoning philosophical literature on models in the last couple of decades, there is still much to be written about models and how they are used, before a 'translation exercise', as I intended it, can be carried out. For a long time while writing this dissertation I doubted my ability to understand the philosophical literature on models. I felt that the fact that I didn't have a solid background in philosophy was the reason why I couldn't really make sense of how a great deal of this literature could inform the practice. Only late in this process I realised that perhaps it was not me who was trapped between two disciplines, neither of which I had a good grasp of. It was possible instead that not all the philosophical literature on models has the purpose of informing the practice or of being relevant to it. Indeed, the philosophical literature on models is massive, but not all of it has had the purpose of having a dialogue with the practice. Part of it pursues instead projects that are mostly of philosophical interest and respond to questions that have been traditionally addressed by philosophy of science.

This is ultimately what this dissertation became about. The suggestion is that in addressing questions that are traditional of philosophy, rather than a response to how models are used in practice, much of what is actually relevant has not received as much attention as it should. In relation to the original project I had, I came to realise that before philosophers can engage in a project addressing economists, a more comprehensive characterisation of the practice is necessary. Another way to put this is that, far from the philosophical literature having addressed everything there is to address about models, there are important aspects of the use of models that still require attention, in particular by what can be regarded as the mainstream literature in philosophy of models. As such, the dissertation became a rather critical project of current philosophical accounts that doesn't fully satisfy me. I wish I could have been more constructive.

* * *

As a symbol of the culmination of a project like this one, the book, the material outcome, fails in many ways to convey the significance it has. My journey was sweet, salty, bitter, sour, umami. Exciting, inspiring, painful. Unique. There's simply no way to convey the significance of such a rollercoaster. A non-trivial way to ameliorate this deficiency though, to be able to somehow convey the elaborate convolutions of the path and add to the significance of the final product, is to stamp on it the names of those that, in one way or another, contributed to the carving. In my case, the added significance is massive.

There are some without whom I doubt I would have managed to finish this journey. To them I owe that I made it this far. My partner Menno Rol, for his unconditional support. His patience and confidence that I was capable of doing this were, for long while, the only source of energy I had. He built my home when I had none and kept me fed with Michelin-star creations and matching libations. Rolf Viervant, Monique Pietermaat, Liesbet van Zoonen, and Marcel Boumans were there to support me in the darkest hours. Rolf helped me to navigate and protect myself from the nastiest bits of Dutch bureaucracy, incompetence and arrogance. He also introduced me to and helped me understand the poetry of Martinus Nijhoff and Annie Schmidt, the short stories of Herman Pieter de Boer, and the Dutch version of Hamlet. Monique constantly checked on me to make sure I wouldn't falter. I showed up one day at Liesbet's office feeling lonely, unsupported and mistreated. Liesbet barely knew me and accepted to listen and counsel me. Later, when I got a teaching job at the University of Amsterdam, she and her partner Jaap welcomed me in their house. I felt at home. Marcel, when in Rotterdam, taught me a lot about Macro. He was also one of the very few patient enough to let me practice my Dutch when it was still incipient. He recommended me for teaching jobs, first at Utrecht University and then at the UvA. Thanks to Marcel I managed to sustain myself financially and stay in the Netherlands to work on my dissertation. Rolf, Monique, Liesbet, Marcel: misschien beseffen jullie niet hoe ontzettend belangrijk jullie steun voor mij is geweest. Er zijn geen woorden om uit te drukken hoe dankbaar ik ben.

My sister Sharon, almost nine thousand kilometres away, supported me over the phone. She never lost her faith in me and kept reminding me of it. She also took care of my debts in Colombia when I couldn't pay them anymore so I could focus on my dissertation.

My lawyer Berber Swart kept we away from the claws of the immigration service. When the IND notified me that I wasn't welcome in the European Union anymore and that I'd have to pack my life back and leave, she let them know I wouldn't be going anywhere—at least not back to the tropics. I did move, up north, beyond what once was known as the Dutch Siberia. Her understanding of my situation, the "vanzelfsprekendheid" with which she accepted and handled it were the most comforting and reassuring.

My supervisor Julian Reiss left EIPE six months after I started my PhD. I'm pretty sure it would have been much more convenient for him to ask me to look for another supervisor. I am indebted to him for not having abandoned me. I think having him close would have had helped me to write a better dissertation; I probably wouldn't have got lost as much as I did, and I would have had the opportunity to share and refine some ideas before they actually hit the paper. Instead, Julian had to read and comment on ideas that were dead-ends since the beginning. The ones that actually made it to this book are definitely much better and refined thanks to him. I must admit I'm not sure how much of Julian's thinking I've actually "inherited"; the distance didn't help and the maturity to see that is something I feel I've just recently began to acquire. But I'm sure that if I've become sharper at all, it's because I knew Julian would not be impressed with flimsy ideas. Julian, thank you.

My friends and paranymphs, Attilia Ruzzene and Luis Mireles-Flores have been an extraordinary support and source of inspiration. They are the example I have followed since I, as a kid, arrived in the Netherlands. At the office, in our tiny student living rooms, and at a bunch of Rotterdam bars, we've had all kinds of conversations, profound and shallow. Some are more significant than others. But they all allowed me to piecemeal discover what philosophy can do and the one I learnt to cherish. Attilia has been a great influence in my view about the role we can and should play as (women) philosophers and academics in 21st century academia and society. It's a question that concerns me deeply and one that has defined the path I've decided to walk from here onwards. I owe that to her. When everyone else slept, Luis was there to read my work. His support is invaluable. Luis has been crucial in helping me build up the self confidence that once faltered.

Arthur van der Laan shared his life with me. He allowed me to keep in perspective what I was doing and to cherish the often short but magical moments that life brings. We shared hundreds of adventures that would refuel my soul to continue pursuing this lonely project. He also did the graphic work of this dissertation.

Colleagues I had at EIPE allowed me to sharpen my ideas and to learn from theirs. François Claveau, my first and only co-author so far, taught me a lot during our collaboration, perhaps much more than he's aware of. He also introduced me to Zotero, surely the most precious tool one can have in academia—if it's compatible with your word processor. With Philippe Verreault-Julien I shared ideas, the office (until he got a mechanical keyboard, then I couldn't bear it anymore) and philosophical literature. His generosity (and his partner's) are much appreciated. Çaglar Dede is the one with whom I shared the most academic interests. Our mutual interest in economics (in the way we're interested in it) lead to great conversations about philosophy, economics and life. It's too bad he kept his office mate James Grayot too much for himself. Otherwise I would have been able to learn more from James's sharpness and enjoy his jokes. Vaios Koliofotis's remarkable genuineness was always refreshing to have around, especially if he was in a good mood.

Conrad Heilmann is probably the one who best knows the turns, loops, ups and downs this rollercoaster took. I appreciate that he was always willing to help and dedicate time for which I'm pretty sure he wasn't properly compensated. His dedication to EIPE, in particular to make it more transparent and organised is really remarkable. I seriously doubt whether EIPE would have made it so far in the last years if it weren't for him. Jack Vromen was supportive all along, and particularly in the very last bits. I remember with great fondness the conversations and laughs we had at conference dinners and EIPE seminars. Ticia Herold, Ursula David, and Judith Gulpers were always helpful. Ticia wrote many last-minute letters and certificates for me. My baby steps in Dutch were thanks to Lizzy Patilaya. Manon Geluk taught me to "sjoelen"—crucial to integrate in Dutch society.

I profited from the comments and discussions with many colleagues at conferences and academic gatherings. To them I owe them their time and dedication. Some that I feel have contributed significantly, in different ways, for what they have taught me about philosophy, economics and/or academia are Deirdre McCloskey, Wade Hands, Alan Kirman, Pedro Garcia Duarte, Marcel Boumans, Roger Backhouse, Harro Maas, David Teira, Till Grüne-Yanoff, Kevin Hoover, Caterina Marchionni, Robert Northcott, Gonçalo Fonseca, Carlo Martini, Chiara Lisciandra, Aki Lehtinen, Emrah Aydinonat, Fred Muller, Ingrid Robeyns, Uskali Mäki, Harold Kincaid, José Edwards, Juan Manuel Durán (who went shopping for shoes with me in Helsinki), Magdalena Malecka, Sonja Amadae, Tom Wells, Mario Huzel, Constanze Binder, Beatrice Cherrier, Roberto Fumagalli, Asgeir Brynjar Torfason, Juan Acosta, and Roel Visser. Veronica Santos welcomed me in her home during my writing retreat in Berlin in the summer of 2017 and Marc Adam helped me to be disciplined. Jimena Hurtado Prieto and Christian Jaramillo are possibly the ones who have had the most influence in me; thanks to them I learnt about philosophy of economics when I was a kid. Federica Russo and Don Ross are two scholars I admire profoundly and from whom I've learnt a lot. I'm not sure whether in this dissertation I managed to do justice to the lessons I've learnt from them. Perhaps not. But they remain a source of inspiration for the work I do. Thank you both.

Thanks to INET's Young Scholars Initiative I have had the opportunity to meet many great scholars, attend good and bad conferences, and meet dozens of inspiring and inspired young economists—some incipient collaborators; yes, Nomi, I'm looking at you! The whole YSI experience was great fun and contributed to my understanding of economics as a discipline, often seen from the perspective of outsiders, who resent the state of the art of the discipline.

Friends who have been part of this rollercoaster and to whom I owe a debt of gratitude are Ana Bobinac, Elisabeth does (amazing office mate), Marco Meyer, Jens van 't Klooster, Nomi Friedman-Sokuler, and Pip Ridley. I will always be indebted to my friend Pedro Llosa Vélez who welcomed me in his home, without knowing me, in the old Rotterdamsche Macondo. Pedro and Joshua Graehl definitely set the winds that would blow for me from the moment I set foot in the Flatlands. You boys are amazing and I love you.

My family, without having much of an idea about why someone would want to write a dissertation in philosophy, or what that is useful for, were supportive at a distance. They would pamper me everytime I'd head back to Colombia, perhaps hoping I'd decide to go back for good. Without knowing that, no matter what happens, they will always be there, I wouldn't even have dared to set foot in the airport to start this journey.

Finally, I want to thank my small and large committees for accepting to be part of this process. I appreciate dearly the time and commitment dedicated to this dissertation. My small committee, Mary Morgan, Roger Backhouse and Marcel Boumans posed important and challenging questions particularly about the disciplinary boundaries that I attempt to call into question in the dissertation and that I, inadvertendly, continued to reproduce. I've learnt. Thank you all.

1

Introduction

Almost ten years have passed since September 15, 2008, the day on which the investment bank Lehman Brothers, after 158 years of operations, filed for Chapter 11 bankruptcy protection. Governments around the world, attempting to avoid a complete financial debacle, bailed out many other banks. Nonetheless, the Great Recession ensued. Since then, mountains of pages have been written about what went wrong, including how economists and their models were incapable of foreseeing the confluence of events that was under way, and how macroeconomic theory needs to be modified to prevent financial meltdowns of such a magnitude from happening again.

It is particularly significant that the confidence in economic models and the authority economists using them enjoy, based on their command and understanding of the economy, has dwindled. Prior to the financial crisis the economics discipline enjoyed a long period of credibility and self-confidence. At least in macroeconomics, according to some commentators such as Robert Lucas (2003), the central problem of the macroeconomics of depression prevention had been solved for all practical purposes. Many macroeconomists thought they had sufficient understanding of the economy and technical means to influence it, to be able to generate steady growth and avoid deep recessions; the policy of inflation targeting, for instance, has managed to bring inflation under control in many developed economies for a sustained period of time. Not only due to the devastating consequences of the crisis, but also due to this excessive confidence that arose as a consequence of The Great Moderation, the discipline has experienced massive criticism from both within and outside. This has been directed specially towards macroeconomic models, their realisticness—or rather lack of it—and the implications that follow from them for public policy[1].

Interestingly, in the nearly ten years since Lehman Brothers filed for bankruptcy protection, the philosophical literature on economic models has had little to say, if anything at all, about this situation. To be sure, philosophers have long discussed issues such as the realisticness of assumptions—e.g. Mäki (1998, 2009)—or the use of idealisations—e.g. Hoover (2010); Weisberg, (2007)—which are issues that come to light in the claims made by critics. But the causes of the crisis and some issues raised by the models that were used by monetary authorities, governments, rating agencies, banks, etc., are broader than just issues of idealisation or the realisticness of assumptions. The neglect is particularly worrisome if, as Hands (2015) has suggested, economic methodology is an inferior good—one that is consumed more when incomes fall. The years of the

[1] Some of the criticisms of current models are their failure to account for asymmetric information, or the use models in which there is a single representative agent. See Stiglitz (2011, 2015) for a discussion.

recession were an excellent opportunity to sell some philosophy of models, and yet little of it was produced.

This situation can be attributed, at least partly, to the fact that philosophers of economics, in general, have seldom engaged with macroeconomics (Hands, 2015; Ross, 2014). In his attempt to urge for a philosophy of macroeconomics, Ross (2014) suggests that the lack of interest in macro is partly the result of an interaction between the belief of mainstream economists that macroeconomics should have microeconomic foundations and an "occupational bias" on the part of philosophers to engage with the logical and epistemological foundations of the sciences. If the foundations of economics are indeed of microeconomic character, Ross suggests, then philosophers have no reason to engage with macroeconomics.

I think the "occupational bias" that Ross (2014) has suggested is particularly manifest in the current modelling literature. The interest of philosophers in models arose when these began to be seen as vehicles of scientific knowledge; something that was previously attributed to theories alone. Since philosophers upgraded models in the ranks of scientific significance, the challenge has been to explain what makes these former second-class citizens of the scientific world worth of their ascension. They clearly deviate from the phenomena they are meant to describe[2] due to the many idealisations and abstractions used in building them, and yet they allow model users to learn about phenomena.

Models climb up the ladder of significance

Models, as objects of philosophical analysis have a rather recent history. Nancy Cartwright's (1983) *How the Laws of Physics Lie* had a profound impact on how models were seen in relation to theories[3]. An important claim she made in this book was that it is not because of their truth that laws are explanatory. To the contrary, Cartwright argued, laws such as those of quantum mechanics fail to give true accounts of phenomena in the same way that phenomenological laws do. Theories are thus not about concrete phenomena. Instead, a model, "a specially prepared, usually fictional description of the system under study" (1983, p. 158) is used, in order be able to apply the theory to reality, and be able to explain a phenomenon (which in Cartwright's view is to find a model that fits into the basic framework of the theory). With this new account of explanation, the simulacrum

[2] I'm here using the term 'describe' following Bailer-Jones's (2009) characterisation of models as interpretative descriptions of phenomena.

[3] Another, earlier, work of importance is Hesse (1966), but her book was very exceptional at the time.

account, Cartwright argued that models are essential for a theory to be explanatory; this was a role that models at the time were not widely believed to have.

At the same time, there was a conscious attempt by some to demote theory from the central role it had. There was an ongoing "revolution against a theory-dominated" view of science that prevailed among some philosophers, sociologists and historians (Cartwright, Shomar, & Suarez (1995)). The idea of this movement was to reconsider the view that philosophy of science was concerned with scientific theories; instead, they claimed, it should be concerned with scientific knowledge in general, and knowledge not only comes in the form of theories. Under this new conception of philosophy of science as a tool-box (Cartwright et al., 1995), theory was just one of the tools available to build models, which are the ones that ultimately represent phenomena (Cartwright et al., 1995, p. 140):

> I want to urge that fundamental theory represents nothing and there is nothing for it to represent. There are only real things and the real ways they behave. And these are represented by models, models constructed with the aid of all the knowledge and techniques and tricks and devices we have. Theory plays its own small important role here. But it is a tool like any other; and you can not build a house with a hammer alone.

With this new, tool-box view of philosophy of science, Cartwright argued for an instrumentalist view of science. That is, that our scientific understanding and its corresponding image of the world, which is present in our theories, as much as in our instruments, mathematical techniques, methods of approximation, etc., should be seen as adaptable tools in a common scientific tool-box, instead of as claims about the nature and structure of reality, expressed as propositions that are true or false (Cartwright et al. (1995)). In this view, theories had been demoted to be just a tool in the scientific edifice and models had been attributed a higher status, being the ones that actually represented the world and thereby provided the means for the justification of scientific knowledge.

The final spur given to models, in which they were given not only a high status but also autonomy from theory came with the work of Morgan & Morrison (1999). In this edited volume, commentators, and in particular Morrison & Morgan (1999), argued for the partial independence of models from theory, by models being autonomous agents that function as investigative tools that enable learning about both theory and the world. Morrison & Morgan (1999) offered an account of models as mediators, in which they are partially independent because of the way in which they are constructed; they function as tools, in the same way that construction tools allow houses and bridges to be built. Models are, however, more sophisticated; they are investigative devices that involve some kind of representation either of theory, the world, or both. This representational faculty is what allows model-users to learn about, again, theory, the world, or both. The way in which learning occurs, given that there is representation of the object of enquiry, is by manipulating models—toying around with them.

This account can be said to have established models as a separate epistemic genre (Morgan, 2012)[4], receiving the attention of philosophers because of what this new account represented for the role of models in science. In fact, commentators such as Kuorikoski & Marchionni (2015, p. 381) attribute the centrality of models in the epistemology of science to the work of Morgan & Morrison (1999):

> The recognition that models and simulations play a central role in the epistemology of science is about fifteen years old. Although models had long been discussed as possible foundational units in the logical analysis of scientific knowledge, the philosophical study of modelling as a distinct epistemic practice really got going in the wake of the Models as Mediators anthology […].

Arguably, therefore, the anthology of Morgan & Morrison (1999) nearly two decades ago, marked the triumph of the revolution that intended to remove theory from its throne. There's no question now among philosophers that models have a primary role in science. But, although it became clear that models have this primary role, it has been much less clear what precisely explains models' success. Previous accounts, including Cartwright's mentioned above and those in the *Models as Mediators* volume, have been based on single cases that, though enlightening, are not easily generalisable. There are too many kinds of models used for many purposes. Since then, therefore, the new revolution or, the new quest, philosophers have set for themselves has been to understand whether there are some properties that models have in common and, if so, what are those properties which have gained them their status as vehicles of scientific knowledge and how precisely do they accomplish this.

The mystery of models

Models are mysterious. At least for philosophers. Philosophers have come to recognise that models are at the centre of the scientific practice, that in the cases they have studied these are the objects that scientists use as vehicles of scientific knowledge, but that, given some known properties of models such that they generally fail to represent their targets accurately, it is unclear how they play such a significant role. Although models in many disciplines puzzle philosophers, in economics the mystery of models is particularly evident. Misrepresentation by models is

[4] See Hausman (2015) for a critique.

ubiquitous and for many even absurd—e.g. models with perfectly rational agents who possess all available information. Yet, at the same time, economists place a lot of confidence in their models, the discipline has significant authority in public discourse, and allegedly has some superiority over other social sciences, especially those that rely less on models and quantitative methods.

Some philosophers of science have been dedicated to solving the mystery of models by trying to identify the source of the mystery. They have placed bets on two sources in particular: their ontology (what models are) and their semantics (the representational relation of models to what they represent). Given how many kinds of models there are—e.g. scale models, sets of mathematical equations, etc.—one of the questions addressed in the ontology of models literature is what it is that makes a model a model and, specifically, whether there is any abstract property that unifies all these different things as models. Some philosophers dismiss this question as unimportant for the purpose of philosophy of science or for understanding scientific practice (e.g. French (2010)). Other philosophers have suggested that models' *functional role* is what unifies them and have tried to build an ontology around their function, describing them as functional entities (Gelfert, 2017). With respect to the semantics of models, the focus has been on in virtue of what a model represents its target. There are several views in this respect, the most popular being that models represent in virtue of being similar to their targets (Giere, 2004; Teller, 2001).

Other philosophers have attempted to solve the mystery by concentrating on the epistemology of models and, in particular, on the form in which the epistemic import of models comes. There are also 'sceptics' who argue that models are merely heuristic devices (Alexandrova, 2008; Alexandrova & Northcott, 2009; Hausman, 1992, Chapter 4). Some specific questions attempted in this literature are whether models can count as evidence (Reiss, 2008, Chapter 6; This dissertation, Chapter 2); whether they are explanatory (Alexandrova & Northcott, 2013; Bokulich, 2011; Reiss, 2012) and whether they yield understanding (Aydinonat, 2007; Kuorikoski & Ylikoski, 2015; Ylikoski & Aydinonat, 2014).

The underlying thought in this philosophical work is that models are a success story that still requires explanation as to where the success is coming from. For instance, Margaret Morrison (2015, p. 1) motivates her latest book on models and simulations as follows:

> How do reconstructive methods and practices of science, specifically those associated with mathematics, modelling, and simulation, impart information and knowledge of the world around us? An answer to that question defines my goal for this book.

The emphasis of her new book is on the relation between mathematics and modelling and addresses the related topic of simulation, which has begun to play a significant role in science, given how cheap and efficient computation has become. A few pages later, she continues (2015, p. 4, emphasis in the original):

> My interest in explicating that relation [between mathematics and physics] stems from an attempt to understand how the abstract nature of mathematics can nevertheless yield concrete physical information; how a reconstruction of reality by abstract mathematics can help us solve physical problems. I take it as a given that this relation is mysterious, but the mystery needn't prevent us from attempting to uncover at least some of its features. In other words, we can hopefully understand *how* mathematics can deliver physical information without understanding *why* it does.

Not all commentators are so explicit about the mystery of models. In great many cases the debates have become so specific, and so many different views concerning models have emerged, that the reason for which philosophers got interested in models in the first place, is no longer mentioned. Still, philosophers continue to attempt to explain the success of science, nowadays in the form of models and simulations. There are now different philosophical accounts of scientific representation specifically or, more generally, of modelling, whose aim is to explain how models are capable of making us learn about the world, and thereby explain their success.

The use of models: a look at the praxis

In this dissertation, I argue that philosophical accounts of modelling ought to look at the practice of modelling. Obviously, this doesn't sound like something new. It isn't something new. Thomas Kuhn's (1996) defence of history as a means to ground philosophical accounts in actual scientific practice has been taken on board by philosophical accounts. In fact, Cartwright's legacy of placing models at the centre of scientific enquiry in the philosophy of science discourse is a consequence of her method in philosophy, namely that arriving at a position in the philosophy of science is based on the observation of scientific practice (Bailer-Jones, 2008), which is itself another of the legacies of the so-called Stanford School of Philosophy of Science, of which Cartwright was a member. Attention for scientific practice is thus at the forefront of philosophical accounts and they generally aim at some descriptive accuracy by relying on at least one case study or example that illustrates or provides evidence for a philosophical claim.

My proposal here is for philosophical accounts of models to look at the practice as an end in itself. As I discussed above, the interest of philosophers in models has arisen because they are objects in scientific practice that answer an old philosophical question, namely, what the source of knowledge is. Models are seen as the objects that answer that question, and the attempt has thus been to find how they do it. Another way to put this is that, interest in models arises out of an *optimistic bias* to explain the scientific endeavour as a success. My proposal is instead for philosophical accounts of models to leave this optimistic bias behind and focus on the practice of modelling—in this particular case, economic modelling—as it is. This means that the practice is observed in all its complexity, which includes, aside from the epistemic aspects that justify the success that we attribute to science, the many non-epistemic elements that govern the use of models. In more concrete terms, getting rid of the optimistic bias implies among other things:

- Exploring more fields and subfields. As I mentioned at the outset, macroeconomics has been largely neglected in the philosophy of economics and in the modelling literature in particular[5]. Why doesn't ignoring an entire field, especially one with so many social consequences, generate anxiety among philosophers?

- Exploring a larger set of models than is currently done. In general, in the philosophy of models, a few workhorse models have been used as a basis for philosophical claims. In economics, in particular, only a handful of models has been explored—Schelling's model of spatial segregation in an old favourite.

- Exploring research questions rather than individual models. Philosophers have tended to focus on the representational properties or the epistemic import of a single model. But models generally don't stand on their own. Scientists and model users are mostly interested in a particular question or phenomenon for which different models and kinds of models are used. Why focus on just one?

- Exploring the different realms or scenarios in which models are used. Perhaps before it was easier to talk about science referring to academia and a few research institutions. Such a view fits with the underlying assumption that science's only aim is to learn about the world. But, is that still the case? Didn't Elon Musk's SpaceX, a private enterprise, just launched the Falcon Heavy, and has multiple contracts with NASA for resupplying the ISS? What about the research carried out by the pharma industry?

[5] There are some clear exceptions like Boumans (1999, 2005); Hoover (2010); Morgan (2012). However, the approach of Boumans and Morgan, in particular, is arguably more historical than philosophical. More on this in chapter 4.

- Exploring model failure. Not all uses of models are successful. What constitutes model failure? What lessons can be learnt from these failures?

Delving into these explorations might allow more accurate and nuanced philosophical claims—so not just, 'Are models explanatory?'—and at the same time allow us to explore new questions that arise out of this practice. I should thus emphasise that I do not mean to suggest that the old philosophical questions should be eschewed; we have learnt quite something from them, but I think we might be able to learn even more if the practice is investigated with less baggage than we currently do.

Let me offer a brief but telling example. I discussed above how models in philosophy have been upgraded from second-class citizens to role models (pun intended). I didn't mention explicitly, however, what being second-class actually meant. For at least the first half of the twentieth century, models were considered to matter merely temporarily; they were supposed to reflect ongoing thinking that was still imperfect and that could therefore not yet be considered a theory: "The received opinion was that good theories rendered models theoretically and practically redundant" (Bailer-Jones, 2009, p. 82). Another way to say this is that models were considered to play a mere heuristic role. Given this background, now that models have been upgraded, philosophers have not only tried to argue for the epistemic import of models, but also do some of these philosophers seem to feel rather uncomfortable with the thought that (some) models may just have that heuristic role, and perhaps nothing else. Here's Grüne-Yanoff (2013, p. 851):

> Philosophers, if they treat such cases at all [non-representational models], have by and large appraised such modelling practices as playing merely a heuristic role […] This heuristic justification is weak because success criteria for such functions are unclear in the extreme. Furthermore, it places the use of such models in the same category as taking a walk, reading the newspaper, or whatever else scientists do in order to inspire themselves to further theory development. Bunching important kinds of scientific modelling together with practices that cannot be rationally accounted for seems an unsatisfactory state, which this article seeks to repair.

Why is it unsatisfactory that perhaps, just perhaps, some models are as effective for scientists as reading the newspaper or taking a walk? Why anticipate that it *just can't be the case* that some models might be merely useful for thinking? Morgan (1999, 2012) is a commentator who has insisted on the importance of model manipulation and of first understanding "the world in the model", in order to later understand the world. And, in some of the interviews that Bailer-Jones (2009) conducted with scientists, some of them concede as much (p. 11):

> You've got some plausible models which comforts you because you can think, well, this is not a total mystery to me; I can imagine what might be going on here. I don't actually know what's going on here, but this is all right, we got some ideas. *Barrie Jones, physicist and planetary astronomer.*

The example is meant to show how some old philosophical baggage—in this case that to call a model a heuristic is met with anxiety—drives our questions and perhaps even the conclusions we draw from the look we give to the practice. After all, Grüne-Yanoff doesn't seem to be prepared to concede that non-representational models might just have the same effect on scientists as reading the newspaper for their scientific activity.

Before I go on to outline the contents of each of the chapters that follow, let me say something about why this fresh look at the practice is important for the philosophy of models. At the outset I mentioned the bankruptcy of Lehman Brothers. The fact itself is not as significant—though it is the biggest bankruptcy in U.S. history, with USD 613 billion debt (Mamudi, 2008)—as it is being a symbol of the biggest recession since the Great Depression in 1929. This is important because of the grave implications it has had socially, economically, and politically. And, it is widely accepted that the financial crisis and the poor handling of the recession was the making of bad models or, ultimately, of poor science. The point is that limiting philosophical accounts of models to epistemological concerns presupposes that science is only in the business of learning true things about the world. Obviously this is not the case. Science is *used*, often in the form of models, and learning about those uses is important to understand and assess the effects that those models might have.

In her "Philosophy of Science for the Twenty-First Century", Janet Kourany (2003)[6] urges a socially responsible philosophy of science. She makes reference to the logical positivists of the Vienna Circle, who advanced their philosophy of science with sight towards the interaction between science and society and the interest they had in the social movements of the time. She laments and criticises the purely epistemological concerns of philosophers of science—"The 'social', for these philosophers, stops at the doors of scientists' immediate environments" (p.5)—but contends that a new, socially responsible philosophy of science is emerging at the hands of feminists. Feminists are concerned with science because of the deleterious effects that scientific knowledge has had on the struggle of women for equality and because of the positive effects that science may have on that struggle. I don't want to claim here that I bear a feminist flag with this dissertation—perhaps regrettably, my interest in this literature came too late for that—but I do

[6] See the exchange that followed in Giere (2003); Kourany (2003b).

want to advocate a more socially responsible philosophy of science and, specifically, a more socially responsible philosophy of models. These mysterious objects are being used in an enormous number of domains, and understanding how they are being used is paramount to understand their effects on people and the ways in which they can be put to socially responsible uses.

The contents of this dissertation

The chapters that follow are observations. That is, each chapter has been written independently of the others and addresses different aspects of the extant philosophical literature on models to which I think the new look at the praxis mentioned above can bring important insights. Two main topics are discussed: what the unit of analysis of philosophical accounts of models is and should be, treated in chapters two and three, and model failure, treated in chapters four and five.

In chapter two, I address the question what unit of analysis of models philosophers should investigate. I argue that the philosophical investigation of models should be focussed on clusters or research questions, rather than on single models and their components, as has generally been done. I suggest that two specific philosophical questions, namely the representational character of models and the attempt to frame discussions of realism in terms of models may have guided the interest of philosophers towards individual models and model components. However, the practice as well as philosophical arguments that maintain that our models are incapable of fulfilling all the purposes we might have for them at once, are compelling reasons to explore how models are related. I discuss some of the literature on New Economic Geography (NEG) for a two-fold purpose. First, to address the claim that modelling in economics is mainly an endeavour to generate robust theorems about causal mechanisms by practising robustness analysis. I argue against this position. Second, to expand the analysis of models to other areas of enquiry.

In chapter three, François Claveau and I discuss three epistemic roles that models might play and offer sufficient conditions for a model to actually play that epistemic role. We use the traditional definition of knowledge as true justified belief (KATJB) as a basis to define learning and thus establish the conditions that a model would have to fulfil. The motivation of the chapter is that, while there has been a long interest from philosophers to defend the epistemic success of models, there has seldom in the literature been a clear definition of what precisely this epistemic benefit is. The attempt here is thus to, having defined learning as ‘coming to know’, establish the sufficient conditions that a model would have to fulfil in order to determine whether it can be said to have epistemic benefit. The three epistemic roles we discuss are, evidential, which states that models can count as evidence for a claim about the world; stimulating, which states that models can be a stimulus for carrying out empirical research, and revealing, which states that models can generate new hypotheses about the world.

Although this chapter might be considered to fall prey to the "occupational bias" that Ross (2014) refers to, it has three important features that correspond to the look at the praxis that I endorsed above. First, it is an exploration of a cluster of models or, rather, of a research question, instead of a single model. In this sense, it can be taken as an example of what was argued in chapter two. We use the Diamond-Mortensen-Pissarides (DMP) model[7], which is part of the search and matching theory that has been developed in economics since the 70s, as a case from which we pick out the potential epistemic roles models can have. Second, since our analysis is on a research question rather than on a single model, we are able to rely on the relations that exist between purely theoretical models, statistical models, and data. So, in our example, justification for believing a particular proposition about the world coming from the model is possible thanks to an network of beliefs that agents have come to have thanks to other sources such as empirical data. In this sense, models are not mysterious but rather another tool, among many, that are used to understand the world. Third, it is an enquiry in the field of labour economics, which has important relations with and implications for macroeconomics. As mentioned above, this is a field that has seldom been investigated by philosophers of economics concerned with modelling.

In chapter four I move to a different subject. In an attempt to make sense of what it means for macroeconomic models to have failed, which many commentators have argued in light of the financial crisis, I explore the philosophical literature for guides as to how the accounts that have been offered so far, can elucidate these claims of failure. My conclusion after surveying the literature is that there is little that has been explored by philosophers with respect to model failure and little in their accounts that can be used for making sense of this aspect of modelling. In general terms, this exercise suggests that, despite the great interest that models have received from philosophers, especially due to the acknowledgement that models play a major role in scientific practice, the reach of the literature has been constrained to comprise only three specific aspects. One of them is the almost exclusive focus on theoretical models (as individual units). This is particularly remarkable in economics, given the transformation that the discipline has gone through in the last decades to a more empirical (or applied) science. This transformation is not reflected in our philosophical accounts of models and raises important questions about how close philosophical enquiry actually is to the practice. The other two are the focus on explanation and

[7] The DMP model is known as a model (singular). However, strictly speaking it is a class of models that were developed throughout the years by the three economists mentioned above.

understanding, and on the identification of causal mechanisms. Surely these aspects are important for science and the use of models more generally. But they are not the only aspects driving science and the use of models. In studying only these, other aspects of the practice are underestimated.

In chapter five, the last one, I continue with the subject of model failure. I argue for the need of an explicit analysis of model failure and, specifically, for a pragmatic account of models. Such a pragmatic account, I argue, is capable of accommodating aspects that determine the outcomes of the modelling activity, and that cannot be accommodated by extant accounts of models. I discuss Uskali Mäki's (2017) account of models "[ModRep]" for two reasons. First, because, though introduced as an account of representation, it has been extended over the years to include pragmatic elements. Second, and more importantly, because Mäki (2017) suggests that [ModRep] is sufficient to accommodate model failure. I argue, based on a few examples of the practice, that some potential sources of failure could not be accounted for by such an account and therefore suggest three additional elements.

To conclude, two notes are in order. First, the focus of this dissertation is on economic models. This means that, whenever I make references to theory, I refer to non-empirical models, unless otherwise specified. While I recognise that the theory-model relation is an important subject in general philosophy of science, this is not the object of this dissertation. 'Theory' is thus used rather casually, without intending to take sides on the debate of what precisely models and theories are. This, I believe, should have no consequences for the points I try to develop in this dissertation.

The second note is about grammar. While it has by now become more or less standard in philosophy of science to use the feminine pronouns as a means to attempt to correct gender imbalances—at least in writing—the use of gender-neutral singular 'they' is not as common or perhaps even considered incorrect. I have made use of gender-neutral pronouns in some chapters of this dissertation because I think this is what gender balance (in writing) should ultimately be about. English is a language that allows neutral expression and I profited from that. I haven't used gender-neutral language throughout the dissertation because it has been a learning process for myself as well—I used to write in terms of 'he' alone.

References

Alexandrova, A. (2008). Making Models Count. *Philosophy of Science*, *75*(3), 383–404.

Alexandrova, A., & Northcott, R. (2009). Progress in economics: Lessons from the spectrum auctions. In H. Kincaid & D. Ross (Eds.), *The Oxford handbook of philosophy of economics*. Retrieved from https://philpapers.org/rec/ALEPIE

Alexandrova, A., & Northcott, R. (2013). It's just a feeling: why economic models do not explain. *Journal of Economic Methodology*, *20*(3), 262–267. https://doi.org/10.1080/1350178X.2013.828873

Aydinonat, N. E. (2007). Models, conjectures and exploration: an analysis of Schelling's checkerboard model of residential segregation. *Journal of Economic Methodology*, *14*(4), 429–454. https://doi.org/10.1080/13501780701718680

Bailer-Jones, D. (2008). Standing up Against Tradition: Models and Theories in Nancy Cartwright's Philosophy of Science. In S. Hartmann, C. Hoefer, & L. Bovens (Eds.), *Nancy Cartwright's Philosophy of Science*. New York, N.Y.: Routledge.

Bailer-Jones, D. (2009). *Scientific Models in Philosophy of Science*. University of Pittsburgh Press.

Bokulich, A. (2011). How scientific models can explain. *Synthese*, *180*(1), 33–45.

Boumans, M. (1999). Built-in justification. *IDEAS IN CONTEXT*, *52*, 66–96.

Boumans, M. (2005). How economists model the world into numbers. Routledge.

Cartwright, N. (1983). *How the Laws of Physics Lie* (First Edition). Oxford University Press, USA.

Cartwright, N., Shomar, T., & Suárez, M. (1995). The tool box of science: Tools for the building of models with a superconductivity example. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, *44*, 137–149.

French, S. (2010). Keeping quiet on the ontology of models. *Synthese*, *172*(2), 231–249. https://doi.org/10.1007/s11229-009-9504-1

Gelfert, A. (2017). The ontology of models. In *Springer Handbook of Model-Based Science* (pp. 5–23). Springer.

Giere, R. N. (2003). A new program for philosophy of science? *Philosophy of Science*, *70*(1), 15–21.

Giere, R. N. (2004). How Models Are Used to Represent Reality. *Philosophy of Science*, *71*(5), 742–752. https://doi.org/10.1086/425063

Grüne-Yanoff, T. (2013). Appraising Models Nonrepresentationally. *Philosophy of Science*, *80*(5), 850–861. https://doi.org/10.1086/673893

Hands, D. W. (2015). Orthodox and heterodox economics in recent economic methodology. *Erasmus Journal for Philosophy and Economics*, *8*(1), 61–81.

Hausman, D. M. (1992). *The Inexact and Separate Science of Economics*. Cambridge ; New York: Cambridge University Press.

Hausman, D. M. (2015). Much ado about models. *Journal of Economic Methodology*, *22*(2), 241–246. https://doi.org/10.1080/1350178X.2015.1037546

Hesse, M. B. (1966). *Models and analogies in science* (Vol. 7). University of Notre Dame Press Notre Dame.

Hoover, K. D. (2010). Idealizing reduction: the microfoundations of macroeconomics. *Erkenntnis*, *73*(3), 329–347.

Kourany, J. A. (2003a). A philosophy of science for the twenty-first century. *Philosophy of Science*, *70*(1), 1–14.

Kourany, J. A. (2003b). Reply to Giere. *Philosophy of Science*, *70*(1), 22–26.

Kuhn, T. S. (1996). *Structure of Scientific Revolutions* (Third). University of Chicago Press.

Kuorikoski, J., & Marchionni, C. (2015). Broadening the Perspective: Epistemic, Social, and Historical Aspects of Scientific Modelling. *Perspectives on Science*, *23*(4), 381–385. https://doi.org/10.1162/POSC_e_00179

Kuorikoski, J., & Ylikoski, P. (2015). External representations and scientific understanding. *Synthese*, *192*(12), 3817–3837. https://doi.org/10.1007/s11229-014-0591-2

Lucas, R. E. J. (2003). Macroeconomic Priorities. *American Economic Review*, *93*(1), 1–14. https://doi.org/10.1257/000282803321455133

Mäki, U. (1998). Aspects of Realism about Economics. *Theoria: An International Journal for Theory, History and Foundations of Science*, 13(2(32)), 301–319.

Mäki, U. (2009). Realistic Realism about Unrealistic Models. In H. Kincaid & D. Ross (Eds.), *Oxford Handbook of the Philosophy of Economics*. Oxford University Press.

Mäki, U. (2017). Modelling Failure. In Hannes Leitgeb, I. Niiniluoto, P. Seppälä, & E. Sober (Eds.), *Logic, Methodology, and Philosophy of Science: Proceedings of the Fifteenth International Congress*. College Publications. Retrieved from https://pdfs.semanticscholar.org/5332/6e9790dc24be8d3597ec98b8a2fdda541bde.pdf

Mamudi, S. (2008, September 15). Lehman folds with record $613 billion debt. Retrieved 16 February 2018, from http://www.marketwatch.com/story/lehman-folds-with-record-613-billion-debt

Morgan, M. S. (1999). Learning From Models. In M. S. Morgan & M. Morrison (Eds.), *Models as Mediators: Perspectives on Natural and Social Science* (pp. 347–388). Cambridge University Press.

Morgan, M. S. (2012). *The World in the Model*. Cambridge University Press.

Morgan, M. S., & Morrison, M. (1999). *Models as Mediators: Perspectives on Natural and Social Sciences*. Cambridge; New York: Cambridge University Press.

Morrison, M. (2015). *Reconstructing reality: models, mathematics, and simulations*. Oxford ; New York: Oxford University Press.

Morrison, M., & Morgan, M. S. (1999). Models as Mediating Instruments. In M. S. Morgan & M. Morrison (Eds.), *Models as Mediators: Perspectives on Natural and Social Science* (pp. 10–37). Cambridge University Press.

Reiss, J. (2008). Error in Economics: Towards a More Evidence–Based Methodology. Routledge.

Reiss, J. (2012). The explanation paradox. *Journal of Economic Methodology*, *19*(1), 43–62.

Ross, D. (2014). Philosophy of Macroeconomics and Economic Policy. Retrieved from http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935314.001.0001/oxfordhb-9780199935314-e-47

Stiglitz, J. E. (2011). Rethinking Macroeconomics: What Went Wrong and How to Fix It. *Global Policy*, *2*(2), 165–175. https://doi.org/10.1111/j.1758-5899.2011.00095.x

Stiglitz, J. E. (2015). Reconstructing Macroeconomic Theory to Manage Economic Policy. In É. Laurent & J. L. Cacheux (Eds.), *Fruitful Economics* (pp. 20–56). Palgrave Macmillan UK. https://doi.org/10.1057/9781137451057_3

Teller, P. (2001). Twilight Of The Perfect Model Model. *Erkenntnis*, *55*(3), 393–415. https://doi.org/10.1023/A:1013349314515

Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy*, 639–659.

Ylikoski, P., & Aydinonat, N. E. (2014). Understanding with theoretical models. *Journal of Economic Methodology*, *21*(1), 19–36. https://doi.org/10.1080/1350178X.2014.886470

2

More Models: Clusters as the Unit of Analysis

More Models: Clusters as the Unit of Analysis

Introduction

For a great part of the 20th century, theories were regarded by philosophers as the vehicles of scientific knowledge. Models, instead, were regarded as merely tentative and unfinished scientific products. This changed in the last few decades, in which models became more and more prominent in scientific practice and philosophers became more attentive to what actually happens in scientific practice. They thus turned their attention to these objects, specifically trying to understand their ontology, their epistemology, and their relation to the world in terms of both reference and truth.

Philosophers of science have focussed on a limited number of models in this quest to understand the role of models in science. That is, some models have become "paradigmatic" in being the object of attention of philosophers and on which different accounts of models have been based. The Lotka-Volterra model in biology, Schelling's model of spatial segregation in economics, and the San Francisco Bay model in fluid dynamics are a few examples of these paradigmatic models. Two important reasons for why these models in particular have received attention are that they are relatively simple, and can therefore be easily explained and discussed without a need for technicalities and, more importantly, that these models represent some of the different kinds of models that have been so far identified: mathematical, abstract, and scale models, respectively.

This approach of studying individual, simple, and representative models has some advantages. It has, for instance, allowed philosophers of different fields to communicate with each other and thereby contribute to a common, general understanding of models, without necessarily having the same background knowledge. However, it is remarkable that little attention has been given to the relationship among models. Models are seldom built from scratch or used in isolation. Models usually respond to a specific research question, which is shaped by the development of earlier models, and which in turn determines the inputs or ingredients with which a model is built, its purposes, and its contribution.

In this chapter I shall argue that acknowledging the ways in which a model is related to other models may have important implications for the assessment that we as philosophers make with respect to the epistemic import of models. By focussing exclusively on the components of individual models, we might be overlooking epistemic import coming from models as clusters.

In the next section I discuss two philosophical questions that have been the object of enquiry of philosophers concerned with models and which I take to be important in having guided the attention of philosophers to the internal components of a model. Then I briefly discuss a few characterisations of models that take for granted the importance of the relationship between a model and its target. In section III, I discuss two exceptions in this literature: a view that has characterised theoretical modelling as mostly derivational robustness analysis and a view that defends how understanding is possible with clusters of theoretical models. Even though I agree with the spirit of the contributions, specifically their attempt to study clusters rather than individual models, I think their diagnosis of epistemic import is hasty. I suggest that before we attempt to draw conclusions about the epistemic contribution of models, we need to study more of them and the relations they have to other models. Any conclusion drawn from the study of a handful of single models should be merely tentative. Finally, in section IV, I offer a sketch of how a view of understanding in the literature in epistemology can contribute to our assessment of the epistemic import of models.

2. One model

Models do not generally stand on their own. Instead, they are related to other models. Sometimes this will be because they borrow convenient functional forms from other models, or because their results are consistent with those of other models, or because they respond to a specific research question that is treated from different perspectives. Kevin Hoover (1990), for instance, though in a different context, has traced the links and history of the models that are regarded as belonging to the New Classical school. The same could be done for a number of schools in economics and probably for different disciplines. Furthermore, individual models are generally quite limited for all the functions we would like them to perform. In terms of model qualities, there seems to be a trade-off between generality, precision and realisticness (Levins 1966; Odenbaugh 2003)[1]. And, in terms of purposes, some cases suggest that mechanistic models, because they aren't as adaptable, might perform worse at prediction exercises than simpler, non-mechanistic models (Reiss 2007).

However, philosophers concerned with understanding models in scientific practice have paid almost exclusive attention to individual models. That is, most of their accounts have been based on the examination of a single model and their properties. This situation is striking considering that one of the reasons for why philosophers have turned to models is because of their ubiquity

[1] Orzack and Orzack and Elliot have contested this claim.

in scientific practice and the interest that philosophers have in understanding their role in this practice.

I think there are at least two important philosophical questions that have prompted philosophers to pay more attention to individual models than to the relations that they have with other models. The first question is related to the representational character of models.

As argued by commentators such as Cartwright (1983) and Giere (2004), it is models, instead of theories, that represent phenomena. This feature of models is now accepted and therefore, in the last decades, interest has turned to more specific details about this relation of models with the world. Specifically, these have been questions that seek to understand the constituents of scientific representation, or broadly speaking, what scientific representation is[2]. Naturally, since these questions are concerned with the representational relation that stands between a model and its target, attention has been given to models and their properties and that of their targets. The literature has grown quite fast, partly because many different accounts have been offered, and all of them face objections and counterexamples. This has thus given impetus to the literature as amendments and new proposals are brought to the table.

Furthermore, some of these accounts, in particular the reductive ones, which are those that intend to explain away scientific representation in terms of more 'basic' notions, require models and their targets to be analysed in terms of their properties. For instance, a popular reductive view is Ronald Giere's account of representation as similarity. According to this view, scientists pick some specific features of models that are taken to be similar to features of the designated real target. And, it is the existence of these specified similarities that makes possible the use of the model to represent the real system, in the way specified by the modeller. (Giere, 2004). This account, and any other that is reductive inevitably demands that attention be placed in models, their properties, and those of the target system. In other words, it demands an inward look into models and their targets. This, for the simple reason that to determine whether the model is similar (or, say, isomorphic) in the required respects to its target, a comparison has to be made between some properties of the model and the target.

The second question concerns issues of realism, such as the role of idealisations in models and what this entails for scientific realism. Debates on realism with respect to science have traditionally taken place in relation to theories and laws. However, since models are now regarded as vehicles of scientific knowledge, their false assumptions raise questions about the exact role they play.

[2] Frigg & Nguyen suggest that there are, in fact, at least five questions related to scientific representation in the literature. I discuss these in Chapter four and will therefore not discuss them further here.

Galilean idealisations, for instance, have been characterised as removing disturbing factors for the practical purpose of making models tractable, and that can later be de-idealised for more accurate representation—and thereby more accurate predictions. The implication of this characterisation is a defence of realism, as the aim and the success of these models is ultimately rendered by accurate representations. Some counterarguments to this view are that scientists seldom de-idealise their models and, particularly in economics, it is often unclear how assumptions would have to be changed in order to de-idealise them. Since models are considered to be successful in conveying true insights about the their targets, the role of idealisations and the epistemic import of models raises philosophical questions that philosophers have been eager to answer.

In the attempt to provide answers to these and related questions, philosophers have turned their attention to models and their components. Let me briefly discuss some of of the accounts that have been defended by commentators that address them. I'll focus on discussions in the philosophy of economics because this is ultimately what interests me in this dissertation. Uskali Mäki has contributed to both of the debates mentioned above and, specifically in economics, has defended the method of isolation as a central method in economics (1992), arguing that unrealistic, Galilean assumptions are what allow modellers to isolate a causal mechanism. He has also defended the functional decomposition approach (2009), which focusses on the individual components of models and the functions they have, and by which he has defended a realist position, locating truth inside the model (2011). Mäki (2009, 2009, 2011) has offered an account of scientific representation in which his view of idealisation and realism come together.

Other commentators have focussed on the epistemology of models, addressing issues of representation or idealisations more indirectly. Bob Sugden (2000, 2011) has addressed the question of how economists make inferences from models and what kinds of inferences these are. He has argued that the credibility a model-user places in a model is what determines whether inferences can be made about the world. How credible the model is depends on the relation of similarity there is between the model and the target. Morgan & Morrison (1999) have focussed on the nature of models as autonomous objects. They suggest that when a modeller builds a model, she uses elements that do not necessarily come from theories or data, but from "outside". This gives autonomy to the model to function as an investigative instrument. A model can be manipulated by the model user to learn about implications that hold within the model. This manipulability and the model representing its target are, according to Morgan and Morrison, sufficient conditions to learn about the world by using a model. Grüne-Yanoff (2009), in addition, has argued that a representational relation is not a necessary condition for learning about the world. He characterises as minimal models those that do not satisfy the condition of having a link to the real world, either by means of being similar, having a relation of partial resemblance, or adhering to natural laws. He suggests that it is possible to learn from these minimal models because they have the capacity to affect a modellers' confidence in impossibility hypotheses about the world.

In general, the focus on models and the epistemological claims have both been about individual models. So, not only the claims have been defended with examples of a single individual model—e.g. von Thünen's Isolated State model in Mäki, or Schelling's model of segregation in Sugden and Grüne-Yanoff—but the epistemological question itself, namely what models' epistemological import is, has been framed in terms of a single model. To be sure, both aspects are important: the first one because of its economy—it's easier to communicate a claim with a simple, known example—and the second because whether a single model has any epistemic import is in itself an important philosophical question. But this not need preclude questions of how models are actually used, whether they could collectively have other representational relation with the world, or whether they may have another kind or a different degree of epistemic import. However, attention has been given almost exclusively to individual models.

In the next section I discuss two views of models that are an exception to this single-model approach[3]. First I discuss a view that tries to characterise economic modelling as robustness analysis. Then I discuss a view that explicitly endorses model clusters as the unit of analysis and that thereby afford understanding. Even though I share the view that model clusters is the most appropriate unit of analysis, I shall argue that robustness analysis is too limited as a general characterisation of economic theoretical modelling (against the first view) and that the analysis of how models afford understanding by means of analysing them as clusters is too hasty (against the second view).

3. More models

Kuorikoski, Lehtinen, & Marchionni (2010) toy with the idea of models as more than single units; they acknowledge that models are related to other models and suggest that this entails a somewhat added epistemic import in the form of higher confidence in the inferences drawn from models. They argue that a substantial part of modern theoretical economics is devoted to "deriving known results from alternative or sparser modelling assumptions", which they suggest to call a form of robustness analysis: Derivational Robustness Analysis (DRA). The idea here is that a set of

[3] Another exception is Bokulich (2003) who offers a view of 'horizontal model construction' as an alternative to what she calls 'vertical approaches' which take models as mediators between theories and data. Her alternative is useful as a means to facilitate inter-theoretic relations and understanding phenomena at the limit of two theoretical frameworks—e.g. quantum chaos. I don't discuss this work here because her the inter-theoretic relations is not an aspect that is treated in this dissertation.

different means, in this case models, arrives independently at the same result, thereby providing reasons to believe that the result is reliable, despite the falsities and/or idealisations that may have been employed in the models. Although this is certainly not a method of empirical confirmation, it makes the inferences that are drawn from models more reliable. Kuorikoski et al. (2010) distinguish three kinds of assumptions, namely, substantial assumptions, Galilean assumptions, and tractability assumptions. The first set is supposed to track the causal factors that bring about the causal mechanism a modeller is interested in. Galilean assumptions, are used to isolate the mechanism that substantial assumptions capture—they are used to eliminate the disturbing factors that may affect the causal mechanism of interest. Finally, tractability assumptions are introduced to make the model mathematically solvable; they are typically false, but they are expected not to have any influence in the result of the model. Robustness analysis is practised to make sure that the tractability assumptions are not responsible for driving the result. It is thus these assumptions that are adjusted in different models, to verify whether versions of the same model with different tractability assumptions derive the same result. The aim of robustness analysis is thus to "distinguish the real from the illusory; the reliable from the unreliable; the objective from the subjective; the object of focus from artefacts of perspective" (Wimsatt, cited in Kuorikoski et al. p. 542). The aim of their paper is to attempt to demonstrate that even though robustness analysis is not an empirical confirmation procedure, it does have some epistemic import. This, in turn explains, or rather justifies, why a substantial part of theoretical economics is dedicated to this endeavour. If robustness analysis had no epistemic value, they argue, the practice of a substantive part of theoretical economists would have very little justification.

Kuorikoski et al. (2010) start their analysis with a model, described by Paul Krugman in 1991 called the Centre Periphery (CP) model. Their claim is that the activity which followed this model can be regarded as robustness analysis because many other economists started building models that "appear to be checking whether the main conclusions of the CP model remain valid when some of its unrealistic assumptions are altered" (p. 553). In Krugman's model, two opposite forces develop which determine whether the bulk of the economic activity locates in one of two regions, giving rise to a core-periphery pattern. Kuorikoski et al. (2010) identify the model result [RCP] and the causal mechanism [CCP] there isolated. Their claim is that this causal mechanism is the one being captured by other versions of the model, using other tractability assumptions.

> [RCP] Ceteris paribus, spatial agglomeration occurs when economies of scale are high, market power is strong, and transportation costs are low (that is, when the centripetal forces are stronger than the centrifugal forces).

> [CCP] In the presence of immobile and mobile activities, the interaction among economies of scale, monopolistic competition and transportation costs gives rise to centripetal and centrifugal forces.

Among the assumptions that are varied are the form of the utility function, which in Ottaviano, Tabuchi, and Thisse (2002) is substituted for a quadratic form, and the transportation costs, which are changed from the 'iceberg' form, to a linear form. Similarly, other models assume that the manufacturing industry makes use of both skilled and unskilled labour, as opposed to Krugman's original model in which only skilled labour is used. The assumptions that remain constant throughout the set of model variations are those that correspond to what they call the substantial assumptions, namely, the presence of monopolistic competition, economies of scale and transportation costs—the causal mechanism. Despite the variations, most of these models reach the same result, which suggests that it is the substantial assumptions which drive the result and not the tractability assumptions; the results of the models are robust. From this Kuorikoski et al. (2010) conclude that the economists who built these models were engaged in robustness analysis.

Their contribution is important because it draws attention to the analysis of models in clusters instead of single models and thereby to the entertainment of the possibility that more epistemic import resides in clusters of models. The idea of robustness suggests that there is epistemic import in clusters that otherwise would not be observable[4]. If the epistemic import of models is assessed according to the strength of the relation that a single model has with its target alone—by looking for instance, as Mäki suggests, at how much a model resembles its target—then any potential epistemic gain emerging from the group of models as a whole is not considered. In robustness analysis, the increased reliability of the inferences drawn from models is only evident once a cluster is taken into account.

There are, however, at least two difficulties with the defence by Kuorikoski et al. (2010) of robustness analysis in economics. First, as Reiss (2012) has argued, an ideal robustness test would require the permutation of each and every one of the assumptions. This is a very different process than what Kuorikoski et al. (2010) have described and therefore the reliability of the process doesn't seem to be guaranteed. Reiss (2012) also claims that in the few instances in which robustness analysis is done in economics, robustness tests tend to fail. Understanding the implications of these failures seems paramount to understand the practice of robustness analysis in economics.

[4] This is not to say that Kuorikoski et al. (2010) argue against the need for representation in modelling. Their standpoint, regardless of what theory of representation they endorse, is one that looks beyond the individual properties of models, even if their defence of robustness highlights precisely that models are composed of smaller parts. (see section 4 of their (2010)).

Second, in their paper they submit that this is what a significant portion of the economics profession does. In the introduction they say that "[m]odern theoretical economics largely consists of building and examining abstract mathematical models. A substantial portion of this modelling activity is devoted to deriving known results from alternative or sparser modelling assumptions" (p. 541). The idea is thus that economists presume some causal mechanism to be at play in an economic phenomenon, and, because their modelling requires making tractability assumptions, they devote themselves to checking that it is not the tractability assumptions but their causal intuitions that drive the observed result. The implication of characterising "a substantial portion of this modelling activity" as robustness analysis is that the only work that is left to do is to agree on the epistemological import of robustness analysis. That is, if we know that this is what most economists do, as philosophers we just need to understand the epistemology of that practice in order to understand the epistemology of economic modelling.

Perhaps it could be argued that this way of referring to the practice is just an idealisation, that the claim was not intended as an accurate description. This claim is indeed milder; instead of saying that this is what the profession does, the authors would be saying that it could be understood as if it did so. However, it is unclear whether this characterisation would help to convey what the epistemic contribution of economic models is. If our purpose as philosophers is to understand scientific practice and how models, as they are used, are able to convey information about the world, we cannot rely on characterisations that do not relate to science as it is practiced. This criticism was made to philosophy already decades ago, against views such as the Received or the Semantic Views of theories, which were more interested in a logically defensible reconstruction of what theories and models were, than in accounting for the actual practice of science. Things complicate further if on top of providing descriptively inaccurate reconstructions, the reconstruction itself is problematic, as Reiss (2012) argued and I indicated above.

It seems to me that Kuorikoski et al. (2010) characterise economic theoretical modelling as largely DRA as a way to grant further legitimacy to economic modelling and, simultaneously, to DRA as a scientific activity: economists practice DRA because of it's epistemic import and DRA must have a significant epistemic import given that it is what a lot of economists do. These are two separate issues. The value of DRA as a scientific activity is worth understanding from a philosophical perspective regardless of how pervasive it is in a discipline. Likewise, though separately, it is of philosophical interest to understand what exactly economists do when they model and why. Before I go on to show with a case that models are used in other ways than just to practice robustness analysis, let me discuss another view in the literature that endorses the study of models as clusters.

Ylikoski & Aydinonat (2014) is an explicit proposal to analyse clusters of models in order to properly understand their epistemic import. In their paper, Ylikoski & Aydinonat (2014) attempt to elucidate the way in which clusters of theoretical models are used in order to gain understanding about a particular phenomenon. They take the Schelling's model of segregation and the research that followed as their case study, emphasising that the Schelling model is not a single model but a family, developed by Schelling himself and other modellers over the years considering different but related problems.

Ylikoski & Aydinonat (2014) argue that the epistemic import of models can only be fully understood in the context of a cluster of models relevant to the explanatory task at hand. In their view, models are related to each other by genealogical origin and similarity, constituting families of models like Schelling's. Theoretical models like Schelling's are not devised for explaining a specific empirical phenomenon but to merely enquire into theoretical possibilities. This is why, according to them, philosophers should look at clusters of models, instead of at a single model. They recognise that, while it is a good start of Kuorikoski et al. (2010) to look beyond a single model, robustness analysis is just one of the things economists do with models. Economists typically make variations to their models in order to understand what happens when certain conditions change which are deemed central. In general, they suggest, the main focus of research that the Schelling models (and theoretical models more generally) allow for is the exploration of what-if questions. This is not a case of robustness analysis because variations in this case are often of core assumptions, and not of tractability ones.

Furthermore, a theoretical model is so abstract that it can't be said to provide "possible causal scenarios"—historical causal scenarios that tell how an explanandum could have come about. Instead, they provide skeletons of these causal scenarios, or "causal mechanism schemes", as Ylikoski & Aydinonat (2014) call them. These are concerned with what is causally possible, without referring to a phenomenon in particular. This idea of providing just skeletons of the causal scenarios is very similar to what Odenbaugh & Alexandrova (2011) have in mind[5]. Theoretical models like Schelling's offer causal mechanisms that, in principle could bring about the effect that is being investigated, such as segregation. This occurs in a setting in which scientists have an array of competing explanations for a phenomenon, meaning that Schelling's contribution is that it changes the menu of possible explanations that scientists may entertain for the causes of segregation.

[5] Odenbaugh & Alexandrova refer to 'open formulae', an idea that was already developed by Alexandrova (2008).

The epistemic import of models has two sources according to the authors. On the one hand, epistemic import comes from clusters of models, for the increased number of what-if inferences that a scientist can make with them. On the other hand, epistemic import comes from the cluster of competing explanations that emerge from different models. If scientists want to be able to explain a phenomenon, a good strategy is to have a set of possible explanations and piecemeal discard those that the evidence contradicts—a process of eliminative induction they refer to as a weak version of inference to the best explanation. The larger the set, the more difficult it is to find the right explanation, but also the higher the probability that the right explanation is within the set. A more complete set of explanations, in turn, imposes higher demands on the evidence required to exclude possible explanations. The higher the demands on the evidence the more warranted the inference to the best explanation is. In addition, Ylikoski & Aydinonat (2014) contend that models have a formal structure that raises the standards that are imposed when a mechanism scheme can be modelled in a rigorous way.

While this analysis is helpful to understand what a cluster of models may be capable of doing in terms of epistemic import, it only brings us this far. In other words, it doesn't allow us to distinguish between a cluster with epistemic import from one without. Recall that the sources of epistemic import in this analysis are the what-if inferences that a cluster allows—a cluster allows more inferences than a single model, given the model variations—and the competing explanations that emerge from a more complete meta-model of explanations of the phenomenon. From additional what-if inferences, we might be able get a broader insight of a particular phenomenon. Whether the inferences are correct is another matter. With respect to the contribution of the models to the possible causal mechanism that brings a phenomenon about, the gain we get from clusters of models is less clear. It seems that entertaining a new hypothesis from a single model could be enough for this purpose. The point is therefore that the analysis suggests that just any cluster, for being a cluster, yields understanding. Ylikoski & Aydinonat (2014) seem to use their analysis to justify why social scientists consider Schelling's models explanatory despite the many shortcomings and criticisms that have been offered in the literature. They thus take for granted that Schelling's models have epistemic import. They also claim that their analysis and diagnosis is likely to apply to other theoretical models in economics and biology, "that are taken to have explanatory import". How does this analysis help us those models that aren't as explored, or that haven't received as much attention as Schelling has, and therefore we don't know whether they have any epistemic import?

4. More models: vertical and horizontal complementarities

I shall show below that there are important relations additional to the ones established by carrying out robustness analysis that are obscured if the practice of theoretical economics is characterised as such. I use the same example that Kuorikoski et al (2010) use, namely a group of models in the branch of new economic geography (NEG) for two reasons. First and foremost, I want to lend legitimacy to my case: I'm using the same case that Kuorikoski et al. (2010) claim to be a case of robustness analysis in order to show that there is much more at play than mere robustness analysis. Second, to expand the analysis to models that so far have not been as explored as others.

An obvious place to start investigating how a model fits together with other models is to look at the context in which it was developed. This point may sound trivial, but the context of models is seldom considered in the extant analyses of models. Kuorikoski et al. are arguably no exception, since they don't discuss this in their paper. All they say is that NEG "is a recent approach in spatial issues developed within economics, the aim of which is to explain the location of economic activity." (p. 553).

The context of Krugman (1991) is an interesting one. Krugman (1991) is regarded as the precursor of New Economic Geography because there used to be a long schism between economic geographers and economists with respect to explanations of regional development (Martin & Sunley, 1996). For economic geographers it was strange that economists reduced the explanation of why there is trade or what the advantages of it are, to differences in comparative advantage, treating countries as dimensionless, in which only factor endowments matter. For geographers, naturally, the role of geography in determining trade has always been important. Yet, economists had seldom paid attention to these insights, partly because economic geographers use different methods and partly because explaining the core-periphery patterns, which emerged more prominently in the second half of the 20th century, can't be explained by factor endowment theories. Instead, imperfect competition, specifically increasing returns to scale, is necessary. But between the 1940s and the 1970s it was the theory of general competitive equilibrium that dominated economic thinking. The treatment of imperfect competition was not as theoretically and formally developed as its perfectly competitive counterpart (Krugman, 1990, p. 2):

> If we ask why so much of the American economy is concentrated in a few coastal strips, we are immediately driven to speak about economies of scale and externalities. Yet economies of scale internal to firms imply imperfect competition, which until recently was regarded as too difficult to model rigorously, while purely technological external economies seem both implausible and too elusive to have useful empirical content.

Krugman (1991) thus marks the point at which economists 'discovered geography' and attempted to explain the location of economic activity taking into account the role of geography. His model attempts to show that a simple economic setting, which features monopolistic competition—thereby the possibility of firms of facing increasing returns—and transportation costs, is capable of giving rise to a core-periphery pattern of economic activity.

> The purpose of this paper is to suggest that application of models and techniques derived from theoretical industrial organisation now allow a reconsideration of economic geography; that it is now time to attempt to incorporate the insights of the long but informal tradition in this area into formal models. In order to make the point, the paper develops a simple illustrative model designed to shed light on one of the key questions of location: why and when does manufacturing become concentrated in a few regions, leaving others relatively undeveloped? (p.3)

This context is important, not only to understand Krugman's contribution, but also the contributions that followed. For expository purposes, I shall only focus on one of the models that Kuorikoski et al. (2010) use for their argument that economists carry out robustness analysis with economic models. Such a limited focus is enough to illustrate my point that there are other kinds of relations than merely those established for practising robustness analysis. As mentioned above, Ottaviano, Tabuchi, and Thisse (2002) arrive at the same results as Krugman (1991) while they substitute the CES utility function for a quadratic form and the 'iceberg' transportation costs for a linear form. According to Ottaviano et al. (2002), their model has two purposes. The first purpose is to introduce a model that exhibits the same features as those of the rest of the literature, including Krugman (1991), while allowing for the derivation of analytical results using simple algebra. This first part of their purpose could very well be interpreted as an attempt to do a robustness test: to replicate the results that Krugman obtained, but in this case deriving the results analytically, instead of numerically, as other contributions had done.

However, the reason for why it was desirable that the model exhibit the same features as the rest of the literature was because it was an attempt to solve some technical limitations that were present in the literature at the moment. Until then, NEG models had relied on a set of strategies that were considered problematic. One problem was that most of the models developed until then required numerical computations in order to be solved. This was seen as a disadvantage and therefore there was a strong interest in being able to derive the model analytically. Another problem was that the literature had relied on the Dixit-Stiglitz model of monopolistic competition. The CES utility function was part of this strategy to express the consumers' love for variety, as was required. Ottaviano et al. (2002) used a quadratic function instead that enabled them to express this love for variety and that allowed them to adopt a broader concept of equilibrium than what the Dixit-

Stiglitz strategy allowed. Yet another difficulty was that 'iceberg' transportation costs included transportation costs in a rather rudimentary way: the assumption of 'iceberg' costs is that part of the product melts on its way to its destination. As a result, equilibrium prices turn out to be independent of the spatial distribution of firms and consumers. This result conflicts with spatial pricing theory, which says that demand elasticity varies with distance and prices change with the level of demand and the intensity of competition (Ottaviano et al., 2002).

The second purpose of the model is to show that their new specification allows them to investigate further aspects of the same phenomenon than what the extant models had so far investigated, namely, to make a welfare analysis of the agglomeration process; to analyse the role of history and expectations in the emergence of economic clusters; and to analyse the impact of urban costs on the spatial distribution of economic activities (Ottaviano et al., 2002). These aims are more in line with the claims of Ylikoski & Aydinonat (2014) and Odenbaugh and Alexandrova's view that robustness analysis is a tool of discovery. In a reply to Kuorikoski et al. (2010), they argue that robustness analysis should be seen as a tool of discovery and not as a tool of confirmation, as Kuorikoski et al., (2010) seem to suggest. Odenbaugh & Alexandrova (2011) argue that robust theorems do not confirm the hypotheses that later appear in causal explanations of phenomena. Instead, hinting to Paul Humphreys and William Wimsatt's "templates", they argue that models are useful for building good explanations. More specifically, they suggest that robust theorems "provide us with open formulae that can be used to build hypotheses about mechanisms, and robustness analysis is a way of chiselling out these open formulae" (2011, p. 769). In the case of Ottaviano et al. (2002), obtaining a robust result indicated that their results of the agglomeration process were consistent with the rest of the literature. This consistency allowed them to make additional inferences about another aspect of the phenomenon under investigation, namely a welfare analysis of the agglomeration process. They could thus be thought of as building new hypotheses. In that particular model, for example, the welfare analysis shows that there is a range of trade costs for which the equilibrium doesn't align with the social optimum, giving room to potential regional policy interventions. Similar kind of inferences with respect to the role of path-dependence and agent expectations and the impact of urban costs on spatial distribution of economic activities were also possible with this model. These were additional to what was possible to investigate with previous models.

Ottaviano et al. (2002) were not engaged in a robustness analysis exercise and therefore shouldn't be characterised as such. I'd like to suggest instead that what seems to be going on here is an attempt to provide some sort of horizontal complementarity to the models. A phenomenon, although regarded as a phenomenon—in this particular case, a pattern of agglomeration of economic activity—can usually be studied from different perspectives and explored with extensions, complications or implications of the phenomenon under study. So models can be developed such that, strictly speaking, they are about the same phenomenon, but in a somehow varied form. They are thus complementary to other models in a horizontal way, in the sense that the cluster of models put together provides an analysis that is broader, more comprehensive. Each individual model treats a very specific aspect of a phenomenon. In the case of Ottaviano et al. (2002) they come up with the inclusion of a welfare analysis and the impact of expectations on the agglomeration of economic activity. A more comprehensive analysis is thus available if these contributions are considered.

Moreover, while the new model allowed to make new what-if inferences, such as "what are the welfare consequences for the population if economic agglomeration takes place in the way described by NEG models"—and thus lending support to Ylikoski & Aydinonat's analysis—the interest was not in the causal mechanism behind core-periphery patterns. In fact, according to Ottaviano (2003), the mechanism that brings about core-periphery patterns has been well known for more than a century. After describing the consolidation of the New Economic Geography (NEG) begun by Krugman, Ottaviano (2003, p. 667) states: "This mechanism is not new. For example, it is carefully described by both Marshall, 1890, and Ohlin, 1933". The interest was instead in being able to generate the known pattern bringing together two kinds of models and techniques that until then had been separate. Ottaviano continues: "The crucial contribution of NEG is that it is translated into a general equilibrium model with solid microeconomic foundations" (p. 667). The point here was therefore not to understand the world. The question was not, "can I obtain the same model result with slightly different assumptions?" nor "what happens in my model if this or that assumption is incorporated?" There was no question about what is the mechanism that brings about this pattern of economic activity. Instead, the point was to be able to improve economic theory: to be able to generate a known result with the tools that were considered acceptable to the discipline.

Neary (2001) and Ottaviano (2003) recognise that a pitfall of the approach initiated by Krugman was that the results were obtained using very specific functional forms and numerical methods. It is for this reason that Ottaviano et al. (2002) modify some of the original assumptions. Ideally, the models should be as realistic[6] as possible, and in this emergent literature this was hardly possible. For instance, Krugman, in an assessment of the literature states:

[6] 'Realistic' is likely to be controversial, considering the extent to which economists have been lambasted precisely for the lack of realisticness of their models. I shall not engage in such a debate here nor question

> Because the main obstacle that economists have traditionally faced when trying to confront issues involving increasing returns is one of tractability, overcoming that obstacle depends crucially on technical tricks: on strategic assumptions that may be unrealistic but make a model easier to build, on clever new ways of solving models that might otherwise seem too complex to deal with. To date, the new economic geography has depended heavily on the tricks summarized in Fujita, Krugman and Venables (2002) with the slogan "Dixit-Stiglitz, icebergs, evolution, and the computer" (Krugman, 1998, p. 164).

More generally, the observation just made raises two related points. First, some economic models are not explanatory. We just learnt that at least some of the models in NEG were not in this business. This is not a tragedy. I also don't think this turns economists into idiots savants who know nothing else than to show off with their mathematical trickery. Economists here seemed to be compromised with a long-term project, in which being able to model the phenomenon in question with the tools and methods that were known and acceptable to them was just a first step. The Kuhnian idea of normal science that puzzle-solving is about scientists testing their own capacity to solve puzzles while taking for granted the theory was here turned on its head. Economists were in fact testing the theory. Krugman made a first attempt, with many technical concessions, that were later modified to bring more realisticness and acceptability to the theory being developed. Once there was convergence and a basic framework was accepted, the next step was, according to Ottaviano (2003), to come up with a coherent necessary framework that connects the theory with the policy implications of NEG. They attempted to take the models literally at their policy implications, in order to start contributing to building the coherent 'organisational framework'.

> The point of this paper is rather the opposite: what is needed at this stage is precisely to take the models literally and ask what their exact policy implications are. This is a necessary preliminary step to provide a model-grounded benchmark for more realistic extensions of NEG insights to the policy domain (Ottaviano, 2003, p. 666).

Above I suggested that the type of relation that arises between two or more models which gives rise to a more comprehensive understanding of a phenomenon may be called a relation of horizontal complementarity. Analogously, digging deeper into a phenomenon, analysing the conditions under which certain result obtains, or attempting a more realistic model for a particular phenomenon, is something that makes two or more models vertically complementary. Whereas

their understanding of what realisticness is. The point is simply that certain assumptions in the first NEG literature were not satisfactory for economists for not adhering to certain standards. (As to the idiosyncratic term 'realisticness', this has entered the debate due to Uskali Mäki, who wanted to distinguish it from 'realism' as a philosophical stance.)

horizontal complementarity aims at a broader comprehensive understanding of the phenomenon, vertical complementarity is more related to the method and the way a phenomenon is modelled.

This brings me to the second point that arises, which is that in order to understand the epistemic import of models, we still need to explore a wider array of economic models. This could help us find out how the distinction between vertical and horizontal complementarity helps to make assessments of the epistemic import of such models. It would also enable us to assess the epistemic import of clusters of models compared to that of individual models and to see whether it is possible to assess the individual contribution of models within a cluster. Ylikoski & Aydinonat do not distinguish individual models from clusters of models. So it is unclear whether according to their analysis it is possible to assess the epistemic contribution of a cluster and, independently, the epistemic contribution of a model within a cluster. This is to me a legitimate question to ask. Ylikoski & Aydinonat seem to defend the stronger claim that only models as clusters have epistemic import. In this regard, I think we first we have to learn more about more models. My much weaker claim, then, is that more models have to be analysed in clusters in order to properly understand their epistemic contribution.

5. Many models and understanding

Given the analysis I have made above, I'd like to suggest that a way to look into the epistemic contribution of economic models is to look into a debate that has more or less recently taken some force in the philosophy of science, namely the debate on understanding as a notion independent from explanation. Ylikoski & Aydinonat (2014) are contributions to this debate. While it is obvious that to understand phenomena is what we're after—rather than simply produce neat, scientific explanations for the sake of it—the notion of understanding had been largely dismissed in the philosophy of science because of its evident subjective aspect. Some philosophers have gotten interest in this divide and have brought enlightening perspectives to the debate that are worth exploring, particularly if one favours a descriptively accurate philosophy of science. In this section I discuss Catherine Elgin's defence of a broad conception of understanding in which both factive and non-factive notions are simply part of a continuum of what understanding is. It seems to me that such a view of understanding is helpful in the analysis that philosophers of science are interested in when it comes to the epistemic contribution of economic (and other social sciences) models.

Elgin (2007) has attempted to develop a more comprehensive conception of understanding than one that regards it only as factive. That is, one that can only obtain if based on facts, or truths. She thinks that this view is too restrictive for scientific practice, in the sense that it states that we only have understanding when the propositions that express it are true. This restrictive notion does not reflect current scientific practice because in science we ascribe understanding to certain enquiries that are not necessarily true. Furthermore, idealisations—she calls them felicitous falsehoods—are ubiquitous in science and therefore, strictly speaking, scientific understanding of phenomena is not obtained from true propositions. Her general motivation is that, if epistemology concerns only with a factive conception of understanding, then epistemology cannot accommodate scientific understanding—at least in its entirety—which, according to her, is necessary, considering that science is one of humanity's greatest cognitive achievements.

Elgin defends a conception of understanding as "a grasp of a comprehensive body of information that is grounded in fact, is duly responsive to evidence, and enables non-trivial inference, argument, and perhaps action regarding that subject the information pertains to" (2007, p. 39). I'll highlight two points about this conception. First, the unit of understanding, or the "primary bearer of understanding's epistemic entitlement" is a body of information and not individual propositions. It is thus not about 'merely' knowing a single proposition or a bunch of them. It involves being able to reason with them, making inferences, or, more generally, using this body of information. It also involves that not all the propositions that comprise the body of information need to be true. One can understand a subject, say the last 50 years of armed conflict in Colombia, while entertaining some propositions about it that are strictly false.

Second, her conception of understanding admits of degrees. Elgin argues that there are three dimensions that are shared with a factive account along which degrees are admitted, namely breath, depth and significance. Breadth states that a person may have a greater degree of understanding of say, the armed conflict in Colombia than another if this body of information is embedded in a greater context of Colombian history—e.g. if the former knows about the colonisation processes in the different regions that took place after the Spanish colonisation and the implications this had for how land was distributed[7]. Depth states that the propositions that comprise the body are more tightly connected, that is, the body has more propositions or there are more relations among propositions. In the case of the armed conflict in Colombia this could amount to propositions about time periods of the colonisation processes or whether certain land was distributed according to institutions imposed by the Spaniards such as the *encomienda*. Finally, significance has to do with whether a particular proposition or subset of propositions is given more importance than others given their significance in the matter—e.g. understanding the significance of "Operación

[7] As with many civil wars, the Colombian war has roots in land property and distribution.

Marquetalia"—a Military Operation that caused a peasant uprising in 1964, which is marked as the origin of the Revolutionary Armed Forces of Colombia (FARC), one of the guerrillas involved in the current armed conflict.

Besides these three dimensions, there is another one that a factive account of understanding cannot accommodate, according to Elgin (2007). This is one in which a body of information from which understanding is derived is strictly false—a person has false beliefs—but these are somehow closer to beliefs that are closer to the truth. Here Elgin (2007) gives the example of a second grader who believes that humans are descendants of apes; a strictly-speaking false belief, but that, as an understanding of human evolution is closer to the truth, and thus cognitively better, than someone who denies evolution altogether. Of course, the crux of the matter is how to distinguish those false-but-cognitively-better beliefs from others that aren't cognitively or epistemically better in any way. Elgin (2007) doesn't go as far, but her general point is still important and relevant: early steps in a sequence from false beliefs to beliefs that may be true should fall within the ambit of epistemology, just because they are often cognitively valuable. An analogous case in economics is probably Jevons's rather crazy belief that business cycles were caused by cycles associated to sunspots, but closer to the truth than the belief entertained by scholars before him that cycles weren't actually cycles but caused by completely exceptional events like war (Morgan, 1990, Chapter 1). Although misguided in considering sunspots as the cause of business cycles, Jevons was closer to the truth in his understanding of fluctuations in economic activity than scholars before him.

How does all this relate to the point of this paper? How does Elgin's treatment relate to analysing models as clusters rather than as single units? Insofar as understanding admits of degrees and stretches from a continuum of say, "very little but in the right direction" to greater extents of understanding, to judge whether a model provides understanding, it is more helpful to judge it in relative terms. A model analysed in isolation can't tell us much more about a phenomenon than the inferences we're able to make in relation to how well it represents its target.. Instead, if a model and its context are considered, and the model under consideration is judged relative to previous or similar models, a more comprehensive notion of understanding is likely to emerge. We are able to tell how many more inferences we can make in comparison to our understanding without the model. In the NEG case, Krugman's model and its features can be thought of as being benchmarks that allow other models to expand horizontally or vertically. The former allow for a more comprehensive understanding of the phenomenon, whereas the latter allow for more comprehensive understanding or refinement of the tools.

Moreover, it is possible to grasp not only the epistemic contribution of the cluster of models, say how much we understand a phenomenon in general, but also how each individual model contributes to that general understanding. So, with an analysis of clusters of models, a question like "What is the epistemic contribution of this particular model?" can be answered. Such question is likely more relevant for the practice than the question of what the epistemic contribution of models in general is.

Naturally, it all hinges on how the understanding of the model is related or can be extrapolated to the world, which is what we're ultimately interested in. This is obviously an empirical question that can be answered only on a case by case basis. But understanding the relations that exist between models, which includes those with their empirical counterparts, is certainly a way to assess the relevance of theoretical models. This is, in fact, explicit in Ottaviano's claim above that once a certain theoretical framework was more or less consolidated, it was time for NEG to start consolidating a framework useful for the empirical questions that include policy. These are most likely the ones that warrant confidence in determining whether understanding derived from a theoretical model is indeed understanding, grounded on fact, and duly responsive to evidence.

Conclusion

In this paper I argue that the extant accounts of economic theoretical models have focussed mostly on the analysis of what a single model can contribute epistemically, ignoring relations between models that may prove useful in such analysis. Two characterisations of models in the literature are exceptions, recognising that models are neither built nor used in isolation and thus analysing the epistemic import of clusters. The first one, by Kuorikoski et al. (2010, 2012) characterises modelling as derivational robustness analysis. The idea here is that such a practice makes more reliable the inferences that are made about a causal mechanism by arriving at the same result with different model specifications. Using their own example, namely a cluster of models in New Geographical Economics, I show that models in such a cluster had different purposes than to just confirm the robustness of a causal mechanism. The second view is one developed by Ylikoski & Aydinonat (2014), in which they defend the view that models as clusters alone have epistemic import, and specifically defend the epistemic import of Schelling's models of segregation. Here I argue that, by their account, there is no way to identify an explanatory cluster from a non-explanatory one. It all seems to rest on choosing the 'right' cluster. There's also not a way to tell whether individual models contribute within the cluster. I thus suggest that until we have investigated more economic models, we're unlikely to find out what their epistemic import is. Finally, I use an account of understanding developed in the literature on epistemology to briefly suggest how the analysis of models as clusters may elucidate the actual contribution of models. This is offered merely as a suggestion and does not pretend to close the debate. To the contrary, the primary aim is to open it by offering an alternative avenue worth exploring.

References

Alexandrova, A. (2008). Making Models Count. *Philosophy of Science*, *75*(3), 383–404.

Bokulich, A. (2003). Horizontal Models: From Bakers to Cats. *Philosophy of Science*, *70*(3), 609–627. https://doi.org/10.1086/376927

Cartwright, N. (1983). *How the Laws of Physics Lie* (First Edition). Oxford University Press, USA.

Elgin, C. (2007). Understanding and the facts. *Philosophical Studies*, *132*(1), 33–42.

Giere, R. N. (1990). *Explaining Science: A Cognitive Approach*. University Of Chicago Press.

Giere, R. N. (2004). How Models Are Used to Represent Reality. *Philosophy of Science*, *71*(5), 742–752.

Giere, R. N. (2006). *Scientific perspectivism*. Chicago, Ill.; Bristol: University of Chicago Press.

Grüne-Yanoff, T. (2009). Learning from minimal economic models. *Erkenntnis*, *70*(1), 81–99.

Hands, D. W. (2015). Orthodox and heterodox economics in recent economic methodology. *Erasmus Journal for Philosophy and Economics*, *8*(1), 61–81.

Krugman, P. (1990). *Increasing Returns and Economic Geography* (Working Paper No. 3275). National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w3275

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, *99*(3), 483–499.

Krugman, P. (1998). Space: the final frontier. *The Journal of Economic Perspectives*, *12*(2), 161–174.

Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2010). Economic Modelling as Robustness Analysis. *The British Journal for the Philosophy of Science*, *61*(3), 541–567.

Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2012). Robustness analysis disclaimer: please read the manual before use! *Biology & Philosophy*, *27*(6), 891–902.

Mäki, U. (2009). MISSing the World. Models as Isolations and Credible Surrogate Systems. *Erkenntnis*, *70*(1), 29–43. https://doi.org/10.1007/s10670-008-9135-9

Martin, R., & Sunley, P. (1996). Paul Krugman's Geographical Economics and Its Implications for Regional Development Theory: A Critical Assessment. *Economic Geography*, *72*(3), 259. https://doi.org/10.2307/144401

Morgan, M. S. (1990). *The history of econometric ideas*. Cambridge [England]; New York: Cambridge University Press.

Morgan, M. S., & Morrison, M. (1999). *Models as mediators: perspectives on natural and social sciences*. Cambridge; New York: Cambridge University Press.

Morrison, M. (1999). Models as autonomous agents. In M. S. Morgan & M. Morrison (Eds.), *Models as Mediators: Perspectives on Natural and Social Science* (pp. 38–65). Cambridge: Cambridge University Press. Retrieved from https://www.cambridge.org/core/books/models-as-mediators/models-as-autonomous-agents/08BB4E28A5914BF3E29ECD8397E0E4FD

Morrison, M. (2015). *Reconstructing reality: models, mathematics, and simulations*. Oxford ; New York: Oxford University Press.

Morrison, M., & Morgan, M. S. (1999). Models as mediating instruments. In M. S. Morgan & M. Morrison (Eds.), *Models as Mediators: Perspectives on Natural and Social Science* (pp. 10–37). Cambridge: Cambridge University Press. Retrieved from https://www.cambridge.org/core/books/models-as-mediators/models-as-mediating-instruments/10737C6DD4744A65E4B5B89B3D489B21

Neary, J. P. (2001). Of Hype and Hyperbolas: Introducing the New Economic Geography. *Journal of Economic Literature*, *39*(2), 536–561.

Odenbaugh, J., & Alexandrova, A. (2011). Buyer beware: robustness analyses in economics and biology. *Biology & Philosophy*, *26*(5), 757–771. doi.org/10.1007/s10539-011-9278-y

Ottaviano, G. (2003). Regional Policy in the Global Economy: Insights from New Economic Geography. *Regional Studies*, *37*(6–7), 665–673. https://doi.org/10.1080/0034340032000108750

Ottaviano, G., Tabuchi, T., & Thisse, J.-F. (2002). Agglomeration and trade revisited. *International Economic Review*, *43*, 409–436.

Reiss, J. (2012). The explanation paradox. *Journal of Economic Methodology*, *19*(1), 43–62.

Ross, D. (2014a). *Philosophy of economics*. Palgrave Macmillan.

Ross, D. (2014b). Philosophy of Macroeconomics and Economic Policy. Retrieved from http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935314.001.0001/oxfordhb-9780199935314-e-47

Stiglitz, J. E. (2010). *Freefall: America, free markets, and the sinking of the world economy*. New York: W.W. Norton & Co. Retrieved from https://www.overdrive.com/search?q=BF982A35-BCDF-4950-9430-70C807B14F6B

Stiglitz, J. E. (2011). Rethinking Macroeconomics: What Went Wrong and How to Fix It. *Global Policy*, *2*(2), 165–175. https://doi.org/10.1111/j.1758-5899.2011.00095.x

Suárez, M. (2004). An Inferential Conception of Scientific Representation. *Philosophy of Science*, *71*(5), 767–779.

Sugden, R. (2000). Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology*, *7*(1), 1–31. https://doi.org/10.1080/135017800362220

Sugden, R. (2011). Explanations in search of observations. *Biology and Philosophy*, *26*(5), 717–736.

Ylikoski, P., & Aydinonat, N. E. (2014). Understanding with theoretical models. *Journal of Economic Methodology*, *21*(1), 19–36. https://doi.org/10.1080/1350178X.2014.886470

3

Epistemic Contributions of Models: Conditions for Propositional Learning

Epistemic Contributions of Models: Conditions for Propositional Learning[1]

Introduction

Models are powerful tools that can make us learn. Few contemporary observers of science doubt that and economists agree; the highest honours of their discipline go to the most influential model builders. Among a long list of modellers who are Nobel laureates, we count Peter A. Diamond, Dale T. Mortensen and Christopher A. Pissarides, who were awarded the prize in 2010 as a recognition of their work in developing a model of the labour market—the DMP model.[2]

While researchers agree that models make significant epistemic contributions in science, judging whether a specific model made us learn is no easy matter. The recent literature on models, though rich in insights, is not as helpful as one might hope in dealing with this issue. Much energy has been spent arguing that models *can* be highly useful and there are today lists of what they *can* do (e.g., Morgan and Knuuttila 2012, p. 73). Unfortunately, these lists give us little handle when it comes to analysing claims about whether we have learnt from a specific model and in what sense.

The main goal of this article is to help with such analysis. In particular, we highlight three epistemic roles that models can play in our learning about the world. In addition, we provide conditions that are sufficient for each role to be actually played by a given model. A secondary contribution of our paper is to connect more tightly the discussion on 'learning from models' to general epistemology. We connect the two by using the traditional account of propositional knowledge to analyse how models can help us learn about the world. Our explicit epistemological perspective allows us to structure the relationship among our three epistemic roles and to articulate how learning from models fits into a more general picture of knowledge acquisition.[3]

The scope of our project must be properly delimited. We do not claim that the three roles identified are the only ones models might play. We are also not the first to try to supply conditions for learning from a model. For instance, the proposals by Alexandrova (2008), that models supply open formulae and Grüne-Yanoff (2009), that they falsify impossibility hypotheses, can be understood in terms of attempting to provide sufficient conditions. Yet, the present article goes

[1] With François Claveau. Both authors have contributed evenly to the chapter.
[2] DMP stands for the initials of the three modellers.
[3] By using an epistemological concept of learning, we are not suggesting that other perspectives—e.g., cognitive—are not fruitful or important.

beyond these contributions by identifying conditions for a number of epistemic roles and by articulating these conditions with the help of the traditional account of propositional knowledge.

Our general account of learning is presented in the next section. We then discuss our three epistemic roles. Finally, we present a case study of the DMP model, which is meant to illustrate how our conditions can help in structuring a fruitful debate over the epistemic contributions of a given model.

1. On learning

To be able to characterise precisely potential epistemic contributions of models, we need to be clear on what we take learning to be. For the purpose of this paper, we propose to take learning to be the process of "coming to know" (Audi 2011, p. 162), and to rely on the traditional account of knowledge as true justified belief. According to this account, which is about knowledge of propositions, an agent knows a proposition if and only if three conditions hold: (i) the proposition is true, (ii) the agent believes the proposition, and (iii) the agent has an appropriate justification for this belief.[4]

Propositional knowledge is only one type of knowledge, which excludes other types of knowledge such as knowledge-how (see Fantl 2012). This restricted focus of ours might be a significant omission when we think about models since it is very likely that models contribute to know-how besides contributing to know-that (i.e. propositional knowledge). For instance, through exercising with models, one might develop abilities to better react to various real-world happenings much in the same way an aircraft pilot develops intuitions and reflexes in a flight simulator. Although we recognise that a significant amount of learning can be related to knowledge-how, we think that providing explicit conditions for learning in terms of propositional knowledge is already a significant contribution, to which we limit ourselves here.

The traditional account of knowledge as true justified belief (KATJB) is not without its faults. Since Edmund Gettier's famous article (Gettier 1963), it is largely granted that the three conditions stated above, though apparently necessary, are not fully sufficient for knowing a proposition. Once the general structure of Gettier's counterexamples is understood, it is easy to produce thought

[4] Some terms in this definition of knowledge—foremost 'truth' and 'justification'—could be given a variety of interpretations. We do not need to commit to specific interpretations for the purpose of this paper. For the major contending theories of truth and justification see entries in the Stanford Encyclopedia of Philosophy (e.g., Glanzberg 2013; Ichikawa and Steup 2012) and readers like Bernecker and Dretske (2000) and Bernecker and Pritchard (2011).

experiments in which a true justified belief can intuitively not count as knowledge (Zagzebski 1994). Although this implies that there could be cases that our account would regard as involving learning—acquiring knowledge—when in fact knowledge is not acquired, Gettier cases are scarce. By their very nature, Gettier cases can amount to only a small proportion of the elements in the set of all true justified beliefs (Hetherington 2011, p. 121). Since our goal is not to provide a definition of knowledge, an account that reliably, but fallibly, distinguishes between knowledge and non-knowledge is satisfactory.

There are good reasons to stick to KATJB in this article despite its drawback. First, the account focusses on what epistemologists still believe to be the concepts most tightly connected to propositional knowledge: truth, belief and justification. In fact, most of the recent accounts of propositional knowledge try to modify KATJB just enough to avoid Gettier cases (Hetherington 2011; Ichikawa and Steup 2012). Second, KATJB is simpler than most other accounts, since others include other elements such as infallibility or the elimination of luck as attempts to shield against Gettier cases. And third, none of the alternative accounts are free of problems; they all seem to fail to provide necessary and sufficient conditions for knowledge. There is simply no account that perfectly distinguishes knowing from not knowing.

What is clear, however, is that knowing is a state of an agent: at a certain point in time, an agent knows or not a proposition. By contrast, learning is a process of passing from a state of not knowing a proposition to the state of knowing it—it is coming to know. Thus, we characterise a process as learning if an agent starts the process lacking belief or justification (or both) in a true proposition, and ends it with both belief and justification in the proposition.[5] Before turning to models, we want to discuss the possible instances of learning implicit in the previous statement.

Learning can involve the process of 'coming to believe a true proposition'. We want to distinguish between two possible ways in which this process of belief generation occurs. First, the agent can change her mind—change her doxastic attitude—with respect to this proposition. In this case, the agent starts the process either disbelieving the proposition or withholding judgment with respect

[5] In our account, the process of learning ends with knowledge. Some might want to work with a more permissive account for which learning is 'coming closer to know' instead of 'coming to know'. The concept of 'closeness' on which this alternative account relies is however difficult to pin down. It leads to difficult questions: Are we learning if we come to be justified in believing a false proposition? What if we come to believe a true proposition for entirely crazy reasons? Though we do not try to develop such a weaker account in this article, it might be possible to do so successfully; we therefore present our account as supplying only jointly sufficient (but perhaps not jointly necessary) conditions for learning.

to it and finishes the process believing it.[6] Second, the agent might start the process without even having a doxastic attitude for the proposition. Indeed, an agent holds, at any point in time, doxastic attitudes for only a tiny fraction of all the possible propositions she could envisage. In the 18th century, no one had a doxastic attitude for the value of Planck's constant. The process of coming to believe a true proposition can thus involve forming a doxastic attitude rather than simply changing it.

In addition to, or instead of, coming to believe, learning can occur through the process of 'coming to be *justified* to believe a proposition'. The concept of justification relies on the distinction between adequate and inadequate evidence: an agent comes to be justified to hold a certain doxastic attitude if and only if her evidence for this attitude crosses the threshold for adequacy. Evidence for a proposition suggests that the proposition is true; if the evidence is adequate, truth is indicated reliably. But even adequate evidence is fallible; truth and justification should not be conflated.

It is helpful to think about epistemic justification in terms of a network of doxastic attitudes for propositions connected to each other. Propositions can stand in an evidential relation to each other—believing one proposition warrants, to some degree, believing another. Since Paula believes that 'the clock indicates 14.00 local time', she feels confident that 'it is not night'. If we locate the doxastic attitude for the proposition 'it is not night' at the centre of our network, Paula's belief that 'the clock indicates 14.00' will be connected to this central node, together with many other doxastic attitudes.

The set of doxastic attitudes having an evidential relation with the doxastic attitude at the centre of the network constitutes the evidence for this attitude. This evidence will be adequate or inadequate depending on properties of the network such as its evidential density. This property summarises the number of doxastic attitudes connected to the central attitude. The density of Paula's network centred at the belief in the proposition 'it is not night' would be higher if, on top of believing 'the clock indicates 14.00 local time', she also had a doxastic attitude for 'the sun is shining through the window'.

To sum up, we take learning to be about 'coming to hold a justified belief for a true proposition'. Learning y means that, at the start of the process, the agent does not *know* the proposition. Depending on what the agent is missing—belief or justification—learning involves either 'coming to believe' or 'coming to be justified in believing' (or both). In any case, the process ends with the

[6] Here we conceptualise doxastic attitudes in a trichotomous framework: disbelief, withhold judgment and belief, but it could also be rephrased in terms of, say, 'degree of belief'.

three conditions for knowledge being met: truth, belief and justification. In the rest of this paper, this account will be used to answer the following question: How can models make us learn?

2. Learning with models

Something that can be easily granted for most models is that by constructing and manipulating a model, an agent learns propositions *about* the model that she works with. We call these 'model propositions'. Morgan (2012) refers to this learning as 'enquiring *into* the model'.

Two conditions must hold for it to be the case that an agent has learnt *with* the model *about* this same model. First, the agent must be in the proper end state: there must be some true model proposition that the agent justifiably believes. In other words, the agent must, at the end point, *know* this proposition. Second, the agent's knowledge must have been acquired thanks to the modelling exercise. In particular, there are two relevant counterfactual dependencies: either the agent would not have *believed* the proposition had it not been for the activity of generating the model, or she would not have been *justified* in believing the proposition (or both).

A model must thus make the agent believe the proposition, or make the agent be justified in believing the proposition, or both. How does a model make an agent believe a proposition? Modelling arguably generates beliefs in the two ways discussed in the previous section. Toying with a model makes the agent *form* doxastic attitudes for many model propositions that were not even on her radar before. That is, prior to the modelling exercise, the agent plausibly possessed doxastic attitudes just for a few model propositions—based on intuitions or theoretical considerations. Likewise, modelling might also lead the agent to *revise* previously-held doxastic attitudes with respect to some model propositions.

Regarding justification, the manipulation of a model typically provides justification for its model propositions. For instance, the fact that Arrow and Debreu (1954) *derived* the existence of an equilibrium in their general equilibrium model looks like adequate evidence for their belief that 'an equilibrium exists in this model'. It is also plausible to say that they did not have adequate evidence for their belief in this proposition prior to their derivation since the effort put in the derivation would make little sense otherwise. Note that this derivation and the belief that Arrow and Debreu are competent modellers are solid grounds for observers like us to grant one aspect of the end-state condition: this model proposition must be true. In short, in cases like the general equilibrium model of Arrow and Debreu, it seems implausible to deny that a model contributes to learning about itself in that agents come to believe and come to be justified in believing true propositions about it.

Granting that models make us learn about themselves is obviously not granting much. Now we turn to how the agent's learning about a model can be a stepping-stone to learn about other target systems, especially phenomena in the real world.

2.1. Evidential role: the model contributing to justification

We start with what is perhaps the most contentious—and most discussed—potential epistemic contribution of models. Roughly, the idea of the evidential role is that *model* propositions, by contributing to justify *real-world* propositions, can contribute to learning about the world.

To begin, let us take the following real-world proposition: 'Low employment protection is a cause of the low unemployment rate in the USA'. At a certain point in time, an agent might lack justification—might have inadequate evidence—to believe this proposition and consequently develop strategies to increase the strength of her evidential network. The agent might, for example, investigate whether countries with more employment protection typically have higher unemployment rates. By doing similar empirical research, she will increase the chance of being justified in holding her doxastic attitude for the initial proposition.

The question at issue when it comes to discussing the plausibility of an evidential role for models is whether *model* propositions can have the same function of strengthening one's evidential network for a *real-world* proposition. There are three conditions that must hold jointly for a model to play an evidential role. First, an end-state condition: there is a true real-world proposition *p* that the agent justifiably believes. Second, there is at least one model proposition *q* that is part of the agent's evidential network for *p*. Finally, if the agent did not have the doxastic attitude she has for *q*, she would not be justified in believing *q*. In other words, at least one model proposition makes a difference to knowledge: given the context, the doxastic attitude for model proposition *q* is necessary for justification.[7] This condition is meant to rule out situations in which the evidence is already adequate to justify the belief in *p*. In such situations, even if it were granted that a model

[7] In contrast to the counterfactual dependence involved in learning about a model (see above) and to most of the ones discussed for the other roles below, the counterfactual dependence here is not causal, but rather "logical" or "analytical" (Kim 1973, p. 570). At the end state, the doxastic attitude for the model proposition is necessary for the evidence to believe p to pass the threshold for adequacy. When counterfactual dependence is causal, assessing it requires going back in the causal process resulting in knowledge to judge whether this process (and its end state) have been causally dependent on the model.

proposition is part of the evidence for *p*—the second condition—there would not be an epistemic contribution since the model proposition would be redundant for justification.

Whether and how often the second condition holds for economic models is the most contentious issue in the discussion of the evidential role in the literature. An influential view is that some propositions known to be true of the model are evidence for real-world propositions if, and only if, the model appropriately isolates the key features of the real-world system (e.g., Cartwright 1989; Mäki 2009). This view, however, leads some scholars to a sceptical conclusion (e.g. Reiss 2008; Alexandrova 2008): it seems that many specific assumptions are built in economic models that are doing more than cleanly isolating the 'key features'.

Nevertheless, it can be argued that there are ways to avoid the conclusion that propositions about economic models are never part of the evidential network for real-world propositions. To start with, a model can indicate the falsity of particular types of real-world hypotheses—e.g., claims that something can never be the case—even though the model does not cleanly isolate key features of the real-world (Grüne-Yanoff 2009).

More generally, asking for a clean isolation of the target's key features appears too severe when we think of models as experiments in analogy to the experiments that we perform on one part of the world in order to learn about another part of it. For instance, we routinely test drugs on mice to assess their potential toxicity for humans. We run these experiments because we think that their results are evidentially relevant to our doxastic attitudes for claims about drug toxicity for humans. This source of evidence is far from perfectly reliable—a lethal drug for mice might be beneficial for humans and vice versa—which comes from the fact that mice do not share all the 'key features' of a human organism. But it can hardly be denied that propositions about these experiments are often part of the evidential network of propositions about humans. The same might hold for models as credible worlds (Sugden 2000). Model economies are unlike real economies in many respects, much like mice are unlike humans. But the similarities shared by the two economies might be enough for model propositions to be counted as part of the evidence for real-world propositions.

We will not provide here a general, philosophical account of what it is for a model to be similar to a real economy, similar in the right way such that model propositions can become part of the evidential network for real-world propositions. Yet, the conditions provided in this subsection can help in structuring an argument to the effect that a *specific* model played, or not, an evidential role. This usefulness of our framework is illustrated below in our case study of the DMP model. Our illustration will also show, however, that these arguments are typically hard to uphold.

2.2. Revealing role: the model as hypothesis generator

We now turn to a potential epistemic contribution of models that is less discussed and sometimes simply referred to as a heuristic contribution. As we said above, one arguably learns about the properties of the model by constructing and manipulating it. In consequence, one comes to have justified beliefs in a host of true model propositions. One way by which this initial process can contribute to real-world learning is when some proposition about the model is transposed as a proposition about the world—that is, as a hypothesis about a target system of interest—and that the agent, in the end, comes to know this proposition. We would say in such a situation that the agent comes to *form* a doxastic attitude for the real-world hypothesis because of the model.

There are also three conditions to be met by a model to play this role. First, the end state condition: real-world proposition *p* is true and the agent justifiably believes it. The second condition is that there must be a conceptual connection between the model proposition *q* and *p*. More specifically, propositions *q* and *p* predicate the same, or sufficiently similar, properties to their respective systems of interest. For instance, *q* could say that employment protection is a positive cause of the unemployment rate for model *M*, and *p* could say that the USA is such that employment protection and the unemployment rate are similarly causally related. We do not want to be overly restrictive on the conceptual connection required since some amount of interpretation is always necessary in order to take a property in a model to be sufficiently similar to a real-world property. It should however be clear that not any interpretation will do—e.g., propositions true of Bohr's model of the atom cannot legitimately be interpreted as sufficiently close to hypotheses about whether Bashar al-Assad will still be the president of Syria in 2015.

The final condition states a specific counterfactual dependence: if the agent had not known *q,* she would not have formed a doxastic attitude for *p.* In other words, if the agent had already an epistemic attitude with respect to the real-world hypothesis or if this real-world hypothesis was bound to be considered because of other developments—a case of overdetermination—then the model would not be making an epistemic contribution.

Is there a link between the conceptual exploration discussed in many commentaries on models (e.g., Hausman 1992, p. 79; Nersessian 2008; Morgan 2012, pp. 270-72, 368-72) and this revealing role? It is to be expected that many of the hypotheses revealed by a model come up through the creation, exploration and clarification of some concepts through modelling.[8] After all, models are

[8] As we stated above, here we restrict ourselves to propositional learning from models and say only little about other potential types of learning from models. We already noted the favourable

widely recognised for their role in creating, exploring and clarifying concepts. For instance, the advent of game theory brought many concepts to the forefront, one example being the distinction between incomplete and imperfect information or, more famously, a situation having the structure of a prisoner's dilemma. However, there is no necessity in the connection between conceptual exploration and the revealing role; it might well be that some hypotheses are revealed by further manipulation of a model using well-established categories.

2.3. Stimulating role: the model as stimulus for empirical research

Models can suggest more than hypotheses; they can also suggest ways to increase the density of the evidential network for real-world hypotheses that the agent cares about. In other words, models can stimulate empirical research. This possibility forms the core of the last potential epistemic role we want to emphasise.

Part of the purpose of doing empirical research is to come to form doxastic attitudes for more real-world propositions (e.g., experimental results) that are evidentially related to propositions that the agent cares about. Models can help increase the density of one's evidential network because the propositions to investigate in order to justify one's doxastic attitude for a hypothesis are not always evident. By toying with the model, researchers can come to realise that some empirical research would be relevant to conduct. The role of the model here is thus to *stimulate* pursuing novel empirical research.

Again, three conditions need to be satisfied by a model to fulfil this role. First, the end state condition: a real-world proposition *p* is true and the agent justifiably believes it. Second, there is another real-world proposition *r* for which the agent has a doxastic attitude, but the research that generated the agent's doxastic attitude for *r* would not have been pursued had it not been for the modelling exercise. In other words, the model has a causal influence on the generation of a doxastic attitude for *r* and it has this influence through stimulating research.

Finally, a third condition requires that the doxastic attitude for *r* is a non-redundant element of the evidential network for *p*: if it were not for the doxastic attitude for *r*, the agent would not be justified in believing *p*. The evidential network would not reach the threshold for adequacy if the agent did not entertain this attitude.

prospects of an account also looking at procedural learning (i.e. resulting in know-how). A full epistemological account will also include conceptual learning—the introduction of vocabulary, see Audi (2011, p. 162-163).

Like for the revealing role, the stimulating role is tightly linked to conceptual exploration, although we should not see conceptual novelty as being necessary to the revealing role. The link is tight because what makes modelling a particularly effective activity at coming up with new ways to investigate target systems is, perhaps, that modelling makes us conceptualise the world differently.

3. An illustration with the DMP model

The previous section discussed potential epistemic roles of models. Now we turn to looking at different claims made in the literature regarding the epistemic contributions of the DMP model. The main goal of this section is to illustrate that our epistemic roles neatly dissect various assertions about this model and that they indicate what is required for these assertions to be true. In addition, we will argue that the DMP model actually played specific epistemic roles while recognising that our arguments, being based on empirical propositions, are disputable.

The origins of the DMP model go back to the end of the 1960s when many researchers were looking for new "microeconomic foundations of employment and inflation theory" (the title of Phelps et al. 1970). The core idea embedded in this model is that the labour market is a matching system with search frictions. There are frictions because, on the supply side, job seekers are not instantaneously informed about all the job offers and their associated advantages and, on the demand side, potential employers have no direct access to all job seekers and their wage expectations. A match between a job seeker and an employer takes time since they must find each other. When a match occurs, each side has some bargaining power since it would be costly for the other side to break the match and go back to search mode.[9]

The best way to see the peculiarity of the DMP model is to contrast it to the main model of the labour market predating it. This earlier model—still taught in introductory labour economics—depicts the labour market as a standard neoclassical market with price-taking demand (i.e. firms) and supply (i.e. potential workers). The two sides of the market are summarised—as usual—in a downward-sloping demand and an upward-sloping supply. The quantity of labour actually used and the associated wage rate (if nothing interferes) is taken to be the intersection of these two curves—the competitive equilibrium. In this model, 'unemployment' is interpreted as being caused by factors forcing the wage rate to be higher than the equilibrium wage rate, thus implying an over-supply of labour at the given wage. In contrast with the DMP model, there is no idea of time

[9] For a book-length exposition of the model, see Pissarides (2000); for shorter presentations, see Cahuc and Zylberberg (2004, pp. 517-536) and Nobelprize.org (2010a, pp. 12-20).

necessary for a match to occur, and there is no idea of wage bargaining between the two sides of a match.

3.1. Evidential role

Many economists interpret the DMP model as providing evidence for real-world claims. For example, the press release accompanying the announcement of the 2010 Prize in Economic Sciences stated that:

> The Laureates' models help us understand the ways in which unemployment, job vacancies, and wages are affected by regulation and economic policy. [...] One conclusion is that more generous unemployment benefits give rise to higher unemployment[.] (Nobelprize.org, 2010b)

This claim can plausibly be interpreted as asserting that the DMP model played an evidential role with respect to the real-world proposition: 'In real economies, more generous unemployment benefits give rise to higher unemployment' (henceforth '*p*').

Did the DMP model play an evidential role with respect to *p*? In other words, are the conditions presented in the previous section met? Regarding the end-state condition, there are reasons to grant that it is met: first, the vast majority of economists *believe* the real-world proposition *p*; second, a rich literature using different methods and data seems to provide adequate evidence to *justify* this belief (for surveys, see Fredriksson and Holmlund 2006; Boeri and van Ours 2008, ch. 11); third, since justification reliably—yet fallibly—indicates the truth value of a proposition, *p* is likely to be *true*.

There are also some reasons to grant that at least one proposition about the DMP model is part of the evidential network for *p* (the second condition of the evidential role). In this case, the most plausible proposition *q* is 'In the DMP model, more generous unemployment benefits give rise to higher unemployment', which is indeed a known property of the model. Among economists, the argument for the view that *q* is part of the evidential network for believing *p* includes claims about the realisticness of the model[10] and, most importantly, about the fact that many results of the DMP model concord with results independently obtained with empirical methods (i.e. a claim about the degree of output validation of the model). Since the model seems to track the world so well on many aspects, we can, the argument goes, take truths about it as belonging to the evidential

[10] For instance, Pissarides says in his Nobel lecture: "To me, search theory was appealing as a foundation for a theory of unemployment because it appeared realistic." (Nobelprize.org, 2010c) See also Blanchard 2007, pp. 413-14.

network for real-world propositions like *p*. Note that this argument does not imply the dubious claim that one would be justified in believing *p* on the sole ground of knowing *q*. In the present case, *q* is one element among many more propositions in the evidential network for *p*. Other important propositions are concordant results from statistical analyses of various types that do not rely on the DMP model (see Claveau 2011).

Finally, there is a compelling reason for why the justification of *p* counterfactually depends on knowing *q* (the last condition). This model proposition is a novel model result in the sense that, perhaps surprisingly, the model of the previous generation (the standard supply and demand model, see above) did not have the conceptual resources to produce a relationship between unemployment benefits and unemployment.[11] Since the DMP model has these conceptual resources, it would be pretty devastating if there was no way to produce a positive relationship between benefits and unemployment in the various versions of the model. This incapacity could indicate that the statistical results are all artefacts. By contrast, knowing *q* helps support the belief that the empirical results pointing to a causal link from benefits to unemployment are not all spurious.

Although we side with most economists here in believing that the DMP model played an evidential role with respect to this specific *p*, we readily note that there is room for objections. It could be argued that the end-state condition is not met because, for instance, the evidential network for *p* is more sparse and incongruent than we are ready to admit (cf. Howell 2005). One might also wonder why *q* should be taken as even a mildly reliable guide to the truth-value of *p* given that some elements and results of the DMP model are quite *unlike* the world.[12] Finally, the ones reacting against the centrality of the modelling culture in economics (e.g., Lawson 1997) might reply that *q* is redundant, that we have no need of a model proposition in the evidential network for *p*. We think that these objections can be satisfactorily answered, but these answers would require developments unnecessary for the purpose of this paper.

[11] The closest proposition to q one could get in this earlier model is: 'In this model, higher unemployment benefits decrease employment.' Indeed, generous unemployment benefits were modelled as decreasing the labour supplied at any wage; thus decreasing equilibrium employment, not increasing unemployment. See Boeri and van Ours 2008, pp. 230-34.

[12] For instance, even proponents of the model take its depiction of bargaining as being "a very poor description of reality" (Blanchard 2007, p. 414) and recognise that it does not manage to replicate even something as central as the cyclical fluctuations in unemployment (Nobelprize.org 2010a, p. 23; Shimer 2005).

3.2. Revealing role

The DMP model has been praised for being a great platform to think about the labour market. For instance, Olivier Blanchard (2006, p. 26), current chief economist of the International Monetary Fund, wrote that, compared to the earlier model, the DMP model is a "richer framework to think about unemployment, a framework based on flows, matching and bargaining." Blanchard is here emphasising the possibility of extensive *conceptual* exploration through the DMP model. The concepts brought to the forefront by this model include a clear distinction between flows and stocks of workers, matching efficiency, search intensity, and wage bargaining. One way by which this conceptual exploration can result in *propositional* learning about the world is the revealing role.

There are many new questions that can be investigated inside the DMP model but could not in the earlier model: what is the relationship between the *stock* of unemployed people and the flows in and out of unemployment? What determines the speed at which potential workers are matched to firms? More specifically, what determines the search intensity of unemployed persons? What matters to the bargaining process between firms and their potential employees? For the DMP model to play a revealing role, a necessary condition is that some answers to these or similar questions with respect to the model be transposed as hypotheses about the real world.

Take, for example, what economists call the 'entitlement effect' of unemployment benefits (Mortensen 1977; Boeri and van Ours, 2008, sec. 11.2.2). While discussing the evidential role, we said that unemployment benefits are believed by most economists to increase unemployment, but Mortensen realised that, in one version of his model, one *group* of job seekers had shorter spells of unemployment when unemployment benefits were higher.[13] This group is the one that is not covered by the unemployment benefit system, but can expect to be covered during its next unemployment spell. Since getting a job also involves a better future as unemployed, this group has incentives to find a job faster.

Once this entitlement effect is shown to exist in the model, economists might entertain a related hypothesis about the world: 'Increasing unemployment benefits in a real country will reduce the length of unemployment spells for at least some uncovered job seekers' (henceforth '*p*'). The DMP model seems to have played a revealing role with respect to learning *p*.

The second condition for the revealing role—the conceptual connection—should be easy to grant in this case. Although the DMP model is highly idealised, we can locate a group of agents in it

[13] It is because the entitlement effect is dominated by other effects at the aggregate level that most economists believe that, for a whole economy, unemployment benefits increase unemployment.

corresponding to the real individuals that are both jobless and unaided by the unemployment insurance system. We can also easily associate a property of the model group to the lengths of unemployment spells in our real group. The conceptual link between the entitlement effect in the DMP model and *p* is thus hard to question.

Is the end-state condition fulfilled? Many economists—especially among the ones specialising in labour economics—*believe p*, which claims only the existence of entitlement effects among *some* job seekers. Although few empirical studies have tested the real-world existence of entitlement effects, the existing results seem sufficient to *justify p*.[14] And *p*, given this evidence and given the weak requirement for the claim, is likely to be *true*.

Can we also grant the last condition that it would not have occurred to economists to believe *p* had it not been for the development of the DMP model? To the best of our knowledge, *p* was not entertained prior to the modelling work of Mortensen (1977) and there is no parallel literature today talking about something like *p* without being aware of Mortensen's work. Though we cannot definitively rule out that *p* was bound to be entertained soon enough independently of the development of the DMP model, the available evidence points toward the fulfilment of this last condition too.

3.3. Stimulating role

A contribution of the DMP model might have been to stimulate empirical research in epistemically valuable directions. The Economic Sciences Prize Committee claims that the development of the DMP model had this effect. According to this committee, the contribution was twofold. First, the model stimulated data collection:

> The early microeconomic models of job search initiated new data collection efforts focusing on individual labour market transitions, in particular transitions from unemployment to employment (Nobelprize.org 2010a, p. 20).

By changing the modelling focus from stocks to flows, the development of the DMP model stimulated researchers to request (or, less frequently, actually gather themselves) reliable data on flows.

[14] For instance, Bennmarker et al. (2007) find evidence of an entitlement effect for men, but not for women.

Second, the DMP model gave impetus, according to the Prize Committee, to the use and refinement of some empirical methods, most importantly duration analysis[15]:

> The methodological literature on econometric duration analysis has expanded substantially over the past couple of decades, a development that is to a large extent driven by the growth and impact of microeconomic search theory. (Nobelprize.org 2010a, p. 23)

Can we thus say that the DMP model played a stimulating role? Take, for instance, the following proposition *p*: "the [U.S.] private-sector (gross) job creation rate began declining well before the 2001 recession and continued to slide until the middle of 2003." (Davis et al. 2006, p. 24) It can hardly be doubted that the first and the last conditions hold with respect to *p*.

To start with, Davis et al. *believe p*. They base this belief on the analysis of two data sources: the Job Openings and Labour Turnover Survey (JOLTS; see Clark and Hyson 2001) and the Business Employment Dynamics (BED) data (Pivetz et al. 2001). The authors put forward two propositions in their analysis. First, "[f]igures 2 and 3 [plotting BED data] show a long downward slide in job creation rates before, during and well after the 2001 recession." (p. 12). Second, "[t]he hires rate [from the JOLTS] declines from 3.8 per cent of employment in December 2000 to 3.0 per cent in April 2003" (p. 13). We denote these two propositions q_1 and q_2. Note that q_1 and q_2 are about patterns in data, while *p* is directly about the United States. Propositions q_1 and q_2 constitute the main evidential ground for *p*. It is thus hard to deny that believing them is a *necessary* condition for being justified to believe *p*. Furthermore, q_1 and q_2 seem to be *sufficient* to justify believing *p*. In particular, the fact that both data sources produce a similar pattern makes it unlikely that this pattern is driven by an artefact in the data. Finally, since believing *p* seems to be *justified*, we should be tempted to grant the *truth* of *p*. In short, it is highly plausible to affirm that Davis et al. *know p* (first condition) and that they do so thanks to q_1 and q_2 (last condition).

For the DMP model to have played a stimulating role with respect to *p*, the second condition must also hold: if the model had not been developed, would economists be in a position to believe evidential propositions like q_1 and q_2? The opinion relayed by the Prize Committee (see above) is that the model is responsible for the collection of new data like the JOLTS and BED data, and thus, ultimately, for the beliefs in q_1 and q_2. We have no substantial reason to reject this opinion—the data collection and the active development of the methods started after the initial work on the

[15] Duration analysis as applied to labour markets empirically studies the length of unemployment spells and the factors explaining it.

DMP model and the scholars involved in all these developments had significant interactions during the period.

That the DMP model played a stimulating role with respect to learning *p* might not be granting much. This proposition is descriptive and it pertains to a single country for a specific period of time. But the stimulating role of the DMP model might become impressive if we can be convinced that many other propositions were learnt through this role. These propositions will not be necessarily local and descriptive; they could be descriptive *generalisations* justified by pooling national surveys together or they could be *causal* propositions justified by combining duration analysis and natural experiments. We do not have space to explicitly argue for these epistemic contributions of the DMP model. We simply note that, if we grant these contributions, the DMP model would have stimulated a great deal of learning about real economies.

The same point holds for the evidential and revealing roles. By focussing on specific propositions, we could have given the impression that these contributions amount to little. But the overall epistemic contribution of the DMP model would be impressive if convincing arguments using a wide array of important real-world propositions could be constructed.

Conclusions

A model can make us learn in a variety of ways. This paper discussed several ways by which propositional learning can occur with models. Manipulating the model can obviously make us learn truths *about* the model itself. But, more importantly, a model might also contribute in different ways to make us learn about the world. We discussed three such ways. First, truths about the model might be part of the evidence justifying one's belief in a true real-world proposition—the evidential role. Second, truths about the model might reveal real-world hypotheses that turn out to be true and justifiable—the revealing role. Third, the model might stimulate researchers to undertake new empirical research, the result of which comes to justify beliefs in some true real-world propositions—the stimulating role. For each of these roles, we provided and discussed a list of conditions. We then used this framework to analyse the praises given to the DMP model.

References

Alexandrova, Anna. 2008. “Making Models Count.” *Philosophy of Science* 75 (3): 383–404.

Arrow, Kenneth J., and Gerard Debreu. 1954. “Existence of an Equilibrium for a Competitive Economy.” *Econometrica* 22 (3): 265–290.

Audi, Robert. 2011. *Epistemology: A Contemporary Introduction to the Theory of Knowledge, Third Edition.* New York: Routledge.

Bennmarker, Helge, Kenneth Carling, and Bertil Holmlund. 2007. “Do Benefit Hikes Damage Job Finding? Evidence from Swedish Unemployment Insurance Reforms.” *Labour* 21 (1): 85–120.

Bernecker, Sven, and Fred I. Dretske. 2000. *Knowledge: Readings in Contemporary Epistemology.* Oxford: Oxford University Press.

Bernecker, Sven, and Duncan Pritchard, ed. 2011. *The Routledge Companion to Epistemology.* London: Routledge.

Blanchard, Olivier. 2006. “European Unemployment: The Evolution of Facts and Ideas.” *Economic Policy* 21 (45): 5–59.

Blanchard, Olivier. 2007. “Review of ‘Unemployment: Macroeconomic Performance and the Labour Market’.” *Journal of Economic Literature* 45 (2): 410–418.

Boeri, Tito, and Jan van Ours. 2008. *The Economics of Imperfect Labor Markets.* Princeton: Princeton University Press.

Cahuc, Pierre, and André Zylberberg. 2004. *Labor Economics.* Cambridge, MA: MIT Press.

Cartwright, Nancy. 1989. *Nature’s Capacities and Their Measurement.* Oxford: Clarendon Press.

Clark, Kelly A., and Rosemary Hyson. 2001. “New Tools for Labor Market Analysis: JOLTS.” *Monthly Labor Review* 124 (12): 32–37.

Claveau, François. 2011. “Evidential Variety as a Source of Credibility for Causal Inference: Beyond Sharp Designs and Structural Models.” *Journal of Economic Methodology* 18 (3): 233–53.

Davis, Steven J., R. Jason Faberman, and John Haltiwanger. 2006. “The Flow Approach to Labor Markets: New Data Sources and Micro–Macro Links.” *Journal of Economic Perspectives* 20 (3): 3–26.

Fantl, Jeremy. 2012. "Knowledge How." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2012. http://plato.stanford.edu/archives/win2012/entries/knowledge-how/.

Fredriksson, Peter, and Bertil Holmlund. 2006. "Improving Incentives in Unemployment Insurance: A Review of Recent Research." *Journal of Economic Surveys* 20 (3): 357–386.

Gettier, Edmund L. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23 (6): 121–123.

Glanzberg, Michael. 2013. "Truth." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2013. http://plato.stanford.edu/archives/spr2013/entries/truth/.

Grüne-Yanoff, Till. 2009. "Learning from Minimal Economic Models." *Erkenntnis* 70 (1): 81–99.

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

Hetherington, Stephen. 2011. "The Gettier Problem." In *The Routledge Companion to Epistemology*, edited by Sven Bernecker and Duncan Pritchard, 119–130. London: Routledge.

Howell, David R., ed. 2005. *Fighting Unemployment: The Limits of Free Market Orthodoxy*. New York: Oxford University Press.

Ichikawa, Jonathan Jenkins, and Matthias Steup. 2012. "The Analysis of Knowledge." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2012. http://plato.stanford.edu/archives/win2012/entries/knowledge-analysis/.

Kim, Jaegwon. 1973. "Causes and Counterfactuals." *The Journal of Philosophy* 70 (17): 570–572.

Lawson, Tony. 1997. *Economics and Reality*. London: Routledge.

Mäki, Uskali. 2009. "MISSing the World. Models as Isolations and Credible Surrogate Systems." *Erkenntnis* 70 (1): 29–43.

Morgan, Mary S. 2012. *The World in the Model: How Economists Work and Think*. Cambridge, UK: Cambridge University Press.

Morgan, Mary S., and Tarja Knuuttila. 2012. "Models and Modelling in Economics." In *Philosophy of Economics, 1st Edition*, edited by Uskali Mäki, 49–88. Handbook of the Philosophy of Science. Oxford: Elsevier.

Morgan, Mary and Morrison, Margaret. 1999. *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge, UK: Cambridge University Press.

Mortensen, Dale T. 1977. "Unemployment Insurance and Job Search Decisions." *Industrial and Labor Relations Review* 30 (4): 505–517.

Nersessian, Nancy J. 2008. *Creating Scientific Concepts*. Cambridge, MA: MIT Press.

Nobelprize.org. 2010a. "Markets with Search Frictions" Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2010. The Royal Swedish Academy of Sciences. http://www.nobelprize.org/nobel_prizes/economics/laureates/2010/advanced-economicsciences2010.pdf.

Nobelprize.org. 2010b. "The Prize in Economic Sciences 2010 - Press Release," The Royal Swedish Academy of Sciences. http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2010/press.html.

Nobelprize.org. 2010c. "Equilibrium in the Labour Market with Search Frictions," The Royal Swedish Academy of Sciences. http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2010/pissarides-lecture.html.

Phelps, Edmund S., Armen A. Alchian, Charles C. Holt, Dale T. Mortensen, G. C. Archibald, Robert E. Lucas, Leonard A. Rapping, et al., ed. 1970. *Microeconomic Foundations of Employment and Inflation Theory*. New York: Norton.

Pissarides, Christopher A. 2000. *Equilibrium Unemployment Theory, Second Edition*. Cambridge, MA: MIT Press.

Pivetz, Timothy R., Michael A. Searson, and James R. Spletzer. 2001. "Measuring Job and Establishment Flows with BLS Longitudinal Microdata." *Monthly Labor Review* 124 (4): 13–20.

Reiss, Julian. 2008. *Error in Economics: Towards a More Evidence-Based Methodology*. London: Routledge.

Shimer, Robert. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *The American Economic Review* 95 (1): 25–49.

Sugden, Robert. 2000. "Credible Worlds: The Status of Theoretical Models in Economics." *Journal of Economic Methodology* 7 (1): 1–31.

Zagzebski, Linda. 1994. "The Inescapability of Gettier Problems." *The Philosophical Quarterly* 44 (174): 65–73.

4

What Do Philosophical Theories Say about Model Failure?

What do Philosophical Theories Say about Model Failure?

Introduction

Some models haven't fared particularly well lately. Their inability to anticipate the financial crisis and to come up with policies for treating its consequences has amply been documented. Andrew Haldane, the chief economist of the Bank of England, recently conceded that the models' failure to predict the financial crisis as well as the effects of Brexit on the British economy was due to exceedingly narrow models not coping with irrational behaviour (Inman, 2017). Likewise, in the political domain, the failure to predict the election of Donald Trump to the United States' presidency continues to have laypeople, political scientists, and other pundits aghast, in part because otherwise reliable models from different sources using different methods all pointed in the same direction. How should we understand this situation? How can we judge whether it was models that failed and not that, for instance, economists, or some of them, misused their models, as (Rodrik, 2015, Chapter 1) suggests is mostly the case in economics? Are there model failures that could have been prevented, or from which we can learn to avoid them happening again?

To attempt to answer questions like these, one option is to turn to the extant philosophical literature on models. Interest by philosophers in the use of models in science has increased significantly in the last couple of decades, generating a vast literature. The main motivation for this interest, according to Frigg & Hartmann (2009), is that philosophers have come to realise the ubiquitous role that models play in science. They suggest that despite the many different kinds of models, there are mainly three types of questions that the literature has attempted to answer. First, there is literature on the ontology of models, which addresses questions such as what models are and how they relate to theories. Second, there is literature on the semantics of modelling, which mainly tries to identify how models relate to the things they are models of: their targets. Finally, there is literature on the epistemology of models, which addresses questions related to what and how we can learn from models.

In general, this literature has been nurtured by the 'mystery' that models pose, namely that models, despite their idealisations and abstractions, can still be used for practical, real world purposes. In other words, considering that science is regarded as a successful cognitive enterprise, and models are ubiquitous in realising such a success, the main interest of this literature is to understand the success of models: why they are useful. From this point of view, the three types of questions or categories in which Frigg & Hartmann (2009) divide the literature, can also be interpreted as three different perspectives from which this question can be tackled.

In this chapter I shall explore these categories in order to find out what they can contribute to the discussion of model failure. Therefore, understanding the role of models and the practice of modelling, as philosophers have attempted to do, should inevitably be able to account for this negative aspect of the practice. My aim is therefore to identify criteria in the current philosophical accounts of models that may offer an indication of what may make model failure more likely. Ideally, this should also offer clues that may allow us to prevent future model failure.

The structure of the chapter is as follows. In section II, I briefly introduce the context of the discussion and offer a working definition of model failure. The next three sections discuss each of the branches of the literature introduced above. I begin with a few words on the ontology of models in section III. From this discussion, it should be clear why I don't engage with this literature further, given the aims of this chapter. In section IV I then discuss the semantics of modelling, focussing on the conditions for scientific representation offered by Frigg & Nguyen (2016). I also discuss Weisberg's (2013) attempt to cash out his account of modelling using the notion of similarity as the relation that is established between model and the world. In section V I move on to a discussion of the epistemology of models, focussing on the accounts offered by Aydinonat (2007; Grüne-Yanoff (2009, 2013); Kuorikoski & Ylikoski (2015) and Ylikoski & Aydinonat (2014). In section VI, I raise some possible objections to my (implicit) claim that in this literature we should be able to find elements that say something about model failure. The final section offers conclusions.

2. Context and model failure

Claims of model failure have been quite common recently. Academic economists and political scientists, pundits, and policy makers have all been the target of different sorts of accusations about the use of faulty models to predict the financial and economic crisis, to respond to it and to predict the outcome of political elections, among other things. Surely some of these claims are not new. Economics, in particular, has long been characterised by a sort of partisanship: there is the orthodoxy on the one hand, and on the other a bunch of marginalised heterodox schools of thought that have offered different conceptual and methodological approaches. In turn, these schools have also directed harsh criticism against the orthodoxy. Philosophers, for their part, have also offered their share of critique of both economics and economists. Famous critiques of economics by philosophers include Alex Rosenberg's (Curtain & Rosenberg, 2013) pungent claim that economics is not a science but, at best, a craft[1]. Recently though, the critiques that have been specifically directed at models have become more pressing, partly because of the dramatic consequences that events like the crisis had, and because the criticism, at least with respect to economics, has come this time not only from those at the margins, but from insiders, which include some Nobel laureates such as Paul Krugman and Joseph Stiglitz (Krugman, 2009; Stiglitz, 2011, 2015).

One way to address these criticisms is to turn to the literature on models in philosophy. Even though models have been used in science for at least a century, the philosophical literature about this topic is relatively recent. It has grown to a considerable size only in the last couple of decades. During the first half of the twentieth century, when logical positivism was at its climax, philosophers tended to think of models as temporary heuristic tools, that is, tools for the discovery of new ideas that would eventually turn into proper theories. Given that most of them believed in a neat division between the context of discovery and the context of justification, models were not thought to be part of the realm of concern for philosophers—which at the time was restricted to the context of justification. Since the beginning of the 1980s, though, with work such as Nancy Cartwright's *How the Laws of Physics Lie* (1983), models began to receive much more attention[2], surpassing other, more traditional topics, in the philosophy of science.

Theories of scientific representation, for instance, lie at the crossroads of two branches of literature in philosophy (Suárez, 2010). One of these branches is "analytical philosophy", which is concerned with understanding the relationship of theory with the world. The other branch of the literature is the "philosophy and history of science" with an interest in developing "a proper understanding of the practice of modelling in the sciences" (2010, p. 91). Even though, according to Suárez, the former precedes the latter historically, it is the

[1] Rosenberg is ambivalent, at best, about his views on economics. In his (1992), he challenges the status of economics as a science. In (2009) he concedes that, in part thanks to some developments in economics, it may be regarded as a 'biological' science. In his (2013) he maintains, again, that economics is not a science but a craft.

[2] Hesse (1966) was an exception in its time.

latter that has become relatively more important due to the attention that models have received in the last years. A growing interest has thus emerged in understanding models, the role they play in science, and how exactly it is that they afford their epistemic benefits. Hence, turning to the extant philosophical theories of models seems worthwhile if the goal is to understand model failure.

To try to come to grips with the aforementioned criticisms by looking *only* at the extant philosophical accounts of models might come across as naïve. For such an exploration seems to presuppose that the critics are right and that the causes of the crisis are exclusively related to model failure. Naturally, the situation is much more complex. There was a savings glut, mainly coming from China that drove US government bonds' interest rates down and in turn drove investors to search for riskier and higher returns, in the form of CDOs. There are also issues related to policy and regulation. To have let Lehman Brothers go bankrupt is itself a matter of controversy concerning the effects it had on the real economy. And regulatory measures by central bankers established at Basel could have been stricter in relation to capital ratios, the definition of capital or the share of a bank's assets that should be liquid (see The Economist (2013)). These are just a handful of all the causes that have been discussed—Davies (2010) discusses thirty-eight sets of causes offered—but enough to suggest that an exclusive look at models is an extremely narrow approach.

It is thus important to state clearly from the outset the scope and aim of the chapter. Without pretending to downplay the significance of other causes, the chapter's scope is the question of potential model failure. The reason is that the chapter's aim is to explore current philosophical accounts of models and not the financial crisis. The financial crisis and the subsequent recession are taken as an important motivation to ask the question of what philosophical accounts of models have to say about model failure. This means that, for my analysis, I suspend judgement about the accuracy of the claims that models caused the crisis and simply take for granted that models *can* fail. It doesn't mean, however, that the question is irrelevant for examining the causes of the crisis. If philosophical accounts of models offer insights about model failure in general, for instance by establishing whether there are certain conditions that make model failure more likely, these insights can illuminate a thorough examination of the crisis that considers the role of models and the many other aspects that might have been significant. In short, the chapter takes a rather indirect route to the question of what the significance of models in the latest crisis was. It enquires about what philosophical accounts of models say of failure, in general.

Until now, I have only hinted at what I mean by model failure by alluding to the criticisms that have been made of models in economics and political science. Now let me say a little bit more about what I have in mind when I ask whether philosophical theories of models are able to offer criteria that allow us to identify 'model failure'.

A good starting point for thinking about model failure is to ask whether there is a precise characterisation of what a modelling exercise consists of, which includes a general idea of the things that might go wrong in this procedure. Here I'm thinking along the lines of a common understanding of the modelling process and the identification of "typical" failures very much in the same way an electrical appliance fails. A dishwasher or an oven, have some common failures that are typically acknowledged in the troubleshooting section of the user manual. Some failures, such as that the appliance doesn't turn on, are acknowledged in these sections as more common than say, that screws loosen or that the springs of buttons break down. The important point here is the fact that common failures are identified as vulnerabilities of the appliance.

A similar, perhaps less mundane example of what I have in mind is the analysis that is carried out to determine the causes of accidents in general, and airplane accidents in particular. In the first years of the aviation industry, accidents were likely to be mainly caused by mechanical failures. However, since the 1950s, with important improvements in aviation technologies and training, this trend has been changing, with at least some form of human error accounting for 70 to 80 percent of the accidents (Shappell and Wiegmann, 1996, cited in Shappell & Wiegmann (2003). Accident investigations lead to the identified causes to be catalogued accordingly, having so far generated an important database that allows for the identification of patterns in the causation of accidents. Identification of these patterns allows in turn for the design of preventive measures such as the early replacement of parts, redesign of an interface or specific training for the crew in human factors. This collection and analysis of information has significantly contributed to the decrease in accidents in aviation in the last decades, making airplanes one of the safest modes of transportation.

There are two kinds of data collected: technical and human factors. Engineers are more successful in cataloguing technical data. There is much less ambiguity about the technical causes of an accident, given that it's easier to detect malfunction in components and parts. The human factors, on the other hand, are still an area under development and several frameworks of human error have been offered, all with the purpose of capturing the right categories for identifying human-error causes of accidents. Shappell & Wiegmann (2003) have identified six frameworks or perspectives from which human error can be analysed. For instance, there is the cognitive perspective, which conceptualises the agent's mind—e.g. the pilot—as an information processing system. Failures that are commonly detected under this framework are information related, such as whether the pilot was able to detect changes in the system, and if so, whether on the basis of that information the diagnostic made by the pilot was accurate. Failure to detect a change in the system would be regarded as information error and failure to diagnose the change accurately would be regarded as diagnostic error. Other identified errors are goal error, strategy error, procedure error, and action error. Accident analysis under this framework suggests that failures of judgement are typically associated with major accidents whereas procedural and execution errors are more likely to lead to minor accidents (Shappell & Wiegmann, 2003, Chapter 2). Another framework is the ergonomic or systems perspective, which starts from the assumption that errors occur generally at the interaction between humans, the machines they operate, and the environment in which they operate. This kind of analysis has contributed to improvements in cockpit layout, for instance. Other perspectives are the behavioural, which regards agents and their performance as driven by rewards and punishments; the aeromedical, which assumes that errors are symptoms of fatigue or illness; the psychosocial, which emphasises errors as coming from failure in human relations and communication among teammates; and, finally, the organisational perspective, which emphasise failures in the decision making of managers, supervisors, and the organisation in general. The current challenge of the aviation industry is to be able to find a unified framework that can make use of the advantages of each of these frameworks.

It doesn't seem like here are a priori reasons for why this sort of analysis would not be feasible for models. To the contrary, if we look at how this has worked in aviation, the reduction in airplane accidents has occurred because there is a *thorough understanding* of what flying airplanes entails, which includes not only the technical details of an airplane, but the full process including human interaction and organisational arrangements, and the attempt to use this understanding for accident prevention. Philosophers claim to be attempting to understand the modelling practice, which makes already for half the task. Whether it's possible to develop this kind of framework for models is a question that can only be answered a posteriori.

There is, however, a difficulty with accepting this kind of analysis as our guiding framework. While in aviation there's an obvious candidate for failure, namely accidents (with or without fatalities), in the use of models it is unclear when exactly a model can be regarded as having failed. Model failure is therefore not so easily tracked. It is difficult to say unequivocally and generally what failure constitutes, considering that there are arguably, at least a priori, many ways in which a model might fail. To begin, a model fails relative to a specific purpose. Prediction and explanation are two of such purposes. And then, it is possible that a (epistemic) purpose such as prediction, has different goals—e.g. to evaluate a policy or some financial gain. Another difficulty is that it is not always clear what the purpose (and goal) of a specific model is. Furthermore, even if we observe what could be taken as an analogous case to the accident, say the financial crisis, there might still be many other causes that have nothing to do with models, like I suggested above.

So how can we identify model failure such that it's possible to establish what philosophical accounts of models have to say about failure? There are four obvious options. The most obvious is simply to survey philosophical accounts of models and search for explicit treatments of failure. As I mentioned at the outset, however, the accounts so far offered have mostly focussed on explaining success. With the exception of Mäki (2017), which will be treated in the next chapter, there aren't philosophical accounts of models that treat failure explicitly. A second option is to a priori determine what model failure is and again survey the philosophical accounts to determine to what extent they explain failure in the specified sense. The problem with this strategy was already mentioned above: models fail relative to a specific purpose. It's thus difficult to see how an a priori general definition of failure—that is not as unhelpful as "models fail when they are incapable of fulfilling their purposes"—would be able to say anything specific about how and when models fail. A more specific definition may leave many instances of failure undetected. A third option is to explore models directly in the economic literature, try to determine how they might fail and then measure the philosophical accounts against these models. The greatest difficulty with this strategy (at least for the purposes of this chapter) is that the number of economic models that could be explored would be too limited in order to also make a fair assessment of the philosophical literature. A philosophical account could account for certain kinds of failure that haven't been detected in the economic models and vice versa. This wouldn't be an even comparison between philosophical accounts. The last option is to search for a sort of proxy or something that helps to identify potential sources of failure. Criticisms, if not a proxy for model failure—they could be made for a different reason than the actual failure of models—if well-articulated, might be very helpful in identifying not only those models that have failed but why. Even though this reduces the set of economic models to explore, the investigation would still require going through all criticisms made of economic models, before we can measure these potential sources against the philosophical accounts.

Being left without any satisfactory strategies to pursue this exercise, I will proceed as follows. First, I will simply accept that, for the moment, we can only start with a vague and general definition of model failure such as "models fail when they have been properly used and are still incapable of fulfilling their purposes". Surely, "to have been properly used" is very vague. But it is trying to exclude the easy cases in which there is intentional manipulation of the model with deception and tricks to make it either fulfil its purpose—e.g. contrast it with fake empirical data—or make it incapable of fulfilling its purpose—e.g. a climate change denier who tampers with the model to give non-accurate data about global warming. The second thing I'll do is go through one of the criticisms voiced with respect to the crisis, namely about the failure of models to predict the housing bubble. Going through this case might be useful to refine the definition, even if slightly.

Something to keep in mind, which might be helpful to identify when a model fails, is that we seem to regard a model to have failed when there is an expectation that a model is able to deliver on its purpose. These expectations—e.g. that the housing market should have been predicted—might arise either because this ability to fulfil the purpose has proved to be there before, or because there was a claim that this particular achievement was possible. The first kind is founded on an inductive inference—say, because bubbles in the housing market have been predicted before, so it is expected that it be predicted once again—whereas the latter is founded on a claim about what a model is meant to do and capable of attaining. Either way, an important aspect to understand model failure is to identify why the expectation arose in the first place and whether it is legitimate that it arises. This implies that any attempt to understand why a model may be considered to have failed inevitably requires analysing the antecedents of the model, including the context in which it was developed—otherwise, we can't track how the expectation arose. This seems to be a difference with understanding model success: we might be able to prove the success of a model by contrasting it with data or with background knowledge without knowing what the expectations of the model were. In fact, the success of some models is sometimes attributed to their offering of unexpected results, which become new hypotheses about the world (this dissertation, Chapter 3). By contrast, we can't show a model to have failed if we don't know what was expected of the model in the first place.

With this in mind, let me now discuss the criticism made about the housing market in the US prior to the crisis. The housing market is one of the most significant factors in the run-up to the crisis for at least two reasons. First, it is a massive market; housing investments (residential and non-residential) account for half of all gross private investment, and the liabilities of home mortgages are more or less equivalent to two thirds of the US GDP (Chambers, Garriga, & Schlagenhauf, 2009). Second, the structured financial instruments such as collateralised debt obligations (CDOs), whose trade came to a sudden halt in 2008, were backed up by prime and subprime mortgages. One of the underlying forces driving the surge in housing demand and of CDOs was the belief that house prices would continue to rise, sometimes at incredible rates, as a survey by Case & Shiller (2003) of home owners attests. A great deal of the criticisms that have been made of economists about the failure to anticipate the financial crisis have included criticisms for failing to notice or acknowledge that there was a housing bubble—e.g. Colander et al. (2009); Pettifor (2006). Just like in the investigation of the causes of an airplane accident, in which one of the purposes is to identify whether there was some kind of human error, in the case of the housing bubble we can ask whether economists failed indeed to acknowledge or predict the existence of a housing bubble. If so, the question is whether this constitutes failure.

If economists are now being blamed for failing to recognise that there was a bubble, it is not necessarily because they didn't think about it or did not discuss it. There was, in fact, quite some discussion about this issue in the aughts. An analysis by Gerardi, Foote, & Willen (2010) of the literature that was being published in the run-up to the crisis about whether there was a housing bubble, argues that economists were mostly agnostic about the fact, with a few exceptions on both sides of the debate. The agnostics often found some signs of a bubble but thought that this evidence was not conclusive to assert that this was the case. Those that did take sides on the debate were often relying on evidence that corresponded to different methodological choices with respect to how house prices are measured relative to fundamentals. For instance, a common way to do it is to use the price-rent ratio, which follows the same rationale as equity markets, which uses the price-dividend ratio. Standard theory states that the price of an asset should be equal to the present value of the sum of expected dividends. The dividend of a housing asset is thought to be the flow value of shelter, which is roughly equal to the rental price. The housing pessimists, as Gerardi et al. (2010) refer to those who argued there was a housing bubble, relied on rising price-rent ratio as evidence for the existence of a bubble. In a seven-year period, between 1995 and 2002, the house price index rose by almost 30 percent, whereas the rental index rose by 10 percent (Gerardi et al., 2010). The optimists, on the other hand, did not dispute that the price-rent ratio had risen more than rents in these years but argued that the price-rent ratio as it is calculated is not a good measure to determine whether housing prices accurately reflect fundamentals. In particular, Himmelberg, Mayer, & Sinai (2005) argue that the correct calculation of the financial return of an owner-occupied property is a comparison between the value of living in the property for one year and the opportunity cost of that capital. In consequence, they calculate the one-year cost of owning a house—the imputed rent—which includes six elements representing both costs and offsetting benefits such as tax deductibility on mortgage interests, that can then be compared with rental costs[3]. According to this measurement, there was no housing bubble.

My purpose with this example is not to suggest that one methodological choice was better than the other. Instead, such a debate raises a few questions that are helpful to guide the search for the sources of failure and that can ultimately illuminate the identification of "weak points" or points along the modelling process where failure is more likely to happen. The first thing to note is that there was disagreement among economists with respect to whether there was a bubble and, especially, a widespread scepticism. A set of questions that emerges is, what is the position that, given the tools and background knowledge available, was expected for economists to take? Could we expect economists to accurately predict a housing bubble? If we accept that there was a bubble[4], does this mean that the sceptics and the optimists were wrong? My point here is to question whether it is reasonable to expect that economists predict the bubble or whether perhaps scepticism was a reasonable position to take given the evidence available at the time. On the answer we give to this prior set of questions depends whether we regard the inability to recognise the existence of a bubble as a failure.

Another set of questions is whether we attribute the failure to recognise the existence of a bubble to economists' judgement, to the evidence, or to the models they were using? Before we can do that, we need to answer other questions such as the reasons each side

[3] For details see Gerardi, Foote, & Willen (2010); Himmelberg, Mayer, & Sinai (2005).
[4] Eugene Fama claimed in 2010 that bubbles can't exist since they can't be predicted (see interview by John Cassidy (2010).

had to defend their position. That is, whether the disagreement was ideological, methodological or of any other kind. Above, I noted that the disagreement between optimists and pessimists was methodological; they had different views about what the right way to measure the price of an asset such as housing was. A question that needs to be answered here is whether and why the optimists were wrong and what (methodological) reasons they had to defend their position. This is important because, even if they were wrong about the outcome, they could be right about their methodological choices. The same holds for the pessimists: they were right in the outcome, there was a bubble, but not necessarily because their methodological choice was the most appropriate. With respect to the agnostics, was evidence the only reason for defending this position? Gerardi et al. (2010) make a very interesting observation about the practice that suggests that beliefs and incentives play a rather important role.

In their discussion of the literature, Gerardi et al. (2010) point out that the "Fundamental Theorem of Asset Pricing", the basis of modern asset pricing theory, which states that the evolution of asset prices is unpredictable, make the reluctance to commit to one of the two opposing positions in the housing bubble debate unsurprising. They cite three reasons in particular for this. First, with the theorem as a widespread belief, the burden of proof for those who claim that assets are under- or overvalued is huge. That is, the theorem is more or less the default position and to prove that in a particular case it does not apply requires a higher burden of proof than to confirm the validity of the theorem. Second, given the importance of expectations for the performance of the economy, economists at policy institutions might have shun from commenting and taking a position publicly to avoid self-fulfilling prophecies. That is, even if they had opinions about the housing bubble, they might have kept them to themselves. Finally, they suggest that economists might have abstained from taking a position for fear of damaging their reputation. In other words, considering that it is generally believed that the fundamental theorem is true, suggesting otherwise was a very risky bet, reputation-wise. This suggests that the reasons for being sceptical about the conclusiveness of the evidence, at least publicly, are not purely epistemic or methodological.

This case illustrates the complexity of what the assessment of the modelling practice involves and the many dimensions that need to be considered for a thorough understanding of model failure. More specifically, the previous example brings to the fore the need for discussion about how expectations of the performance of models are formed, the role that evidence plays in the judgements economists make and the role of beliefs and private incentives of economists and policy makers in the claims they make. Another important aspect is how to account for disagreement among the profession with respect to an economic phenomenon. This shows that it is not even clear to determine what the purposes of a model are. Something like reputation (the attempt to preserve it) might determine how evidence is interpreted and thereby the purpose that a model is meant to fulfil.

Given that this seems to complicate the task I have for this chapter much more rather than simplify it, the only alternative left to have at least a preliminary answer to what philosophical accounts of models say about failure, is to take these philosophical contributions for what they are. That is, since these accounts have mostly focussed on accounting for model success, in terms of say, representation or explanatoriness, we can try to explore whether the criteria that have been offered for success can also be used as criteria for failure. Hence, the survey that follows of philosophical accounts of models starts from the premiss that, in principle at least, the elements that have been highlighted by the literature as explaining the success of models should, at the same time, be able to say something about failure, even if failure hasn't been explicitly addressed.

3. Extant philosophical theories of models

As I said above, I will use the characterisation of Frigg & Hartmann of the extant literature to guide my exploration of the literature. I will discuss the ontology of models very briefly, since this is the branch in which it is less likely that criteria related to model failure is found. Still, it is important to briefly discuss what this literature has been mostly concerned with. Afterwards I shall continue with the other two branches, namely the semantics and the epistemology of modelling.

3.1. Ontology of models

Insofar as models are physical, things we can see, touch, and, in general, manipulate, there isn't an ontological conundrum about what they are or about the epistemological implications of such kind of object. Properties of the model are easily comparable with its target. For instance, in the San Francisco Bay - Delta model, it is relatively easy to distinguish in which aspects exactly the model differs from its target. So, the scale, the material, the mapping of the bay floor, etc. are precisely known and easily comparable with the target (Weisberg, 2013). Likewise, model results obtained from a scale model are relatively straightforward since we can know, even if sometimes with difficulty, why a particular result holds: we can trace the way in which the model is manipulated. In the Phillips machine, even if not quite a scale model, we know that movements in output that are drawn on the sheet of paper are the result of the water levels in the tanks of consumption, investment and government expenditure. In addition, the representational relation in which the model stands with respect to its target is usually easier to grasp. The floor of the Delta model represents the floor of the actual bay and the amount of water in the investment tank in the Phillips machine represents the stock of investment of the UK economy.

Other non-physical kinds of models, by contrast, do pose ontological conundrums. Firstly, they are not tangible; only in our heads. A question that arises is therefore what kind of entities these models are. Several attempts have been made in the literature to answer this question. Some commentators have argued that models are set-theoretic structures; others have suggested that models are equations; and, more recently, it has been suggested that they are fictional entities very similar to novels or films (Frigg & Hartmann, 2009). Let me just add that all of the accounts offered so far about the kind of entities models are face important objections. Another problem for this literature is that, regardless of the kind of entity that theoretical models are, the fact that they are abstract, poses questions about the implications for their manipulability; a feature often considered essential for learning about the model and thereby about the world (Morgan, 1999, 2012, Chapter 1; Morrison & Morgan, 1999). Furthermore, considering the different kinds of non-physical models, for instance mathematical models, simulations, or thought experiments, another question that arises is whether there are features that they all share and whether they share them with physical models as well.

Surely some of the questions with which this literature deals are not strictly metaphysical. As I suggested above, some commentators explore what the epistemological implications would be if models were one kind of entity or another. Godfrey-Smith (2009), for instance, tries to identify the challenges (some of which are epistemological) of squaring the ontology that model users implicitly attribute to their models, the "folk ontology", with a philosophically sound "external" metaphysics. The point that is important for us here, though, is that regardless of the nature we end up attributing to models, an undeniable fact is that these allegedly mysterious entities are used by modellers to make inferences about the world that often prove to be reliable. Some seem to believe that identifying the "right" metaphysical characterisation of models will allow them to shed light on why these entities allow modellers to make correct inferences about the world, but arguably, an accurate answer to this question is much more likely to come from cognitive science, with respect to the way in which humans use these devices as a sort of extended cognition. For this reason, I leave the ontology of models aside and move on to the other two branches, in which there might be aspects more relevant for my interests in model failure.

3.2. Semantics of modelling

The literature on the semantics of models can be identified mainly with the attempt to provide clues with respect to the kind of relationship that obtains between models and their targets. A general assumption of this literature is that the clue to understanding why it is possible to learn about the world from models that are false (or that, in general, misrepresent their targets) is to be found in the representational relation models stand with respect to their targets. Philosophers have tackled different aspects of this relationship and different views have been offered about what precisely constitutes this relationship. The general aim of this literature can be summarised as attempting to provide a theory of scientific representation.

Based on the extant literature on scientific representation, Frigg & Nguyen (2016) have compiled a set of the minimum requirements that a theory of scientific representation ought to have. They rely on the different accounts offered and the objections raised to them to propose five specific issues that any general theory of scientific representation has to be able to respond to.

1) The Representational Demarcation Problem.

This requirement is that a position must be taken with respect to whether, and if so, how, scientific representations are different from other kinds of representations. Most philosophers have endorsed this position. However, the requirement arises mostly because some philosophers, in particular Callender & Cohen (2006), have argued that there's nothing special about scientific representation; they give precedence to representation as something that goes on in the mind and that therefore belongs to the domain of philosophy of mind. Scientific representation is derivative of this primitive form. The requirement is thus that if there's something special about scientific explanation, this needs to be made explicit in a way that demarcates it from other kinds of representation.

2) What counts as a scientific representation?

Depending on the answer that has been given to the Representational Demarcation Problem, the analyst must also provide an answer to either of two further questions. Those who demarcate scientific from non-scientific representations have to provide an answer to the Scientific Representation Problem, which can be summarised as 'What fills the blank in "S is a scientific representation of T iff ___"?'. Those who reject the Representational Demarcation Problem have to address the Epistemic Representation Problem, which addresses the question of what constitutes a representation in the cases in which it's possible to learn about a specific target, indirectly, by means of the object (model) doing the representation. This is summarised as what fills the blank in "S is an epistemic representation of T iff ___".

At the same time, the answers provided to this question need to fulfil five requirements of adequacy. First, that the representation allows for surrogate reasoning. That is, that the representation allows for generating hypotheses about the target system. The representation is used as a stand-in for the target. Second, that there's room for misrepresentation. That is that the proposed criterion allows for the distinction between an inaccurate representation and a non-representation. My drawing of a purple sun with mountains should be able to be regarded as a misrepresentation of a sunny day at noon, rather than as a failure to represent the sun at all. Third, that there's the possibility for representation regardless of whether there is a specific target (target-less models). Many models are not built with a concrete target in mind, sometimes they are meant to capture a generic mechanism, but nothing concrete in the world. Target-less models should still be able to represent. Fourth, that there is a sense of directionality—representation is a one-way relationship. My drawing of the sun represents the sun, but not the other way around. Finally, this answer needs to be explicit about how this theory of representation is reflected in the mathematics often used in models.

3) The Problem of Style.

This problem addresses the need to acknowledge and accommodate that there are many ways in which a single target can be represented. So, a building can be represented by a 3D render, a 2D blueprint, or a physical scale model. This suggests that there are different representational styles and a theory of scientific representation should be able to account for this possibility—and not, for instance, take such a building as a different target or a different representation altogether. Whether it's possible to determine a priori all possible representational styles is an open question, but the theory should be able to distinguish that a target may be represented in different styles.

4) Standards of Accuracy.

How do we identify what constitutes an accurate representation? The answer offered to the problem of scientific/epistemic representation has to be able to provide a standard of accuracy. It is not sufficient that such an answer allows us to distinguish between inaccurate representation and non-representation. A certain standard of accuracy is needed, which allows us to distinguish between two models that represent a target in terms of which one is a more accurate representation.

5) The Problem of Ontology.

Any answer provided above will have to say something about the kind of objects that serve as representations. For physical objects there's no conundrum, as explained above in the section on ontology. But for other kinds of models—e.g. mathematical—this is more difficult to establish. This requirement seeks to have clarity about metaphysical commitments. For instance, it's important to determine whether the entity representing something is indeed in capacity to perform a representational function.

Whether a representation is regarded as scientific or not is not something that determines whether a model fails or not. The models under investigation here are already considered scientific. The attempt to demarcate scientific representation seems thus orthogonal to the discussion of failure. This excludes requirements one and two as potential criteria that might help identify model failure. Style of representation and the problem of ontology are also not very helpful: there appear to be both successful and unsuccessful models in different representation styles and ontologies. That is, a model is not necessarily more likely to fail because it represents in a particular way or is physical or mathematical. And, even if this were the case, since most models in economics are mathematical, we're interested anyway in more fine-grained criteria. This excludes requirements three and five as potential leads for the identification of failure and leaves only requirement four for consideration.

The fourth requirement states that a theory of scientific representation has to formulate standards of accuracy. This means that any account that tells us in virtue of what a model represents its target will also have to say something about the extent to which such a representation is accurate. So, if an account states that a model represents its target in virtue of being similar—e.g. Giere (2004)—the standard of accuracy will be able to tell us how similar a model is to its target. We could thus think of a criterion for model failure such as "less-accurate-than-intended" or something along these lines. Likewise, different degrees of accuracy may allow us to compare models and determine whether one represents its target more accurately than another, which might also be helpful in identifying model failure. The problem is, however, that such a requirement doesn't say anything about how accuracy of representation maps into a measure of reliability in terms of say, epistemic import or predictive accuracy. To be sure, to demand that one's theory perfectly maps accuracy of representation into reliability (of prediction, explanation etc.) would be a tall order. But my concern is that if it is not possible to track degrees of accuracy in representation with any epistemic import, it is unclear why this is a requirement for a theory of scientific representation.

Interest in scientific representation as a special category comes from the belief that there is something special about scientific representation that makes it unlike other kinds of representation. In fact, Frigg & Nguyen (2016) suggest as much in their discussion of minimum requirements for a theory of scientific representation. With respect to the first requirement of adequacy, namely, that a theory of scientific representation allows for surrogate reasoning, they argue that this requirement helps to distinguish scientific representation from others such as lexicographical—e.g. preferences or numbers—or indexical—e.g. smoke[5]—representation, which do not allow for surrogate reasoning. Yet, they also argue that surrogate reasoning is insufficient because it allows for other non-scientific kinds of representation: it "does not constrain answers sufficiently because any account of representation that fills the blank in a way that satisfies the surrogate reasoning condition will almost invariably also cover other kinds of representations". They use as example a picture taken by a traffic camera to detect those who speed. The picture allows for surrogate reasoning since it allows its users to make inferences about the speed of cars and charge fines accordingly. This, however, is a case they do not want to include in the category of scientific representation. So, scientific representation seems to be about something more than just being able to make inferences about the target, even if Frigg & Nguyen do not specify what precisely this is. Presumably this has to do with making

[5] See Peirce's indexical signs (Atkin, 2013).

accurate inferences that *add to our understanding* of the world or with affording us some kind of epistemic benefit that we in general value. If this is the case, then such a theory should be able to say how standards of accuracy translate into this kind of (special) epistemic benefit. At the very least, it should make explicit *why* a measure of representational accuracy is something that should be required of a theory of scientific representation. Otherwise, an explanation needs to be provided with regard to what makes *scientific representation* worth of attention as a different kind of representation. Surely it can't be just the kind of representation that is typically done by modellers within the confines of academic institutions.

3.2.1. Weisberg's account

Among the different views of scientific representation that are on offer in the literature, Michael Weisberg (2013) has provided one in which the relationship between a model and its target is one of similarity. Ronald Giere (1988, 2004, 2006) and Paul Teller (2001) have also argued that similarity is what accounts for the model-target relationship. Weisberg, however, is the only one who has tried to cash out this relationship formally. According to Weisberg, this account makes possible to measure how similar a model is to its target. More importantly though, it allows a scientist to compare models (with respect to how similar they are to a common target) and to compare the actual similarity of a model with the similarity expected to be accomplished. Given that it renders the degree to which a model is similar to its target, it is helpful to determine whether such a metric (rather than the more abstract notions of similarity) can help with identifying similarity (or lack of it) as a potential source of failure. seems to be a metric that, in principle, might allow us to measure model failure. But, as I said above, unless we have some idea of how accuracy of representation tracks epistemic import, it is unclear how this would help. Let me present Weisberg's formal account and see whether Weisberg's account deals with this aspect, and if so how. Then we can see whether there's something that might be said about model failure.

Weisberg (2013) builds on a method used by Amos Tversky to capture judgements of similarity or dissimilarity made by experimental subjects. Tversky's contrast account of similarity says that the similarity of two objects *a* and *b* depends on the features they share and those they do not. The former count *towards* the measure of similarity whereas the latter *against* it. Formally, the basic idea is expressed as follows. Δ is a set of features that can be qualitative or quantitative. For two objects *a* and *b*, A is the set of features in Δ

possessed by *a* and *B* is the set of features in Δ possessed by *b*. There's also a weighing function $f(\bullet)$, which is defined over the power set of $\Delta(\wp\Delta)$. Similarity of *a* to *b* is given by the following equation:

$$S(a, b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$$

(8.3)[6]

Where θ, α, β are term weights. Weisberg uses this idea to provide an account of the model-world relationship. Then he goes on to modify the original form in order to accommodate models. Weisberg modifies the form of the equation—making it the ratio of the similarities and differences, rather than the difference—and distinguishes between kinds of features, namely attributes, *a*, and mechanisms, *m*. The idea here is to distinguish the properties and patterns of a system from the mechanisms that generate them, as scientists may be interested in being able to distinguish between representing properties or patterns of a system and the mechanisms that bring them about. The following equation is the one he starts with to make his analysis (Weisberg, p. 147):

$$S\,(m,t) =$$

$$\frac{|M_a \cap T_a| + |M_m \cap T_m|}{|M_a \cap T_a| + |M_m \cap T_m| + |M_a - T_a| + |M_m - T_m| + |T_a - M_a| + |T_m - M_m|}$$

(8.8)

M_a and T_a would therefore be the set of attributes of model *m* and target t, respectively, that are in Δ. Weisberg refers to this equation as the core of his *weighted feature-matching account* of model-world relations. Here the simplest possible weighting function has been assumed, where each element in Δ is weighted equally and therefore each term in the similarity equation takes the value of its cardinality. Weight parameters are all assumed to be equal and are therefore dropped. *S* is a value between 0 and 1.

[6] Weisberg's nomenclature has been preserved for easy referencing.

Naturally, in order to obtain a measure of similarity, it is necessary to determine what goes into Δ, the function f, and the weight parameters, $\theta, \rho, \alpha, \beta, \gamma, and\ \delta$ so the equation that results in such a metric is the following:

$$S(m,t) =$$

$$\frac{\theta f(M_a \cap T_a) + \rho f(M_m \cap T_m)}{\theta f(M_a \cap T_a) + \rho f(M_m \cap T_m) + \alpha f(M_a - T_a) + \beta f(M_m - T_m) + \gamma f(T_a - M_a) + \delta d(T_m - M_m)}$$

(8.10)

According to Weisberg, these elements should be determined as follows. The scope of the model is what determines what actually goes into Δ. The scope consists of the aspects of the target that are intended to be represented by the model and are therefore determined on a case by case basis. The weighting parameters are determined according to the modelling goals of the modeller. The modelling goals that Weisberg discusses are obtained based on different kinds of modelling, namely hyper-accurate modelling, how-possibly modelling, minimal modelling and mechanistic modelling, which are instances of the three different modelling activities that Weisberg (2007, 2013, Chapter 6) has previously identified. For instance, he suggests that if a modeller is interested in hyper-accurate modelling, the theorist would want the model to have all the features of the target $(M \cap T)$ and not to have any distortions $(M - T)$ or approximations $(T - M)$ [7]. With how-possibly modelling, the idea is to find a possible mechanism that recreates a set of properties or patterns, in which case the modeller tries $(M_a \cap T_a)$ to have a high value of similarity whereas $(M_m \cap T_m)$ should have low value. It's the properties or patterns that need to be similar and not the mechanism through which these are reproduced. Weisberg thus defines the goal of how-possibly modelling (again in the simplest possible version) as:

$$\frac{|M_a \cap T_a|}{|M_a \cap T_a| + |M_m - T_m|} \rightarrow 1$$

[7] It is unclear why Weisberg considers hyper-accurate modelling as one of the goals. If such kind of modelling were possible, the question arises why a modeller would want to engage in modelling in the first place instead of experimenting in the target directly––or whether this activity would still be regarded as modelling.

Finally, with respect to the weighting function, which determines the relative importance that each element and combinations of elements in Δ have, Weisberg recognises that the function that would satisfy equation 8.10 above would demand too much from scientists. It demands that the function $f(\bullet)$ be defined over $\wp(\Delta_a) \cup \wp(\Delta_m)$, which means that a modeller would have to be able to express such a function. Instead, he suggests that because modellers usually know which features of Δ are more important, more weight can be given to the *special features* subset of Δ, whereas others continue to return their cardinality. Background theory should be able to help modellers determine which are the special features.

As I mentioned above, this account returns a metric of the similarity between a model and its target. This is certainly an advantage over other commentators who have defended the similarity account in non-formal terms. The question that emerges, though, is how is this metric useful to assess the reliability of the inferences that are drawn about the targets of our models. Is it supposed to guide our judgement about how good a model is? When Weisberg defends his account of model representation in terms of similarity, as opposed to other unsuccessful, model-theoretic accounts[8], he indicates his agreement with commentators such as Giere or Cartwright, who "have argued that *successful* models are *similar* to their targets" (2013, p. 142, emphasis added). Presumably, therefore, a formal account of similarity should help us to understand how similarity explains success. However, when Weisberg discusses the uses of his account, he seems to have other goals in mind. He suggests his account is useful to compare models with respect to how similar they are to a target and, particularly, for scientists to be able to measure how far they are from their modelling goals. The point of reference here is thus not necessarily the target, but any modelling goal such as mechanistic modelling. This is how Weisberg puts the issue (2013, p. 151):

> It is traditional to say that the model-world relation is the relationship in virtue of which studying a model can tell us something about the nature of a target system. But at the same time, scientists are often interested in comparing the relationship that a model actually holds to the world to the one they are interested in achieving between the model and the world […] weighted feature matching allows scientists to assess how close they have come to meeting their goals. It also

[8] Isomorphism, homomorphism and partial isomorphism are such model-theoretic accounts, which are also formal and thus comparable to Weisberg's formal treatment of similarity. They all have important drawbacks and are therefore are regarded as unsuccessful (see Winther (2016) for details).

> recognises that different goals can require different kinds of similarity relations, or at least the emphasis of different kinds of features.

Weisberg's idea seems to be that if a modeller's aim is, say, to build a how-possibly model that accurately represents properties but not mechanisms—as I discussed above—they might want to have a metric of the similarities and dissimilarities between model and target. While this is perhaps interesting for a scientist to know, it is unclear what the added-value is. First, Mary Morgan (1999, 2012) has argued that a crucial feature of models that allows modellers to learn about the world is that models can be manipulated and therefore reasoned with. It is this manipulation, "understanding the world in the model", Morgan argues, which allows scientists to gauge how close they are to meeting their goals. This view is consistent with Rodrik's (2015) account of the modelling practice in economics, in which he suggests that craft and experience are important in deciding what the best model is for a particular situation. Second, it is unlikely that such a metric would replace the scientists' judgement of how successful they have been. I think it is more likely that a modeller would make this judgement simply based on the model result they get. If the result is quantitative, the numerical distance between actual results and an expected interval or the sign of the result would offer a more direct metric of how far a modeller is from meeting their goals. If the result is not quantitative, background theory or knowledge of the phenomenon under investigation should be able to do the job. In general, 'the feeling' that modellers may gain with experience as well as whatever measure of validity they use for their models are more likely to inform scientists about how far they are in their goals than a metric of similarity. The reason is, I think, because there isn't a way to relate degrees of similarity to epistemic import. Put differently, even if the account makes it possible to compare two or more models and determine which of them is more similar to a common target, given a specific modelling goal, it is unclear whether it would be possible to judge the most similar model also as the most successful or the one with the most epistemic import, predictive success, or any other cognitive goal.

It seems to me that the problem with Weisberg's account is that it *presupposes*—rather than demonstrates—that there is a straightforward relationship between representation and epistemic success. By straightforward I don't mean a linear relationship; it is clear that Weisberg acknowledges that there may be different levels of similarity that a modeller may want to achieve, depending on the kind of modelling they are interested in. I mean that it is presupposed that similarity and degrees thereof determine model success even though it is unclear how. To be able to have a metric of similarity doesn't help much in that respect.

Weisberg follows the classical discussions about model-world relationships in which such relation is a model-theoretic analogue to truth. In turn, this presupposes that the less idealised—or the more accurate a model representation is—the more true things such a model can tell us about the world. While this idea may be attractive intuitively, such an assumption should be made based on our experience of how models are used and the purposes they are supposed to fulfil. That is, this should be a consequence of our explorations of different models and their successes and failures and not something we merely presuppose. Particularly because there are cases in which a purpose like empirical success is advanced by fewer shared properties between the model and the target. As Northcott (2017) and Reiss (2007, 2008, Chapter 8) have argued, and I will discuss below, sometimes simpler models, those that are more adaptable, perform better than causally (and thus representationally) accurate models. This demonstrates that it's not always the case that better representation (or more similarity) translates into more reliability or more epistemic import. Therefore, a clear specification of how accurate representation tracks epistemic import would be needed. But because it's clear that such a might be impossible to come by[9], it is unclear, at best, how such an account in terms of similarity is helpful to understand the modelling practice and thereby to assess the ways in which models succeed and fail.

3.3: Epistemology of modelling

The epistemology of models is the most relevant of the three branches for the concern that I have. Literature that is interested in understanding what and how model users can learn about the world by using models determines their function and thereby their limitations. In principle, therefore, this should allow us to say something about model failure. If a model is used in a way that exceeds its limitations, we can expect it to fail. While this kind of model failure is not the only one—a model might fail even when it's 'performing' within its limits—it still allows us to cover some ground in establishing the possible sources of model failure. In consequence, what I shall be looking for in the survey of this literature is the criteria that are used to assess models' functions.

[9] A specification of how accurate representation tracks epistemic import and therefore reliability would amount to solving the problem of induction.

There are two major features of this literature: first, it has focussed its attention on theoretical models. This is not surprising. If there is a mystery about the epistemology of models, it is why theoretical models, in contrast with say, statistical models, are capable of telling us true things about the world despite their numerous false assumptions. Empirical models are, arguably, not mysterious in this way. Second, the literature has focussed on establishing whether models can be explanatory. This is not a trivial question, considering that it is generally thought that only true accounts explain. If models are false, at least to some extent, a question arises whether it is possible that they explain (Reiss, 2012). And, in the last years, an alternative to the explanatory role has caught the attention of commentators, leading to explorations of whether models might provide understanding—instead of or besides explanation.

To a certain extent, focus on this literature is problematic for my purposes. According to number of philosophical accounts of what scientific explanations are, economic models do not fulfil the criteria for any of them (Reiss, 2012). This means that, according to these philosophical accounts, economic models do not provide scientific explanations. If we were to use these criteria for our purposes, we would conclude that economic models are all failures—they aren't explanatory according to the philosophical accounts. This is, of course, unsatisfactory: one of the reasons why there is so much talk about the failure of economic models lately, is because presumably they have succeeded in other occasions. If we turn to the notion of understanding, which has also been explored with respect to economic models, the problem is that understanding is not a dichotomous concept like failure or success. Understanding, reflecting the different depths of understanding that an agent might have about a subject, comes in degrees Elgin (2007) I'll thus concede at the outset that it is unclear how the criteria that are explored in this literature might be mapped into notions of failure and success. Still, considering that recent literature on economic models has focussed mostly on their epistemology, it is still worth exploring the literature in order to see how the criteria used there might be useful to think about model failure.

The current debate with respect to whether models explain or afford understanding is as follows. Most of the commentators have defended models, arguing that they are epistemologically useful. These I'll call the enthusiasts. Others, the sceptics, claim that models are only useful heuristically; models are useful for discovering new hypotheses or as heuristics to formulate causal claims, but by themselves models do not say anything about the world. Let me discuss the sceptics first and then move on to the enthusiasts.

3.3.1. The sceptics

Some commentators have, for different reasons, downplayed the epistemic import of models, attributing them a very minimal heuristic role. Here one finds philosophers like Dan Hausman (1992, Chapter 4), who has claimed that models are mostly conceptual explorations. He suggests that it's only hypotheses that state that a target system satisfies a class of the model's assumptions that are related to the world; not models per se. Another commentator is Anna Alexandrova (2008), who, in an attempt to offer an account of models that accommodates the ways in which models are used in market design, has suggested that models be regarded as 'open formulae'. Only later, via experiments or other methods, these open formulae can be specified further and therefore 'closed' to make causal claims. Even stronger, Alexandrova & Northcott (2013) claim that there's no evidence that supports the view that models 'isolate' causal mechanisms, a view long defended by Uskali Mäki (1992, 2005, 2009), or that they state 'capacities', as argued by the early Nancy Cartwright (1989). In particular, Alexandrova & Northcott (2013) argue that there is no evidence that economic models can do the same kind of things that we could do if models were capable to isolate or state capacities; if they did, models would have a better predictive record than they actually have.

The bottom line of this literature is that models don't have any epistemic function except for their heuristic role, helping modellers in their process of discovering hypotheses. If this is the case, then this literature appears to be of little relevance for my purposes: if models are only heuristic devices and they don't make any claims about the world, then they can't fail (except in the limited sense of not being heuristically useful; but this is not my concern here). With this in mind, there are two positions that advocates of this view might take with respect to model failure. On the one hand, they may claim that model failure, at least in the way in which I mean it here, namely with evident, empirical consequences (recall aviation accidents), can only be caused by the misuse of models. That is, by using a model that only has a heuristic role for making direct claims about the world. In this case, advocates of this view would have to regard model failure as neglect on the part of model users or mere stupidity for using models for different purposes than their heuristic role. If this is the stance, the response to those who claim that the crisis was caused by models would be that it is model users who are at fault. There wouldn't be much more for me to look for and this would be the end of the chapter. On the other hand, sceptics could claim that models, in the way they are used, say to carry out policy, are typically complemented with other kind of knowledge that is being left out of the analysis of the philosophical modelling literature and this is where failure may come from. In this case, if we want to find the sources of failure, then our accounts of models need to account for the way in which models interact with other sources of information. The point is thus that those who defend the view that models are purely heuristic devices also need to explain the source of model failure.

Alexandrova & Northcott (2009, 2013), who defend their view based on models used for market design, argue indeed that there are more elements at play, like experiments, in the whole process of designing and carrying out a successful auction. So, they hold the latter view. In fact, they maintain that the approach to understanding the role of models (theory), should be accompanied by how this theory is applied, in the same fashion it is done in engineering. They even make an analogy with Formula One racing in which they suggest that while theoretical knowledge of Newtonian mechanics, material composition of rubber, and so on, are necessary conditions for the success of racing cars—measured in terms of speed and reliability—so are the experimental trials of drivers in private circuits, new chassis in wind tunnels, etc. They thus maintain that the modelling *process* is more complex than just the use of the theoretical model. However, while they maintain that the contribution of theoretical models can only be assessed instrumentally, with respect to the empirical success of the engineering exercise, they remain silent about whether and how a model can fail or make the whole process fail. To be fair, this has surely not been their purpose; they use the FCC auctions as a success story and highlight how models and experiments worked successfully together. But my claim in this respect is precisely this: that a view of models, regardless of the function that one attributes to them, has to be able to account for why models succeed *and fail*. Alexandrova & Northcott (2013) do offer an error theory of why economists might be mistakenly led to believe that their models explain when they in fact don't, but they are silent with respect to how this may have negative consequences in engineering-type exercises. Furthermore, while they acknowledge the importance of the wider setting in their view of models as open formulae, the philosophical analysis of this larger setting is still missing.

3.3.2. The enthusiasts

In contrast to the sceptics, other commentators have attempted to defend the epistemic import that models have: they highlight, justify, and defend some of the various epistemic roles that models might have, emphasising their advantages or benefits. In the remainder of this section I shall discuss the work of some philosophers of economics whose work represents the two features I mentioned above: they have focussed on the explanatory role of theoretical models. All methodologists I am going to discuss here have analysed the Schelling model of spatial segregation. Economist Bob Sugden is presumably the one who first paid significant attention to Schelling's model in his "Credible Worlds: The Status of Theoretical Models in Economics" (2000), referring to it—and Akerlof's model of lemons—as "the kind of model building to which I aspire" (2000, p. 2). Since then, this is perhaps the model that most attention has received by philosophers of economics and thus arguably the one from which the most comprehensive analysis has been made. It is thus reasonable to expect that these accounts are able to determine whether, and if so why, Schelling's model is particularly special. My take is that if we can identify what are the features that make Schelling's model special, lack of these same features could offer guidelines that could help identify possible model failure. Naturally, analysis of a single model is not enough but it's a start.

Let me quickly describe what Sugden does in his paper. His intention, made explicit at the outset, is to defend the use of theoretical models. He is a theoretical modeller and, even though he is convinced that his models are useful, he says, he's not really sure why. He thus sets himself to understand what theoretical models are capable of offering from an epistemic point of view. In this paper, he analyses Akerlof's famous paper on the market for lemons and Schelling's model of spatial segregation. He says that he has picked these two models for his analysis because he finds them as exemplary theoretical pieces and because they are models that possess features critics usually complain about: a high degree of abstraction and plain unrealisticness. In his analysis, Sugden explores two aspects: first, what the models can possibly be contributing and, second, how one would be justified in inferring that such a contribution applies to the world. With respect to the latter, he suggests that a model is not an abstraction or a simplification of the real world. The modeller creates a parallel world from which they make inductive inferences. These inferences are justified insofar as the parallel world is credible. We do such inductive inferences all the time. About what the models are contributing, he suggests that these models offer a hypothesis that could be considered an explanation, if pursued further.

3.3.2.1. The meta-model: mechanisms and attention

Aydinonat (2007, 2008, Chapters 4, 7) has discussed Schelling's model extensively and has attempted to "explicate the explanatory characteristics of the checkerboard model in order to understand its strengths and weaknesses" (2007, p. 430). The basic idea of Schelling's model is that "non-racist preferences", or a preference for simply not being in the minority defined by a certain threshold, gives rise to segregated neighbourhoods. The model is thus interpreted as providing a mechanism of how individual preferences may give rise to a macro phenomenon—residential segregation[10]. Aydinonat argues that the model provides a "partial potential (theoretical) explanation" of residential segregation. The explanation is partial, Aydinonat argues, following Hempel, because it ignores other factors that may also give rise to the explanandum, like known systematic discrimination by housing developers. It is potential (and not actual), because the explanans are not known to be true; they are only conjectural. Despite the partiality of the model, Aydinonat suggests, the model still contributes epistemically because it can be added to the meta-model or theory (i.e. collection of models) that a scientist may recourse to in searching for an explanation of a concrete case of segregation. The mechanism suggested by the Schelling model contributes to the breadth of the explanatory meta-model, making it applicable to more instances of segregation by offering extra tools by which segregation can be explained. Whereas previous explanations of segregation suggested that the phenomenon was the outcome of explicit segregationist behaviour, the model shows that segregation can be an unintended macro effect of individual preferences and decisions. Such a mechanism does not exclude the presence of more direct segregationist behaviour.

Since Schelling introduced the model, many different specifications have been explored afterwards. Some of these new specifications are purely theoretical, others introduce certain assumptions that are based on empirical findings. The results obtained from these new model specifications are mixed; for some, the "Schelling result" of segregation holds, whereas for others segregation does not occur. Aydinonat suggests that these explorations demonstrate that the hypothesis of the model is interesting and promising, encouraging scientists to explore the model conditions under which segregation occurs.

This practice is not unusual in economics. In his discussion of the functions of models in economics, Rodrik (2015) argues that one of them is to clarify hypotheses and to reveal counterintuitive possibilities and unexpected consequences. Rodrik offers several examples among which is Ricardo's Principle of Comparative Advantage. Intuitively, Rodrik suggests, trade would be considered beneficial for a trading party only when it has

[10] Given how much attention Schelling's model has received, I will simply assume that the model is known to the reader.

an absolute advantage over the other—that is, when one party produces every potentially tradable good more efficiently that the other. Ricardo instead argued that a country benefits from exporting what it produces relatively better and importing what it produces relatively less well. Further model specifications, "tinkering with the model by theorists over generations" showed that this result did not depend on certain specifications such as the number of commodities or the number of countries trading.

In this respect, Aydinonat's analysis seems to capture economic practice well. However, the philosophical analysis he offers, namely his attempt to "explicate the explanatory characteristics of the model in order to understand its strengths and weaknesses" is, in my view, less successful. I understand Aydinonat's main objective as to offer an analysis of the features of the model that account for both its epistemic benefits and its limitations. The analysis Aydinonat offers, though, is rather sanguine about the model strengths. Aydinonat's analysis seems to presuppose that strengths and weaknesses are somehow independent. This, I shall suggest, raises questions about how potential the explanation actually is. I shall also suggest that whether the explanation is partial or not is a question that is better resolved empirically and not philosophically. Finally, I will argue that the only feature that Aydinonat offers as distinctive of Schelling's model that could explain its success is philosophically unsatisfactory.

To suggest that the Schelling model offers a potential explanation of segregation, Aydinonat makes the distinction between potential and actual explanations. As mentioned above, the latter are those whose explanantia are true. Because Schelling's explanantia are just conjectural or fictional, and thus their truth cannot be guaranteed, Aydinonat argues that Schelling offers just a potential explanation. Aside from this feature, Aydinonat offers a substantive overview of the many different model specifications that scientists have tried. As mentioned above, some of these are theoretical and others are based on empirical findings. For some of them the result of segregation holds whereas for others it doesn't. For instance, Aydinonat (2007) cites a study conducted by Bruch and Mare (2003), which suggests that utility functions that best describe real individuals' preferences, which have the property of continuity, yield lower levels of segregation than the ones suggested by Schelling's model, which uses a threshold function. Similarly, although this isn't one of Schelling's model specifications, Aydinonat cites in a footnote a test of the macro implications of Schelling's model performed by William Easterly (2009)[11], which Easterly claims to contradict Schelling's model results.

These tests with contradictory results—that segregation does not occur—are crucial to determine the set of the conditions under which the Schelling result does hold. Especially the model specifications whose conclusion is that segregation does not occur should be of particular importance because they suggest that the model result does depend on certain specifications. If this set is wide enough—e.g. segregation holds regardless of the number of different players, or a very high number—it may be possible that the real-world conditions fall within this set. This would in turn lend support to the possibility that the Schelling mechanism holds in the world and that it may thus be a potential explanation of segregation, as Aydinonat claims. We wouldn't know whether the conditions are true of the world, but we would be able to tell that the world conditions are likely to fall within the set of conditions under which the result of segregation holds. However, if the set of conditions under which the model holds is too constrained, in such a way that any world condition is very much unlike those of the model, the hypothesis loses force as being possible in the world. In other words, an explanation is only a potential one if the explanantia cannot be guaranteed to be true, but they are not known to be false. I don't want to suggest that in Schelling's case they are false; it is not my task to determine this. My point is that this sort of appraisal is necessary for the claim that the model may be applicable to the world, and would therefore be what Aydinonat has to do, if he is to claim that the mechanism described by the model is a potential explanation. However, while Aydinonat acknowledges that segregation occurs under a specific set of conditions (and not others), his analysis doesn't contemplate that the model's mechanism is a potential explanation contingent on the set of conditions under which it holds. Instead, he claims that the model *already* made an epistemic contribution: the original hypothesis suggested by Schelling's model has added an additional potential explanation. That is, the model was causally responsible for an additional potential explanation that is at the disposal of scientists and policy makers to use whenever the explanation may be considered suitable. Furthermore, he claims that the "explorations" or the multiple model specifications are *refinements* to the meta-model of explanation of segregation, suggesting that the exploration can only be beneficial because it expands the meta-model. "If the argument that the checkerboard model contributes to a meta-model of residential segregation is accepted,

[11] Aydinonat (2007) cites a previous version of the paper, which was not yet published then.

explorations of the checkerboard model may be considered as future refinements and expansions of this meta-model. Hence, Schelling's initial hypothesis and his checkerboard model helped researchers to expand their conceptual toolbox to include more explanatory factors" (2007, p. 444). The explorations, however, could suggest that the hypothesis be removed from the meta-model altogether if the conditions under which segregation occurs are too restrictive. Aydinonat considers that the explorations are only improvements, but such a conclusion is not warranted. The model's benefits depend on the model's limitations, which are being tested by the explorations Aydinonat describes.

Leaving the issue just mentioned aside, so, assuming that Schelling's model indeed offers a potential explanation, one may also question its partiality. Recall that a partial explanation is one whose explanans do not fully account for the explanandum. In this respect, Aydinonat argues that the Schelling mechanism that leads to segregation (from micro motives to macro behaviour) could be operating simultaneously with other more direct mechanisms of segregation like outright racism by real estate developers, so the model may actually be capturing a real mechanism, while not excluding others. One may ask, however, whether such a contribution is indeed partial if it has never been able to actually explain, partially, any real phenomena. To be sure, Aydinonat doesn't say that it hasn't, but he also doesn't say that it has. Aydinonat's assumption seems to be that, once the model's hypothesis has made it to the "repository of possible explanations", it can already be counted as both potential and partial. This is surely possible. But considering that there are reasons to think that the other more direct mechanisms are in place, like outright racism, economic and racial disparity, gentrification and other socio-economic and demographic dynamics, one may question what the "marginal contribution" of the Schelling result is to the explanatory meta-model of segregation.

Some may rebut, by arguing that even if such a marginal contribution were infinitesimal, it would still have to be ruled out before an alternative hypothesis (say, outright racism) can be accepted. At the very least, they might add, the live possibility of the Schelling mechanism operating, lowers the warrant of alternative hypotheses. Ylikoski & Aydinonat (2014) make a similar point when they claim that the search for causal explanation has the structure of eliminative induction. Reiss (2015), in the context of his pragmatist theory of evidence, has argued as much.

Reiss's theory is meant to be of direct relevance for the biomedical and social sciences, and especially for those domains that take randomised controlled trials as the gold standard of evidence. His theory is meant to be for causal claims—such as whether smoking causes lung cancer or whether segregation is caused by 'mild discriminatory preferences'. To explain his theory, Reiss sometimes uses the analogy of TV detectives trying to solve a murder case. The idea here, briefly, is that if we're investigating who murdered Maria, and we learn afterwards that Manuel was in the room when she died, our original hypothesis that Eva was the murderer is less warranted—Manuel is now a suspect too. We thus learn from knowing about Manuel, because we now have more hypotheses that need to be ruled out before we settle on one. The more competing hypotheses are discarded, according to Reiss's theory, the more warranted the original hypothesis is[12]. In the case of the meta-model of explanation, the more complete the set is, the more warrant one of the hypothesis has once the others have been discarded.

The problem with Reiss's analogy, though, is that it doesn't apply to Schelling's model of segregation. In the case in which we want to find out who murdered Maria, we *don't know who did it.* So, any clue about who the culprit might be, learning about Manuel, is indeed helpful. By contrast, in the case of segregation, *we know* there are other causes. We know that segregation is caused by racism (Mahler & Eder, 2016), socioeconomic and demographic conditions (Florida, 2014, 2017) and explicit segregationist policies (Duursma & Heck, 2017), just to name a few. Here the right analogy, if we are to stick to murderers and detectives, would be that we know people are murdered with guns (*x* percent of the time), baseball bats (another *y* percent of the time), by suffocation (another *z* percent of the time), etc. The equivalent to finding whether segregation is caused by non-racist preferences would be something like finding whether people are murdered by being quartered with a nail clipper. So, my point about offering a partial explanation is that being part of the meta-model doesn't secure the model's partial contribution. There are several other explanations that we know account for the effect of segregation. The partiality of the explanation, or its 'marginal contribution', is probably best established empirically by using other methods, such as econometric analysis, which allows us to identify how much non-racist preferences contribute to the effect of segregation. This is perhaps difficult to measure confidently, considering the outcome of segregation is the result of preferences that are not directly observable. But there are interesting methods such as instrumental

[12] Reiss (2015) makes a distinction between direct and indirect support. The former is evidence that directly supports the hypothesis in question. The latter is evidence that is incompatible with alternative hypotheses, lending indirect support to the hypothesis in question by ruling out alternatives. Warrant, which allows for degrees, is gained by having both kinds of support.

variables or differences in differences that combined with ethnographic research could be helpful in answering such a question. Such methods might prove to be more reliable than mere speculation that the mechanism suggested by Schelling may or may not play a role.

The points I've made so far about Aydinonat's analysis of Schelling's model have challenged the model's epistemic contributions. Let me now turn to my final point, which is about a feature of the model that, according to Aydinonat's analysis, seems to be a distinguishing feature. That is, a feature that distinguishes the model from others and that is, at least partially, responsible for the epistemic contributions the model makes. I am referring to the attention the model has received by scientists from different disciplines. According to Aydinonat, the attention, and the many different specifications that have later been explored by social scientists, are a sign of the model's contribution to the meta-model of explanation.

> These studies give us enough evidence that the checkerboard model has received considerable attention and its results and implications have been explored and tested in different ways. These explorations give us reasons to believe that Schelling's insights may be relevant for the real world (2007, p. 443).

It thus seems that it's not the result of the explorations or tests, but rather the fact that these explorations have taken place, that give reasons to Aydinonat to believe that Schelling's model might be applicable to the world. But attention as a criterion of relevance is not satisfactory. The role of the philosopher is precisely to come up with independent criteria for why a model is reliable, useful, etc. To judge the 'goodness' or relevance of the model by the amount of attention is it has received, is to beg the question. If the amount of attention Schelling's model has received is significant relative to other models, part of the philosophical analysis calls for understanding why this is the case. Furthermore, it is remarkable that Aydinonat has chosen the attention and exploration of Schelling's model as a criterion to judge its relevance because exactly the opposite claim has been made lately about macroeconomics and specifically about models with microfoundations. A recurrent criticism of macroeconomics and the causes of the crisis is that macroeconomists gave too much attention to DSGE models, which are too restrictive and can't incorporate real world features such as heterogeneity of agents (Colander et al., 2009; Rodrik, 2015, Chapter 3). Obviously, this doesn't make the critics right, but that attention and focus on a particular model or kind of models is considered as beneficial by some and damning by others is not helpful to understand the model's benefits or weaknesses.

3.3.2.2. Robustness

Others who have contributed to the discussion, providing a comprehensive account of modelling in which some adequacy conditions are discussed that can be useful to look at model failure are Kuorikoski & Ylikoski (2015). They defend a very similar view to that of Aydinonat (2007) and define the modelling exercise as a matter of extended cognition. Here models are not mysterious in their capacity to explain despite their many falsehoods; they are simply external devices that help us, cognitively limited beings, to make inferences about a phenomenon of interest. Like Ylikoski & Aydinonat (2014) they argue models afford understanding insofar as they allow us to make correct what-if inferences[13]. These correct inferences are possible by obtaining knowledge of dependencies—causal or other types, they call themselves pluralists in this respect. Explanations are answers to contrastive questions that understanding of a phenomenon affords. Thus, whereas understanding is about the ability to use the knowledge of dependencies to make what-if inferences, explanations are answers to specific contrastive questions.

They also argue that models are representational to the extent that they allow model users to make correct inferences about a target. So, they basically turn the issue of representation on its head: while many philosophers claim that models afford epistemic import insofar as they represent their targets in the right way, Kuorikoski & Ylikoski (2015) claim that it is the amount and quality of what-if inferences that we can make with the use of a model that determine how well a model represents its target. They dismiss the literature on representation I discussed above quite strongly, suggesting that, if there is indeed something interesting to be discovered in the way models represent their targets, the answer is surely not to be found in philosophy. In other words, this is not a philosophical problem, if it is a problem at all.

In their account, therefore, the crux of the matter is the number of inferences that are made based on the model and their precision. This depends on the truth of the assumptions of the model and the reliability of the inferences that can be drawn. These two issues can be tackled by means of conducting robustness analysis, which involves examining whether the same model result holds in different specifications of a model in which the tractability assumptions are varied. If the result holds, it is said to be robust, and

[13] Aydinonat (2007) hints at some of the explorations he discusses as being what-if questions. But suggests 'what-if' kinds of explorations are just a kind and that there are many ways in which modellers can examine the plausibility of a hypothesis.

suggests that it is driven by an identified causal mechanism and not by the typically false tractability assumptions used in the model.

Kuorikoski & Ylikoski (2015) argue that the reliability of the explanations derived from models hinges on the success of the robustness analysis conducted. This thus makes robustness—or lack thereof—a candidate for identifying potential model failure: if a model (or cluster of models) fails the robustness test, this could be a sign that the mechanism identified is not present in the world and that, if we were to make inferences based on this model, they may be incorrect. Lack of robustness would allow us, therefore, to *prevent* model failure.

There are, however, at least three problems with using robustness as a criterion to identify possible sources of failure. The first is that the robustness of a result is not a necessary condition for at least some epistemic purposes. This is something that Kuorikoski & Ylikoski (2015) recognise with respect to understanding. A non-robust result—precisely because it's not robust—may suggest that there is a particular assumption that the model does depend on[14], as in the discussion of Aydinonat's account above. So, they suggest we gain some understanding by learning this fact. Also for prediction, considering that accurate representation is not a necessary condition for successful prediction, robustness also doesn't seem to be necessary. After all, robustness allegedly tracks true causal mechanisms, which are not necessary for accurate prediction. Admittedly though, that robustness is not necessary for understanding or for prediction doesn't imply that it may not be necessary for another purpose. But then, if it were to be a necessary condition, it would be, at the very best, purpose-specific.

The second problem is that it's also not a sufficient condition, partly because of the practicalities of conducting robustness analysis. Lisciandra (2017) has argued that there is a difference between the ideal robustness test and the ones that can actually be performed. Similar claims have been made by Odenbaugh & Alexandrova (2011) and Reiss (2012). So, even if we have a robust theorem or a successful robustness test, this is not a method of confirmation. Kuorikoski, Lehtinen, & Marchionni (2012) admit as much.

Finally, the extent to which these tests are actually performed by modellers is unclear. This is an issue about which Kuorikoski, Lehtinen, & Marchionni (2010) are ambiguous at best,

[14] It should be noted that there are more kinds of robustness than the one referred to by Kuorikoski & Ylikoski (2015).

and contradictory at worst. In their paper, they claim that, "theoretical economic modelling is to be understood as collective derivational robustness analysis" (2010, p. 549). Likewise, the title of their paper is "Economic Modelling as Robustness Analysis". Yet, in a remark about Nancy Cartwright's (2005) suggestion that robustness analysis is rarely performed, they say: "Whether she was right in claiming that this is not sufficiently done in economics is a question we cannot fully address here. Our illustration provides an instance in which tractability assumptions are also modified, and this is not an isolated example" (2010, p. 548). It is thus unclear why they claim that theoretical modelling in economics should be understood as derivational robustness analysis, if they can't simultaneously assert that this is a common practice. That the example they provide is not an isolated example is surely not the same as being a common practice. And, as I have discussed earlier in this dissertation, there are other reasons for why economists model a particular phenomenon using different assumptions than just carrying out robustness tests.

In conclusion, robustness analysis, if carried out, is a practice that may give us more confidence in the existence of an identified causal mechanism. In this respect, insofar as we're able to make what-if inferences about causal dependencies, we may be able to gain understanding about a particular phenomenon. However, robustness analysis is not, at least by itself, a practice that is likely to help identify features of the modelling practice that may allow us to identify possible sources of model failure.

3.3.2.3. Learning

Grüne-Yanoff is another scholar who has analysed Schelling's model, in the attempt to understand models' epistemic contributions. He has been particularly interested in defending the epistemic relevance of models that he regards as non-representational. That is, he has argued that the assessment of models can't be based on the representational capacities of models because not all models are meant to represent specific targets (Grüne-Yanoff, 2009). He takes sides with Sugden, by suggesting that models refer sometimes only to possible processes, background conditions or possible phenomena or properties, without attempting to represent anything specific in the real world. In his paper "Appraising non-representational models", Grüne-Yanoff's (2013) specific concern is that, insofar as the representational capacity of a model is considered a necessary condition for learning about the world and the appraisal of models relies on this representational capacity, models that do not represent specific aspects of the real world can't be properly appraised. The few philosophers who have looked at these models, Grüne-Yanoff claims, have considered them to play, at best, a heuristic role, as some commentators (e.g. Hausman, 1992) have claimed. This heuristic role poses two additional issues, according to Grüne-Yanoff: If models are thought to play at most a heuristic role the criteria to assess their success is "unclear in the extreme". Furthermore, this equates modelling with other activities that Grüne-Yanoff considers can't be "rationally accounted for" like taking a walk or reading the newspaper in order for scientists to gain inspiration. He finds this situation unsatisfactory and thus attempts to repair it. His ultimate aim is to argue that it's possible to learn from models, even if they lack an established representational relation to real-world targets (p. 851).

Grüne-Yanoff's strategy is first to define what he takes learning to be, namely a change in confidence in a particular hypothesis about the world that is justified by reference to the model. Then, using the literature on how-possibly explanations, he suggests five opportunities in which learning may be achieved. Finally, he presents five models from different areas of science, each corresponding to one of the five opportunities of learning discussed before. Schelling's model is one of the models he discusses and argues that the model makes us learn by "affecting impossibility claims". This means that the model justifies changing one's confidence in the hypothesis that racist preferences are a necessary cause of segregation. Because the model shows a mechanism in which non-racist preferences lead to segregation, a model user should be less confident in the claim that racist preferences are a necessary cause of segregation.

I think Grüne-Yanoff is correct in his assessment that non-representational models play a significant role in science and that philosophers have not much considered what their roles are or have indeed lumped them in the category of "heuristic". His claims are valid and important. But, in my view, he confuses the defence of a certain type of models for having an important epistemic role, with the assessment of such models. One thing is to identify their potential epistemic import and defend it. Another is to assess them as "good" or "bad" models. I think he conflates the two. Another way to put this is that while non-representational models might make us learn, this doesn't automatically make them good models. There might be "bad" non-representational models.

At the outset of his paper, Grüne-Yanoff claims that assessing representational models for their "goodness" is relatively straightforward given that it is a matter of accurately representing the target. Criteria differ, he suggests, as there are commentators who would claim this is a matter of resemblance (Mäki), isomorphism (van Fraassen), or similarity

(Giere)[15]. Then he suggests that because non-representational models do not represent anything concrete in the world, but only possible entities, processes or properties, the criteria to judge the kind of inferences drawn from the model can't possibly be how well these inferences represent a target. Instead, therefore, he proposes *learning* as his criterion of appraisal (Grüne-Yanoff, 2013, p. 853):

> In the cases discussed here, however, the knowledge processed by the model is different: it contains beliefs about possible entities, processes, or properties, which cannot be obtained by establishing an adequate representation of the model to actual target systems. Thus, adequate representation is not a useful appraisal criterion for such models. Instead, I propose learning as the appropriate criterion.

So, Grüne-Yanoff justifies the use of learning as a criterion of appraisal based on the impossibility of using adequate representation as a criterion for non-representational models. But he doesn't really argue why this is an adequate criterion. In fact, I don't think it can be defended as a criterion for the assessment of a model. There are four reasons for this.

The first thing to note is that learning, at least as it is defined by Grüne-Yanoff, is also something that is typically achieved by representational models, and that is usually desired. A scientific model represents a target generally because there is an interest to learn about that target. It is thus unclear why other criteria would be used for these models, namely adequacy of representation, as Grüne-Yanoff suggests, and not also learning. After all, it is a live possibility that a model adequately represents its target, but that we don't learn anything from it—for instance, because it's something that was already known. Prima facie, we would thus seem to be better-off with learning as a criterion than with adequacy of representation. Grüne-Yanoff would first have to explain why learning is not a criterion used with representational models.

Another reason why learning could not be the criterion used to assess the goodness of representational models, or at least not the *only* criterion to assess a model, is that it is just too coarse. For non-representational models, it is sometimes too coarse and sometimes simply inappropriate. Let me illustrate this with an analogy. I like cooking very much. Every now and then I buy books and subscriptions to magazines with access to some of

[15] As I argued above with respect to Weisberg's similarity, these criteria do not say anything about the "goodness" of a model, if understood as having some kind of epistemic import. It is generally assumed that this is the case but it's not demonstrated.

the recipes and tricks of cooks I really like and from whom I can learn things. That is indeed something I aim at and I judge my purchases by whether I learn new cooking tricks. If learning were my only criterion, though, I wouldn't be able to appreciate the emphasis on nutrition and health of Sarah Britton, or the exquisiteness and sometimes awkwardness of the ingredient combinations that Yotam Ottolenghi makes, or the emphasis on technique that Naomi Pomeroy makes to enhance even the simplest everyday meal. I would judge them all equally because I always learn from them. In fact, I would judge the "Tasty" bird-eye-view (and sometimes disgusting) short films of recipes that circulate nowadays on Facebook equally well, because I learnt an ingredient combination I hadn't thought about. In addition, I wouldn't be able to judge them by how good they are as recipe books: Ottolenghi, for instance, is too UK oriented ingredient-wise and not all the recipes have pictures—which I find very important. "El Mercado", by the Peruvian Rafael Osterling is a beautiful book, but most of the recipes require advance preparation of other more "basic" recipes. I want to be able to judge them by how good they are as recipe books, even if I can learn from all of them. Goodness (of recipe books) for me, hinges on different criteria than just whether they make me learn.

I want to make two points here: one is that, non-representational models in particular, have other purposes than just learning about the world. They may therefore be judged by other non-epistemic criteria such as simplicity, adherence to a particular modelling style, or the proof of an analytical result. Examples of the latter are the proofs of existence and uniqueness of general equilibrium (Arrow and Debreu) and Samuelson's (in)famous Loan-Consumption model[16]. The other point is that, even if we just appraise those models that indeed are capable of making us learn, we should still be able to judge them by how good they are, which requires different criteria. Learning is here too coarse a criterion.

Another aspect that points to the difficulty of suggesting learning, at least as defined by Grüne-Yanoff, as an appraisal criterion of a model is that it is a subjective aspect. Grüne–Yanoff uses the criterion of learning as if it were an objective property of a model to have the capacity to change model users' confidence in a particular hypothesis. However, doxastic attitudes are subjective and depend on the network of propositions

[16] H. Maas (2014, Chapter 7) argues that despite the non-representational character of this model, we're able to learn something about the world from this model. Yet, the model is considered "foundational" by economist Olivier Blanchard, who suggests that models such as the Loan-Consumption model make "deep theoretical points". See Blanchard (2017) for details.

related to the one in question which a subject (an individual or a community) believe to be true. So, it could be the case that a model that provides a true result—say, correctly identifies a causal relation between two factors—but that, because the network of beliefs of the model-user are totally opposite (and thus probably false) a change in confidence in the proposition under scrutiny does not occur. This is perhaps an unlikely scenario but, shows that the criterion tracks something different than what we would want to assess in a model. To be sure, there is some subjectivity in the appraisal of a model. Models are arguably appraised relative to their purposes, which in turn are defined by subjects. But this is not how Grüne-Yanoff seems to be taking this criterion to work.

Finally, just like Aydinonat, Grüne-Yanoff doesn't seem to be open to the possibility that Schelling's model fails. He argues that we learn from Schelling's model because "the model result thus justified changing one's confidence in the hypotheses about racist preferences being a necessary cause of segregation" (p. 856). That is, it shows that a segregation pattern may occur, given a possible initial condition (preference for not being a minority) and a possible process. Grüne-Yanoff thus suggests that Schelling's model is an example of learning by affecting an impossibility claim by means of offering a how-possibly explanation (2013, p. 856):

> Schelling's model shows that segregation patterns might be produced by another cause, which is an actual property of agents in many real-world populations: namely, the preference not to be in the minority (it shows only that it might be produced because it does so in a merely possible context—with an environment and a process that our knowledge does not rule out but that we by no means can assume to be the actual environment or process).

Grüne-Yanoff thus acknowledges that we can't assume the environment, or the conditions under which the result holds, to be like ours. This is the reason why the Schelling result is merely a possibility. However, as I already discussed with respect to Aydinonat's analysis, social scientists working with the model and with its many specifications have been in fact expanding their knowledge about the model, in order to determine the conditions under which the Schelling result continues to hold. Grüne-Yanoff, nevertheless, attributes epistemic import to the model without considering that such import is contingent on the conditions under which it is established that the model result holds. Surely the work that sociologists have done (some of which is mentioned by Aydinonat) suggests that the model may apply widely, but that is what we have learnt afterwards; Grüne-Yanoff claims that we learnt from the model originally conceived by Schelling.

Let me summarise my points now with respect to Grüne-Yanoff. I have argued that to defend the epistemic import of non-representational models is a different exercise than to appraise them. Appraisal requires different criteria than learning. First, because not all representational models are in the business of learning about the world. Second, because even if they were, learning is too coarse a criterion; we want to distinguish the "good" models from the "bad" ones, even if we can learn from them all. Third, learning as defined by Grüne-Yanoff is a subjective concept, not something we can attribute to models. And fourth, the way in which learning is employed as a criterion is too weak and doesn't leave open the possibility for failure.

3.3.3. Epistemology and the economics context

Now that I have discussed the details of each of the accounts, let me now offer some remarks about the literature in general. The enthusiasts, those whose view is that models by themselves do have epistemic import, would likely agree with the following way of characterising the epistemic import of models. In the cases in which a model result is robust, explanations derived from them are more likely to be reliable. This is the position of Kuorikoski & Ylikoski (2015), which is the most demanding and thus one with which neither Aydinonat nor Grüne-Yanoff would probably disagree with. In the cases in which there is no such robustness, it's possible, according to Aydinonat (2007) and Grüne-Yanoff (2013), respectively: i) to enhance understanding nonetheless, because a broader set of what-if inferences can be made, or ii) to learn, by a change in our confidence in a particular hypothesis. Since Kuorikoski & Ylikoski (2015) concede that it's possible to gain understanding from non-robust results, they would also agree with i). This suggests two things. First, that, at least for these commentators, epistemic import comes relatively cheap. For we can learn from non-representational models and gain understanding from non-robust results. This is, obviously, good news; this means that economists probably gain more understanding or learn just by toying with their models. But it also suggests that, if this is all there is to economic modelling, then it offers very little epistemically. In particular because, as I already mentioned above, it is debated whether economists are in the capacity to carry out proper robustness tests. So, this means that theoretical economic models will usually only be able to afford understanding, as defined by Aydinonat (2007), in the form of a broader meta-model of explanation, or because scientists are obliged to be explicit about the assumptions of their models, and their inferences are more reliable, as Kuorikoski & Ylikoski (2015) suggest[17].

In fact, a concern raised by Northcott & Alexandrova (2014)[18], is that philosophers like Aydinonat (2007) have defended and justified economists modelling activity as epistemically successful—because they have allegedly demonstrated that models do have epistemic import—when in fact the epistemic contribution, if any, is rather meagre. Their specific concern is that economists invest much effort and resources in devising complex mathematical models, that, in return, deliver very little—e.g. potential partial (theoretical) explanations.

My view on this matter is the following. While Alexandrova and Northcott's concern seems to be about (theoretical) economists investing their—and taxpayer's!—resources in an apparently futile matter as modelling, and they offer some evidence of their own of what they call armchair science, the question is, to what extent is their view of economists mediated by the attention that other philosophers have given to theoretical modelling? As I mentioned above, theoretical modelling has been the main, if not the only, kind of modelling that has received significant attention by philosophers of economics in their enquiry about models. The reason, surely, has to do with theoretical models being the ones that pose an "epistemological mystery" and are thus the object of enquiry of philosophers of science. In economics, idealisations such as *homo oeconomicus* or the representative agent have been the object of constant criticism for their lack of realisticness and have prompted interest in trying to find out the extent to which these assumptions hinder or foster the reliability of the models that use these assumptions. The motivation is not minor because economics is a model-based science, and philosophers have been concerned about understanding the practice. This is part of the legacy of the naturalistic turn that, in the modelling literature, has been exemplified by Cartwright (1983, 1999) and Morgan and Morrison (1999).

However, (pure) theoretical modelling, as analysed by philosophers, has been in decline in economics since the mid-eighties. While theoretical modelling played a major role in the

[17] Kuorikoski & Ylikoski (2015) suggest, in their defence of models as extended cognition, that there are at least three ways in which models enhance understanding: to oblige scientists to be explicit about their assumptions; to make inferences more reliable and to expand the scope of correct what-if inferences. The latter is more likely to be reliable if proven to be robust.

[18] This concern has also been raised in conference sessions.

sixties and seventies, afterwards it has declined sharply, at the expense of empirical observational and experimental work. An article published in the *Journal of Economic Literature* (2013) by economist Daniel Hamermesh shows this strong decline of theoretical modelling. He takes the total papers published in one year on the American Economic Review (AER), the Journal of Political Economy (JPE) and the Quarterly Journal of Economics (QJE) for six consecutive decades since 1963 and classifies them by different categories, among which is the "type of study" used, regardless of the topic. The different "types of study" are: theory, theory with simulation, empirical: borrowed data, empirical: own data, and experiments. Among the total papers published in one year in these journals, theoretical papers amounted to 50.7% in 1963, reached a peak in 1983 with 57.6% and then decreased steadily to 32.4% in 1993 and in 2011 only amounted to 19.1%. Here is the table of his findings (Hamermesh, 2013, p. 168):

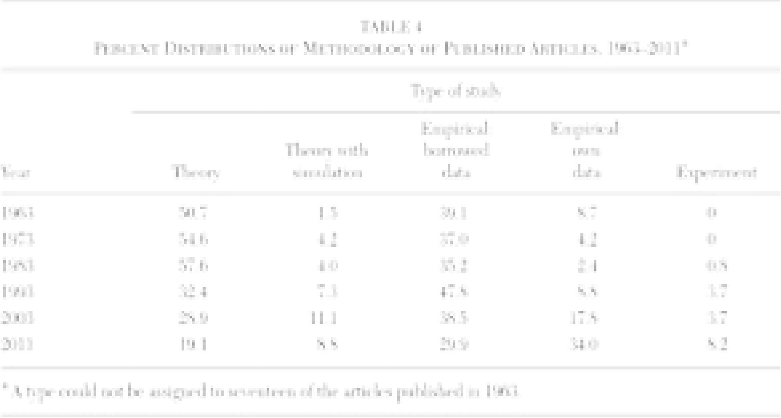


TABLE 4
PERCENT DISTRIBUTIONS OF METHODOLOGY OF PUBLISHED ARTICLES, 1963–2011[a]

	Type of study				
Year	Theory	Theory with simulation	Empirical borrowed data	Empirical own data	Experiment
1963	50.7	1.5	39.1	8.7	0
1973	54.6	4.2	37.0	4.2	0
1983	57.6	4.0	35.2	2.4	0.8
1993	32.4	7.3	47.8	8.8	3.7
2003	28.9	11.1	38.5	17.8	3.7
2011	19.1	8.8	29.9	34.0	8.2

[a] A type could not be assigned to seventeen of the articles published in 1963.

Obviously, while these numbers are very telling, they are just an invitation to investigate further these categories and the possible causes of these drastic changes[19]. Hamermesh speculates that possible causes are: first, theory having become so abstruse that journal editors refuse to publish it, recognising that few of the readers may be able to actually comprehend it; and second, developments in technology that facilitate the processing of empirical data. The latter doesn't necessarily mean that economists have ditched theory

[19] In a more recent and larger study using machine-learning, Angrist, Azoulay, Ellison, Hill, & Lu (2017) show that the trend towards empirical research holds. However, they also show that microeconomics, the largest field in the literature studied continues to be very much theoretical. See their paper for details. Cherrier & Backhouse (2016) contest this 'empirical turn' and argue, in turn, for an 'applied turn'.

altogether; in many cases they might be using this theory to test it empirically[20]. But the data still shows a striking and rather worrisome result, which is my second general point about the literature: philosophers have paid almost exclusive attention to a kind of modelling that, already by the end of the previous century, was substantially declining. A more provocative way of putting this point is that while philosophers claim to be interested in understanding the modelling practice, they have been ignoring most of what the modelling practice is actually about. Alexandrova and Northcott worry about economists wasting time and resources in developing models that do not seem to amount to much; presumably they should be at least as concerned about philosophers of models not having got their object of enquiry quite right.

To be sure, attention has been given to relatively new developments such as simulation (See e.g. Grüne-Yanoff & Weirich (2010) for a review; Frigg & Reiss (2009)); experiments (See Guala (2005)) and to evidence-based policy (see Reiss (2013, Chapters 9, 11) which are all topics that are more or less directly related to models. Furthermore, there is also important work about statistical and econometric models (See e.g. Mayo & Spanos (2004); Morgan (1988, 1990)) and the 'credibility revolution' advanced primarily by economists Angrist & Pischke (2010) and the subsequent discussion of instrumental variables in research design (e.g. Reiss (2005)). I therefore do not want to suggest that there are no other discussions of models. However, the literature is in general quite fragmented. By this I mean that there are few references and connections between the different branches mentioned above. In relation to the literature on theoretical modelling discussed above, it is generally discussed as if this was mainly what economic modelling was about. In fact, the literature is not even qualified as 'theoretical'. Another way to put this is that the philosophical literature on modelling paints a picture of economics as largely theoretical. The story by Hamermesh (2013) above and extended and clarified by Backhouse & Cherrier (2014) and Cherrier & Backhouse (2016) paints a very different picture: one in which economics has mostly relied on empirical models and that there was rather an exceptional period of 'high theory' in the fifties and sixties. In this respect, the sceptics, though they have restricted their analysis to the role played by theoretical models alone, they have at least recognised the role of models within a more complex, broader practice.

A similar point can be made about explanation (or understanding) as a favoured epistemic goal. Philosophers have generally associated theoretical models with the goals of

[20] I thank Emrah Aydinonat for suggesting this point.

explanation and understanding. There are, obviously, good reasons for this: the crucial question is whether it is possible to learn something from these models that are typically highly abstract and unrealistic. For example, Kuorikoski & Ylikoski (2015) write the following in the introduction of their paper (p. 3817-18):

> The importance of model-based reasoning in science has not gone unnoticed by philosophers, and the autonomy and perceived unrealisticness of most models have raised questions concerning the way in which they can provide understanding of the world. This puzzlement can be summed up in two questions. First, how can the manipulation of these surrogate systems provide genuinely new empirical understanding about the world? Second, how can models, which always incorporate assumptions that are literally untrue of the model target, provide explanations, if explanation is taken to be factive?

The issue here, though, is two-fold: one has already been highlighted by Reiss (2008, Chapter 8) which is that among the goals that scientists could pursue, philosophers and some social scientists have favoured explanation. Reiss argues that there are other similarly important, attainable, and methodologically contentious goals such as description, prediction and control (or intervention) and that there are no valid reasons to favour explanation over the others. According to Reiss, the "new mechanistic perspective", NMP, that has become fashionable lately in philosophy, emphasises the importance of investigating causal mechanisms and thereby of explanation as a goal, even when there are good reasons for investigating issues related to other goals. In the case of modelling we could say that the issue is that, while some models aim at explanation, it is not the only goal economic models have. Here the literature on modelling seems to be an instance of the phenomenon that Reiss highlights of systematically ignoring other aims. The way in which some of the authors discussed above define explanation and understanding, namely as knowledge of dependencies that in turn allow for correct what-if inferences, presupposes, in general, a mechanistic model which refers to an underlying structure or process that is causally responsible for the phenomenon of interest[21]. In this way, because philosophers are concerned with explanation as a goal, they have tended to focus their attention on what they regard as mechanistic models. They have done so despite the lack of agreement about what actually constitutes a mechanism in the philosophical literature, with many different views on offer (Machamer, Darden, & Craver, 2000; Reiss, 2007). So,

[21] Reiss (2007) recognises that despite the many views of models available, there are three features that they all share. This one is one of them.

by favouring explanation, mechanistic models are the focus of attention, but it is even unclear how a mechanistic model is to be unequivocally identified.

The other aspect of the issue is that explanation tends to be associated with mechanistic models, but they do not necessarily go together. Arguably, not all models that could be (or have been) regarded as mechanistic, aim at explanation. Tinbergen's models of the Dutch economy or the MIT-Penn-Fed model, for instance, can be considered mechanistic since they attempt to capture the causal structure of the economy, but they are not models that aim at explanation. Some of Tinbergen's models, for instance, had as main purpose the evaluation of different economic policies, among which was the abandonment of the gold standard and the subsequent devaluation of the Guilder (Maas, 2014). Likewise, the MIT-Penn-Fed model was a collaboration between the Federal Reserve, the University of Pennsylvania, and the MIT. The Fed commissioned the model to Albert Ando, at U Penn, and to Franco Modigliani, at the MIT. The model's purpose was forecasting as well as economic policy analysis. In particular, the Fed was interested in a model that would represent the monetary sector in a way that was adequate for the policy needs of the Fed. Existing models at the time did not have this feature (Backhouse & Cherrier, 2017). Surely, these models *presuppose* an understanding of the economy—that is, that the model correctly captures the causal structure of the economy—but they were *used* and *assessed* as forecasting devices (Backhouse & Cherrier, 2017). The problem is thus that among the models that can be regarded as mechanistic, prevalence has been given to those that attempt to explain phenomena, like Schelling's. Kuorikoski and Ylikoski (2015) grant as much in a footnote in which they justify having chosen Schelling's model as their case study (p. 3817):

> The checkerboard model is perhaps also the most used stock example in the philosophy of social science literature, and worries have been raised that using it repeatedly may have created biases in philosophical views. Granted, the checkerboard model might not be representative of economic and sociological models in general. However, as the model has been heralded as an example of a good explanation in social sciences (Sugden 2000; Hedström and Ylikoski 2010), it must embody at least some of the key virtues that social scientists expect their theoretical models to have.

Another way to formulate the issue with explanation and mechanisms is as follows. Reiss (2008) has framed his discussion around *the goals* of science, motivating his discussion by asking whether philosophers have reasons to prescribe explanation as the most important goal and ignore the aims that scientists set for themselves. This, in turn involves, according to Reiss, prescribing the use of mechanistic models. The strongest argument he considers according to which philosophers could be justified in doing so, would be if investigating mechanistic models attained other aims, aside from explanation. In that case, mechanistic models would be the best models of data (description), for prediction and control. Reiss concludes that for neither of those it is the case that mechanistic models are the best and therefore other kinds of models are also worth of enquiry. He defends enquiry into other models that are non-mechanistic because they respond to other aims. My point here is that there is a subset of mechanistic models that has other aims than explanation that has been neglected. If the literature on models only focusses on the mechanistic models that are explanatory, the subset of models being investigated is even more limited than if attention is given to mechanistic models in general.

To conclude, after going through an important part of the literature on the epistemology of modelling, there is arguably very little that is useful as criteria to determine model failure. Finding these criteria has surely not been the purpose of any of the commentators I discussed here, so my aim here has not been to criticise them for this. My critique is instead that an aspect of modelling such as failure, which is at least as important as success, has been neglected in the extant philosophical accounts of models. Such neglect is arguably the result of the limited scope of the literature. I have shown that it is limited in at least three fronts. First, the literature has focussed almost exclusively on theoretical models. As I noted above, by 2011 pure theory represented less than a fifth of the type of research that was done in economics. Even though the interest of looking at models has been partly to understand the modelling practice, it is unclear, at best, that this is being achieved if the object of enquiry is not representative of the actual practice. In this sense, Alexandrova & Northcott (2009) are the only ones who seem to recognise that models are used in larger contexts. Their view is incomplete, because they restrict their analysis to the role of models as open formulae. Second, the literature has largely focussed on explanation and thereby on mechanistic models. Above I argued that the criteria offered by commentators to justify the explanatoriness of the Schelling models aren't useful as criteria that may help us identify model failure. Third, even though there are mechanistic models that aim at other goals, only those that aim at explanations are the focus. In the chapter that follows I will look at some aspects that might be helpful to identify model failure and thereby broaden the scope of extant analyses. But first, I will look at some objections that could be raised against the analysis made so far in this chapter.

4. Objections

In the previous section I discussed the general approach of each of the three branches of the philosophical literature on models and concluded that none of them offer criteria that are helpful to determine model failure. In this section I shall discuss two possible objections to the points I have raised. The first is that my concern about diagnosing and identifying possible sources of failure is not something that the philosophical literature should be concerned with. The second is that my concern is in fact addressed, just under a different name.

4.1. Analysis of model failure is not the business of the philosophy of modelling

The first objection that can be raised against my claim that philosophical theories of models do not offer relevant criteria to assess whether a model has failed (or is more likely to fail) is that this is simply not the business of theories of models. Another way to put it is that this is not necessarily a task that philosophers have set for themselves. If this is the case, the argument would go, then it's pointless to expect that their accounts illuminate aspects of model failure. To be sure, a thorough discussion of this objection would lead me to discuss what the purpose of philosophy of modelling is, or even of the philosophy of science more generally. In particular, one could ask whether philosophy ought to have a normative character. I have a strong opinion on the matter; I think philosophy of science should be instrumental in conducing to better scientific practice. This requires profound understanding of the practice, involvement, and the aim for relevance. In this dissertation I attempt precisely to reconcile philosophical accounts with current practice. But I do recognise that this is a difficult philosophical question in itself and to properly defend my position would bring me too far away off topic. I will thus restrict myself to offer an argument based on a mere observation of the philosophical literature.

First let me offer a reason for why philosophers who are concerned with modelling may not have been particularly interested in understanding model failure. Traditionally, philosophers have been interested in the success of science. As scientific theories are considered the facilitators of this success, understanding their structure became of primary importance for philosophers. In this context, models were discussed only in so far as they were related to theories. In the syntactic view of theories or the Received View, as it is also known, models are structures that satisfy sentences of the formal axiomatic calculus that are theories. So, models don't have any significant role, except as playing a heuristic one, facilitating the understanding of the formal calculus (Portides, 2008). The semantic view, which superseded the Received View, gives more importance to models. In this view theories are identified with classes of models and, unlike in the Received View, models do have representational capacity and are thus considered vehicles of scientific knowledge. However, just like with the syntactic view, the interest here is to define the structure of theories and how such structure may be interpreted. In this case, such a structure is considered to be a family of models that yield the theory true. Arguably, therefore, the literature that has paid attention to models has traditionally attempted to offer a philosophically sound reconstruction of theories, which has little to do with appraising models independently from theories. The analysis of failure of models in this context could thus seem unnecessary or even incoherent.

There are, however, at least two reasons for why it could still be expected that current literature on models has something to say about model failure. The first reason is simply that philosophers have somehow moved on from the syntactic and semantic views and have endorsed a third, pragmatic view. This view, first defended by Cartwright (1983), incorporates many aspects about the practice of science that were previously neglected, such as the fact that models, and not theories, are crucial for understanding science and therefore should be the unit of analysis of philosophers[22]. Morrison & Morgan (1999) gave impetus to this new project, by arguing (and compiling work that also argued) that models are autonomous and only partially dependent upon theory and data. Recent literature has therefore focussed on models as autonomous objects and on understanding the role that scientific models play in science, as discussed above.

The second reason is that, aside from this new approach to understanding and characterising scientific practice, some philosophers seem to have a genuine interest in offering guidance as to how scientific practice may be improved. A good example is, again, Nancy Cartwright, who in her later work has striven to understand the extent to which causal claims are warranted, often offering guidance to policy makers with respect to the extent to which the alleged effects of certain policies may be warranted or not. For sure there are many other philosophers in the causality debates who have the same interest in providing relevant analyses for how causal claims are warranted, but Cartwright is certainly one of the very few who has got the attention of practitioners, at least from the point of

[22] See Winther (2016) for details.

view of economics (e.g. Deaton & Cartwright, 2016). So, this single instance suggests that there are indeed philosophers of science who have normative aims.

In general, it is not easy to identify what the specific purposes of philosophers are or what their stance is with respect to the normativity of the philosophy of science. The philosophy of modelling is surely not an exception. There are some cases in which the purposes of philosophers are, at best, confusing. For instance, in his book Michael Weisberg (2013) mentions in the preface that the book is an attempt to synthesise his 15-year-long thinking about why modellers often use incompatible and highly idealised models. This suggests that he is in the business of understanding and conceptualising the modelling practice. The interest appears to be solely philosophical—"Is there a satisfactory philosophical account of models that explains how models are used by scientists?" seems to be the question Weisberg is trying to answer. Later in the book, when he discusses why he will be considering only three kinds of models—and not more and not less—he corroborates this purely philosophical interest by suggesting that he has opted for an "epistemic level of philosophical theorising", whose purpose is to answer the question of how many categories of models are needed in order to build an account of model-based theorising. Other options, which he has discarded for his analysis are, "face-value practice of science", which would force him to ask the question of how many types of models scientists talk about, or an ontological perspective, in which he would ask the question of how many kinds of models there exist (p. 20):

> When I say that there are three kinds of models, I'm not making a purely descriptive claim about how many categories of models are recognised by scientists, nor am I making a point about fundamental ontology. Rather, I am arguing that a philosophical account of models and modelling needs these three categories to account for modelling as it is practised in contemporary science.

It is unclear whether the philosophical account of models that Weisberg has in mind is something that speaks to scientists. He doesn't say much about how this philosophical account relates to the scientists' perspectives or how it idealises or abstracts from their practice. In other words, it's unclear how the epistemic level of theorising he has chosen relates to the face-value practice of science. Nevertheless, towards the end of the book, Weisberg claims that his account of models offers a framework that may be helpful for scientists to "locate sources of disagreement and to give scientists a way to explicitly formulate their standards of fidelity" (p. 174). A case would have to be made for why such an account that explicitly deviates from the scientific understanding of models is actually capable of speaking to scientists. Weisberg doesn't make such a case and his purposes are, as a consequence, confusing. For he intends to offer a philosophical account and presumably to be relevant for the scientific practice, without really explaining how the two views relate to each other.

Another case that shows at least some ambivalence with respect to the motivations of philosophers to put forward a certain account is the following. In Kuorikoski et al. (2010, p. 541), it is argued that a great part of theoretical modelling in economics is dedicated to carrying out robustness analysis:

> A substantial portion of this modelling activity is devoted to deriving known results from alternative or sparser modelling assumptions. Why do economists spend so much time and effort in deriving the same results from slightly different assumptions? The key to understanding this practise [sic] is, we propose, to view it as a form of robustness analysis […].

One could argue that this is merely an idealisation of the practice. Presumably, such a characterisation allowed them to discuss the epistemic import of robustness analysis. While such an idealisation may offer the opportunity to discuss the epistemic import of robustness analysis, it may be difficult for practitioners to recognise themselves in such an activity, especially if the motivation of the account seems to arise from the description of a situation.

These are clearly just two instances of the literature so it would be inappropriate to draw any strong conclusions from them. But these two instances do show that philosophers may have different aims with their contributions, and these are not always clear. This situation may generate analytical problems, for different conceptual frameworks may be used, depending on whether the purpose is to offer a philosophically sound account of scientific practice or whether it is to offer relevant criteria that may be helpful for scientists to improve their modelling practices. It may also have some practical consequences. One could be that scientists intending to rely on philosophical contributions to illuminate their methodological stance may dismiss the entire literature as irrelevant if they don't identify themselves with the philosophers' portrayal of their practice. This may, in fact, be one of the reasons why despite the vast amount work there is in the philosophy of economics, there is little collaboration or interest from practitioners to look at philosophical work. So, this situation suggests that philosophers, at the very least, should try to make the purposes of their analyses explicit. Otherwise, they run the risk of being targets of analyses like this one, which demands things from them that they may have never set to do in the first place.

4.2. There is philosophical discussion of model failure; just not in the modelling literature.

Another objection is that there is philosophical discussion of model failure, but just not in the modelling literature. Another way to frame this objection is that I have created a straw man to build my case, because there *is* philosophical discussion of model failure, just under a different name and I have ignored it. There is quite some work both in economic methodology and general philosophy of science that deal with related issues. In economic methodology, Boumans (2005), Morgan (1988), and den Butter & Morgan (1998) are three examples of commentators who have discussed models in a broader context (than the limited one discussed above)and who have engaged with their means of appraisal. Morgan (1988) discusses the strategies that econometricians of the first half of the 20th century employed in order to bring together theoretical insights, which didn't have the necessary statistical properties, and statistical techniques in order to find satisfactory empirical models, which meant that they had to work well with the data available. Boumans (2008) does something similar, looking at a longer period and exploring the different ways in which economists (econometricians) attempted to validate their models given the statistical developments brought by the time. Den Butter & Morgan (1998) go beyond strict modelling practices and discuss the role that empirical models play in policy making by means of how policy makers use policy advice and how this feeds back in the modelling process. In a number of case studies, they discuss different institutional arrangements in which the interaction between modellers and policy makers take place; the 'value chain' of the interaction, or how each party attains value from the interaction; and attempt to determine some organisational (institutional) conditions that have to be in place for the interaction between policy makers and academic economists to be successful.

This literature offers important insights about the processes of model construction and use, the difficulties with which modellers might have been confronted, the methodological questions that have emerged given the techniques favoured by economists and the goals being pursued at different times, and the interaction between policy makers and modellers. These are clearly important insights that can contribute to a cogent understanding of model failure. After all, the difficulties raised by say, incompatible theoretical and statistical frameworks, as discussed by Morgan (1988) or the incontestable authority that a model can have over expert advice in policy decision making, as discussed by den Butter & Morgan (1998) seem like good candidates for potential sources of model failure. However, there are two reasons for why this literature hasn't been discussed above and is therefore not the object of my criticism. First, , this work doesn't address the question of model failure explicitly—to the contrary, they are mostly focussed on cases of success—which is my interest here. And, second, in some cases it is more historical- than philosophically oriented. By this I mean that there is more interest in detailed description of (historical), sometimes comparative, cases than in offering more abstract and general accounts of models and modelling, as the ones discussed above[23]. Therefore, while these detailed analyses might offer crucial insights for understanding model failure, they are not general theories of models. There's also the work of those who might be considered the first generation of economic methodologists, such as Terence Hutchison (1988) and Mark Blaug (1980), who dedicated a great deal of attention to the appraisal of economic theories. Their view was heavily influenced by Popper and Lakatos, which in turn lead Blaug to dismiss large chunks of economics as not fulfilling the Popperian/Lakatosian standards. These two approaches have been largely dismissed in philosophy of science, among other reasons, for their approach to the demarcation problem.

In philosophy of science—as opposed to economic methodology, specifically—there are also commentators who have dealt with aspects that may be interpreted as signs of scientific failure. Literature on social epistemology addresses issues about expertise and aggregation of judgement, the reliability (or lack thereof) of peer review and the communication of scientific findings to the public. These are issues that clearly raise questions about the reliability of science and knowledge in general. The literature on climate models in the philosophy of science, specifically on the inconsistencies between

[23] It has been brought to my attention that this claim about orientation towards philosophy or history as characteristic of some work might be contentious. To be sure, especially in interdisciplinary work that reflects upon scientific practice the distinction becomes blurry, which is, quite possibly, a positive state of affairs. However, two points should be noted. First, that the work is considered "oriented towards history" doesn't imply that there is no philosophy there. Second, it is difficult to ignore that we continue to operate within fixed institutional disciplinary boundaries; we read and publish in philosophy-, economics-, sociology-, etc. journals; cite and are cited by clusters of scholars with fixed professional identities—e.g. philosopher, historian—, and attend conferences that, though perhaps in spirit interdisciplinary, are generally frequented by rather homogeneous crowds. Still, I recognise that these are the boundaries that we need to break in order to bring about the kind of disciplinary changes this dissertation argues for.

models and different ensembles (W. S. Parker, 2006) and the role of values influencing uncertainty estimates (W. Parker, 2014; Winsberg, 2012) also have to do with reliability of models and estimates. So, there is surely other literature out there that I'm leaving out.

My aim is, in fact, to a certain extent, to raise the question why philosophical accounts of modelling in economics have not been permeated by these insights, which speak clearly about an aspect of the use of models for practical purposes. So, if anything, what I'm suggesting is that philosophical accounts of modelling in economics should also contemplate these aspects as constitutive of the modelling practice. The philosophical accounts I have discussed above analyse models in isolation and for very specific purposes, such as explanation or understanding; models are used for more than that. A proper understanding of the practice of modelling is incomplete without contemplating what makes it successful and what makes it fail. Above I showed that the attempts that have been made so far to explain the success of models are not sufficient to explain their failure.

Conclusions

The literature on models, perhaps as much as the general philosophy of science, has been motivated by the success of science and the role that models play in such an achievement. Taking as a basis the recent accusations by different types of commentators have made of economic models, my aim has been to explore different aspects of the literature in philosophy, to evaluate the extent to which extant philosophical accounts of models offer lessons or insights about model failure. My approach to model failure has been similar to the way in which accidents in aviation are investigated: finding the weak links in the chains of events that go from the technical to the human factors, all with sight towards accident prevention. I have explored the philosophical literature on models from the perspective of their ontology, their semantics and their epistemology with the aim to find possible accounts of model failure. From my analysis, I have concluded that current accounts of models, especially those concerned with how models relate to the world, namely the literature on semantics and epistemology, have little to offer in this respect. This matters for a literature that is concerned with understanding scientific practice and the role of models. The understanding of scientific practice and the role that models play there cannot be complete if failure, as much as success, is not reliably accounted for.

To engage with the literature on semantics, I discuss the conditions that Frigg & Nguyen (2016) have argued a theory of scientific representation should have in order to be a satisfactory theory of scientific representation. The conditions they establish are based on different accounts of representation and objections raised that so far have been provided in this literature. I also discuss Weisberg's account (2013) of modelling, an account that has formalised its notion of similarity. I argue that, as it is perhaps to be expected from a theory of scientific representation alone, which attempts to track successful and accurate representation, it is not sufficient to judge why or how a model has failed. Successful or accurate representation does not track epistemic import or inferential reliability.

With respect to the epistemology of models, I explore the literature in the philosophy of economics that has offered appraisals of the epistemic import of models. This literature has focussed to a great extent on Schelling's model of spatial segregation. As this model has received a great deal of attention, one might expect that clear criteria have been offered that explain the success of this model, which in turn could offer lessons for identifying failure. I suggest that the extant contributions limit themselves to argue, or rather justify, that it's possible to ,depending on the account, learn or understand from this model, but that the criteria offered for this remains silent about possible reasons for failure. In fact, in some cases, I have suggested, the given accounts for success are invalid. In general, the focus on theoretical models alone and on explanation as an epistemic criterion, at the expense of the neglect of others, make the literature very limited and thereby has little to offer in terms of criteria that can offer insights into how to identify model failure.

Later in this dissertation I will argue that the lack of such criteria, or at least attempts to shed light on model failure, an essential feature of the modelling activity, calls for a new turn in the modelling literature, namely towards the *pragmatics* of modelling. A first stab at this will be offered in the chapter that follows.

References

Alexandrova, A. (2008). Making Models Count. *Philosophy of Science*, *75*(3), 383–404.

Alexandrova, A., & Northcott, R. (2009). Progress in economics: Lessons from the spectrum auctions. In H. Kincaid & D. Ross (Eds.), *The Oxford handbook of philosophy of economics*. Retrieved from https://philpapers.org/rec/ALEPIE

Alexandrova, A., & Northcott, R. (2013). It's just a feeling: why economic models do not explain. *Journal of Economic Methodology*, *20*(3), 262–267. https://doi.org/10.1080/1350178X.2013.828873

Angrist, J. D., & Pischke, J.-S. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, *24*(2), 3–30. https://doi.org/10.1257/jep.24.2.3

Angrist, J., Azoulay, P., Ellison, G., Hill, R., & Lu, S. F. (2017). Economic Research Evolves: Fields and Styles. *American Economic Review*, *107*(5), 293–297. https://doi.org/10.1257/aer.p20171117

Atkin, A. (2013). Peirce's Theory of Signs. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2013). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/sum2013/entries/peirce-semiotics/

Aydinonat, N. E. (2007). Models, conjectures and exploration: an analysis of Schelling's checkerboard model of residential segregation. *Journal of Economic Methodology*, *14*(4), 429–454. https://doi.org/10.1080/13501780701718680

Aydinonat, N. E. (2008). *The Invisible Hand in Economics: How Economists Explain Unintended Social Consequences*. Routledge.

Backhouse, R., & Cherrier, B. (2017). The Ordinary Business of Macroeconomic Modelling: Working on the MIT-Fed-Penn model (1964 - 1974). Presented at the History of Macroeconometric Modelling, Utrecht University.

Blanchard, O. (2017, April 10). On the Need for (At Least) Five Classes of Macro Models. Retrieved 4 May 2017, from https://piie.com/blogs/realtime-economic-issues-watch/need-least-five-classes-macro-models

Blaug, M. (1980). *The methodology of economics, or, How economists explain*. Cambridge; New York, NY: Cambridge University Press.

Boumans, M. (2005). *How economists model the world into numbers*. Routledge.

Callender, C., & Cohen, J. (2006). There Is No Special Problem About Scientific Representation. *Theoria. Revista de Teoría, Historia Y Fundamentos de La Ciencia*, *21*(1), 67–85.

Cartwright, N. (1983). *How the Laws of Physics Lie* (First Edition). Oxford University Press.

Cartwright, N. (1989). *Nature's Capacities and their Measurement*. Oxford University Press.

Cartwright, N. (2005). The vanity of rigour in economics. In P. Fontaine & R. Leonard (Eds.), *The Experiment in the History of Economics* (p. 118). Routledge.

Case, K. E., & Shiller, R. J. (2003). Is there a bubble in the housing market? *Brookings Papers on Economic Activity*, *2003*(2), 299–342.

Cassidy, J. (2010, January 13). Interview with Eugene Fama. *The New Yorker*. Retrieved from https://www.newyorker.com/news/john-cassidy/interview-with-eugene-fama

Chambers, M. S., Garriga, C., & Schlagenhauf, D. (2009). The loan structure and housing tenure decisions in an equilibrium model of mortgage choice. *Review of Economic Dynamics*, *12*(3), 444–468. https://doi.org/10.1016/j.red.2009.01.003

Cherrier, B., & Backhouse, R. (2016). The age of the applied economist: the transformation of economics since the 1970s. *Open Science Framework*. https://doi.org/10.17605/OSF.IO/FGRJF

Colander, D., Goldberg, M., Haas, A., Juselius, K., Kirman, A., Lux, T., & Sloth, B. (2009). The Financial Crisis and the Systemic Failure of the Economics Profession. *Critical Review*, *21*(2–3), 249–267. https://doi.org/10.1080/08913810902934109

Curtain, T., & Rosenberg, A. (2013, August 24). What Is Economics Good For? Retrieved 3 May 2017, from https://opinionator.blogs.nytimes.com/2013/08/24/what-is-economics-good-for/

Deaton, A., & Cartwright, N. (2016). *Understanding and Misunderstanding Randomized Controlled Trials*. National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w22595.ack

den Butter, F., & Morgan, M. S. (1998). What makes the models-policy interaction successful? *Economic Modelling*, 443–475. https://doi.org/10.1111/j.0013-0427.2005.419_6.x

Duursma, M., & Heck, W. (2017, May 19). Nog 45 gemeenten hebben aparte Molukse wijk. *Nrc.nl*. Retrieved from https://www.nrc.nl/nieuws/2017/05/19/nog-45-gemeenten-hebben-aparte-molukse-wijk-9344587-a1559706

Elgin, C. (2007). Understanding and the facts. *Philosophical Studies*, *132*(1), 33–42. https://doi.org/10.1007/s11098-006-9054-z

Easterly, W. (2009). *Empirics of strategic interdependence: the case of the racial tipping point.* National Bureau of Economic Research. Retrieved from https://www.degruyter.com/dg/viewarticle.fullcontentlink:pdfeventlink/$002fj$

002fbejm.2009.9.1$002fbejm.2009.9.1.1492$002fbejm.2009.9.1.1492.pdf?t:ac=j$002fbejm.2009.9.1$002fbejm.2009.9.1.1492$002fbejm.2009.9.1.1492.xml

Florida, R. (2014). *The Rise of the Creative Class–Revisited: Revised and Expanded.* Basic books.

Florida, R. (2017). *The New Urban Crisis: How our Cities Are Increasing Inequality, Deepening Segregation, and Failing the Middle Class—and what we can do about it.* Basic Books.

Frigg, R., & Hartmann, S. (2009). Models in science. *Stanford Encyclopaedia of Philosophy.* Retrieved from http://stanford.library.usyd.edu.au/archives/spr2009/entries/models-science/

Frigg, R., & Nguyen, J. (2016). Scientific Representation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University. Retrieved from http://plato.stanford.edu/archives/win2016/entries/scientific-representation/

Frigg, R., & Reiss, J. (2009). The philosophy of simulation: hot new issues or same old stew? *Synthese*, *169*(3), 593–613.

Gerardi, K., Foote, C. L., & Willen, P. (2010). *Reasonable People did Disagree: Optimism and Pessimism About the U.S. Housing Market Before the Crash* (SSRN Scholarly Paper No. ID 1692761). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=1692761

Giere, R. N. (1988). *Explaining Science: A Cognitive Approach.* University of Chicago Press.

Giere, R. N. (2004). How Models Are Used to Represent Reality. *Philosophy of Science*, *71*(5), 742–752. https://doi.org/10.1086/425063

Giere, R. N. (2006). *Scientific Perspectivism.* Chicago: University of Chicago Press.

Godfrey-Smith, P. (2009). Models and fictions in science. *Philosophical Studies*, *143*(1), 101–116.

Guala, F. (2005). *The methodology of experimental economics.* Cambridge ; New York: Cambridge University Press.

Grüne-Yanoff, T. (2009). Learning from minimal economic models. *Erkenntnis*, *70*(1), 81–99.

Grüne-Yanoff, T., & Weirich, P. (2010). The Philosophy and Epistemology of Simulation: A Review. *Simulation & Gaming*, *41*(1), 20–50. https://doi.org/10.1177/1046878109353470

Grüne-Yanoff, T. (2013). Appraising Models Nonrepresentationally. *Philosophy of Science*, *80*(5), 850–861. https://doi.org/10.1086/673893

Hamermesh, D. S. (2013). Six Decades of Top Economics Publishing: Who and How? *Journal of Economic Literature*, *51*(1), 162–172. https://doi.org/10.1257/jel.51.1.162

Hausman, D. M. (1992). *The Inexact and Separate Science of Economics*. Cambridge ; New York: Cambridge University Press.

Hesse, M. B. (1966). *Models and analogies in science* (Vol. 7). University of Notre Dame Press.

Himmelberg, C., Mayer, C., & Sinai, T. (2005). Assessing high house prices: Bubbles, fundamentals and misperceptions. *The Journal of Economic Perspectives*, *19*(4), 67–92.

Hutchison, T. (1988). The Case for Falsification. In N. De Marchi (Ed.), *The Popperian Legacy in Economics: Papers presented at a Symposium in Amsterdam, December 1985*. Cambridge University Press, Cambridge.

Inman, P. (2017, January 5). Chief economist of Bank of England admits errors in Brexit forecasting. *The Guardian*. Retrieved from https://www.theguardian.com/business/2017/jan/05/chief-economist-of-bank-of-england-admits-errors

Krugman, P. (2009, September 2). How Did Economists Get It So Wrong? *The New York Times*. Retrieved from http://www.nytimes.com/2009/09/06/magazine/06Economic-t.html

Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2010). Economic Modelling as Robustness Analysis. *The British Journal for the Philosophy of Science*, *61*(3), 541–567. https://doi.org/10.1093/bjps/axp049

Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2012). Robustness analysis disclaimer: please read the manual before use! *Biology & Philosophy*, *27*(6), 891–902. https://doi.org/10.1007/s10539-012-9329-z

Kuorikoski, J., & Ylikoski, P. (2015). External representations and scientific understanding. *Synthese*, *192*(12), 3817–3837. https://doi.org/10.1007/s11229-014-0591-2

Lisciandra, C. (2017). Robustness analysis and tractability in modelling. *European Journal for Philosophy of Science*, *7*(1), 79–95. https://doi.org/10.1007/s13194-016-0146-0

Maas, H. (2014). *Economic methodology: a historical introduction.* (L. Waters, Trans.). Routledge.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about Mechanisms. *Philosophy of Science*, *67*(1), 1–25.

Mahler, J., & Eder, S. (2016, August 27). "No Vacancies" for Blacks: How Donald Trump Got His Start, and Was First Accused of Bias. *The New York Times.* Retrieved from https://www.nytimes.com/2016/08/28/us/politics/donald-trump-housing-race.html

Mäki, U. (1992). On the Method of Isolation in Economics. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, *26*, 19–54.

Mäki, U. (2005). Models are experiments, experiments are models. *Journal of Economic Methodology*, *12*(2), 303–315. https://doi.org/10.1080/13501780500086255

Mäki, U. (2009). Models and Truth. In *EPSA Epistemology and Methodology of Science* (pp. 177–187). Springer, Dordrecht. https://doi.org/10.1007/978-90-481-3263-8_15

Mäki, U. (2017). Modelling Failure. In Hannes Leitgeb, I. Niiniluoto, P. Seppälä, & E. Sober (Eds.), *Logic, Methodology, and Philosophy of Science: Proceedings of the Fifteenth International Congress*. College Publications. Retrieved from https://pdfs.semanticscholar.org/5332/6e9790dc24be8d3597ec98b8a2fdda541bde.pdf

Mayo, D. G., & Spanos, A. (2004). Methodology in practice: Statistical misspecification testing. *Philosophy of Science*, *71*(5), 1007–1025.

Morgan, M. S. (1988). Finding a Satisfactory Empirical Model. In N. De Marchi (Ed.), *The Popperian Legacy in Economics: Papers Presented at a Symposium in Amsterdam, December 1985*. Cambridge University Press.

Morgan, M. S. (1990). *The history of econometric ideas*. Cambridge [England]; New York: Cambridge University Press.

Morgan, M. S. (1999). Learning From Models. In M. S. Morgan & M. Morrison (Eds.), *Models as Mediators: Perspectives on Natural and Social Science* (pp. 347–388). Cambridge University Press.

Morgan, M. S. (2012). *The World in the Model*. Cambridge University Press.

Morrison, M., & Morgan, M. S. (1999). Models as mediating instruments. In M. S. Morgan & M. Morrison (Eds.), *Models as Mediators: Perspectives on Natural and Social Science* (pp. 10–37). Cambridge: Cambridge University Press.

Northcott, R. (2017). When are purely predictive models best? *Disputatio*. Retrieved from http://eprints.bbk.ac.uk/18061/

Northcott, R., & Alexandrova, A. (2014). Armchair Science. Retrieved from http://philsci-archive.pitt.edu/10888/

Odenbaugh, J., & Alexandrova, A. (2011). Buyer beware: robustness analyses in economics and biology. *Biology & Philosophy*, *26*(5), 757–771. https://doi.org/10.1007/s10539-011-9278-y

Parker, W. (2014). Values and uncertainties in climate prediction, revisited. *Studies in History and Philosophy of Science Part A*, *46*, 24–30. https://doi.org/10.1016/j.shpsa.2013.11.003

Parker, W. S. (2006). Understanding Pluralism in Climate Modelling. *Foundations of Science*, *11*(4), 349–368. https://doi.org/10.1007/s10699-005-3196-x

Portides, D. (2008). Models. In S. Psillos & M. Curd (Eds.), *The Routledge Companion to Philosophy of Science*. London ; New York: Routledge.

Reiss, J. (2007). Do we need mechanisms in the social sciences? *Philosophy of the Social Sciences*, *37*(2), 163–184.

Reiss, J. (2008). *Error in Economics: Towards a More Evidence–Based Methodology*. Routledge.

Reiss, J. (2012). The explanation paradox. *Journal of Economic Methodology*, *19*(1), 43–62.

Reiss, J. (2013). *Philosophy of economics: a contemporary introduction*. New York, NY: Routledge.

Reiss, J. (2015). A Pragmatist Theory of Evidence. *Philosophy of Science*, *82*(3), 341–362. https://doi.org/10.1086/681643

Rodrik, D. (2015). *Economics rules: why economics works, when it fails, and how to tell the difference* (First edition). Oxford ; New York: Oxford University Press.

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

Rosenberg, A. (2009). If economics is a science, what kind of a science is it? In H. Kincaid & D. Ross (Eds.), *The Oxford handbook of philosophy of economics*. Oxford; New York: Oxford University Press.

Shappell, S. A., & Wiegmann, D. A. (2003). *A human error approach to aviation accident analysis: The human factors analysis and classification system.* Ashgate Publishing, Ltd.

Stiglitz, J. E. (2011). Rethinking Macroeconomics: What Went Wrong and How to Fix It. *Global Policy*, *2*(2), 165–175. https://doi.org/10.1111/j.1758-5899.2011.00095.x

Stiglitz, J. E. (2015). Reconstructing Macroeconomic Theory to Manage Economic Policy. In É. Laurent & J. L. Cacheux (Eds.), *Fruitful Economics* (pp. 20–56). Palgrave Macmillan UK. https://doi.org/10.1057/9781137451057_3

Sugden, R. (2000). Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology*, *7*(1), 1–31. https://doi.org/10.1080/135017800362220

Teller, P. (2001). Twilight Of The Perfect Model Model. *Erkenntnis*, *55*(3), 393–415. https://doi.org/10.1023/A:1013349314515

The Economist. (2013, September 7). Crash course. *The Economist*. Retrieved from https://www.economist.com/news/schoolsbrief/21584534-effects-financial-crisis-are-still-being-felt-five-years-article

Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy*, 639–659.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. Oxford University Press.

Winsberg, E. (2012). Values and uncertainties in the predictions of global climate models. *Kennedy Institute of Ethics Journal*, *22*(2), 111–137.

Winther, R. G. (2016). The Structure of Scientific Theories. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/win2016/entries/structure-scientific-theories/

Ylikoski, P., & Aydinonat, N. E. (2014). Understanding with theoretical models. *Journal of Economic Methodology*, *21*(1), 19–36. https://doi.org/10.1080/1350178X.2014.886470

5

Towards a Pragmatic Account of Modelling—Amending Mäki

Knowledge and error flow from the same mental sources, only success can tell the one from the other.
—Ernst Mach

Towards a Pragmatic Account of Modelling—Amending Mäki.

Introduction

Extant accounts of models have generally focussed on understanding the success of models or, in other words, the features that allow models to have epistemic import. Accounts that focus on representation try to explain the success of models by means of the type of representational relation that they have with their targets. Other accounts try to somehow measure the epistemic import of models—what is it that we can actually learn from models? Do models afford knowledge? Knowledge of possibilities? Is knowledge of possibilities knowledge? Understanding?

However, almost no account has attempted to understand model failure, what it entails, or how to detect it. If philosophical accounts are meant to understand the role that models play in scientific practice, they must be able to account for success as much as failure: both are part and parcel of scientific practice and modelling in general. In this chapter, I argue that accounts of success can't double up as accounts of failure and therefore an explicit analysis of model failure is necessary. Furthermore, I argue that an analysis of model failure demands a pragmatic approach that views modelling as a process. Some commentators have focussed on parts of this process, but not in a sufficiently comprehensive way. Mäki (2017) is the only account that, in order to accommodate failure, has included pragmatic elements. I will argue that at least three further elements must be part of an account that attempts to understand failure.

The development of the chapter is as follows. In section two I argue that accounts of success do not work simultaneously as accounts of failure. In section three I discuss the idea of modelling as a process and some of the accounts that focus on this process even if just partly. Then, in section four, I discuss Mäki's attempt to accommodate failure in his account. In section five I discuss the three elements with which I amend Mäki's account. Section six concludes.

2. Why study failure if you can study success?

There are at least two important reasons why philosophy of science ought to be engaged with understanding model failure. The first is that philosophers of science claim to be interested in understanding scientific practice. Both the success and failure of science are equally significant parts of the practice. An accurate understanding of the latter demands, at the very least, an interest in both. In fact, if one defends a view that philosophy (and science, in general) ought to have societal value, as some university administrators claim nowadays, a focus on failure is perhaps even more pressing: the loss of value in terms of unemployment, foreclosures, bankruptcy, as well as the suffering of many due to the recent economic slump—that was allegedly a failure of models––call for attention to failure as a priority. Furthermore, as demonstrated by engineering practice, specifically the attempt to understand human error in disasters such as *The Challenger* or Chernobyl in 1986 (Reason, 1990), we can learn from the study of failure. A second reason is perhaps more controversial but not necessarily less important. If the role of the philosopher of economics, or of any other social science, is to play a role in the unification of the sciences, or in the way in which the sciences can offer complementary perspectives on the understanding of the social realm, as Ross (2014, Chapter 1) has argued, then the understanding of both success and failure is paramount. Only a comprehensive understanding of a discipline, of its strengths and weaknesses can allow philosophers to compare and gauge what each discipline has to contribute to the cross-disciplinary understanding of social reality. For economics, a model-based science, this inevitably means to be able to make assessments of models in all areas of the discipline and regardless of whether they are mainstream or not.

Some might think that there is little reason to worry about failure, at least explicitly, as I have done here, because any account of success serves equally as an account of failure. Not meeting the criteria for success is failure. If, as I argued in the previous chapter, philosophy of science is concerned only, or mostly, with accounts of why science in general, or models in particular, are successful, some might argue that philosophy is already fulfilling its task because, by implication, the accounts of success are also accounts of failure. While this position might sound intuitive, and might in fact have fed the reason for the neglect of failure in the philosophy of science, it is incorrect.

There are two aspects to consider. First, its logic. Success in science is normally defined using sufficient conditions. That is, whatever conditions have been found in the modelling literature or elsewhere that yield success, it is quite likely that there are other sets (known or not) that also yield success. In chapter two I offered sufficient conditions for learning. Now, if meeting criteria *S* implies that we have success, we can conclude that the absence of success is failure to meet these criteria *S*. It is obviously not the case, however, that not meeting criteria *S* implies failure. Now consider that here I'm suggesting that we might be able to identify a set of criteria *F* that yields failure. Someone who claims that failure can accurately be defined as lack of success would have to show that not meeting criteria *S* is the same as meeting criteria *F*.

The second aspect is something I already hinted at in the previous chapter, namely, the role of expectations in determining whether a model is regarded as success or failure. The possibility to judge failure is expectations-dependent whereas success is not. This means that if there's clarity about what is expected from a model, it is relatively straightforward to assess whether it has met those expectations, in which case it is considered a success and otherwise a failure. However, when there are no expectations, or at least these aren't explicit, a model may be regarded as a success, say, because it accomplished something the model was not expected to do, but not a failure.

Let me offer an example to illustrate this. Every professional cyclist who has a balanced performance in terms of both speed and (climb) endurance, aims to win the three grand rounds of Europe: Tour de France, Vuelta a España and Giro d'Italia. Nairo Quintana is a Colombian professional cyclist, team leader of the Movistar team, who has won two of the three: the Giro d'Italia and the Vuelta a España. He has been second at the Tour de France twice, the most prestigious of all. In 2017, he was expected to fight Chris Froome, Sky team runner and winner of the last three Tours de France, for the first place. Quintana, in the end, did poorly: he finished 12th in the general classification. Pundits thought it was foolish of his team's strategy to participate in the Giro d'Italia, which takes place in May, if the goal was to win the Tour de France, which takes place just over a month later. We can say that Quintana failed, according to the expectations that were set for his performance.

Now take Mikel Landa, a Basque support rider for Chris Froome in the Sky team. He is an excellent rider, who also run the Giro last year and won a stage. But there were no expectations at the Tour de France for him as an individual rider because there were several other contenders for the first place[1] and, more importantly, because his job in the Tour de France was precisely to make sure Froome could keep the *maillot jaune*—the yellow jersey that the first runner in the general classification wears. Landa was in such a good shape throughout the Tour, that on stage 18th, which finished on the Izoard, a long, massive climb, Landa made an attack that raised suspicions among commentators whether he was running as support for Froome or to improve his own standing in the general classification and perhaps take the yellow jersey himself. In the end, he finished the Tour de France 4th in the general classification, just one second away from making it to the podium. In Landa's case, there were no expectations that he would be so strong and finish in such a good place. His performance was considered successful—he got a contract to run in the Movistar team, as leader, at the Giro d'Italia in 2018.

[1] See Fotheringham (2017).

What this example illustrates is that commentators have been able to regard Landa's performance as successful, despite the fact that there were no expectations for him. Instead, the performance of Quintana has been considered a massive failure. The point is thus that appraisals of failure are expectation-dependent, whereas success is expectation-independent: an appraisal of success can be made without there being previous expectations.

Some might argue that having no expectations is itself an expectation. So that the fact that there were no expectations about Landa's performance is an expectation. In that case, expectations are somehow inevitable, the only difference is whether they are made explicit or not. So those about Quintana were explicit whilst those about Landa were not. That might well be the case. But that is, in any case, all I need to make my point. As external observers—of models or professional cyclists—it is necessary to be aware of these expectations in order to appraise something as a failure. In order to have this knowledge, expectations have to be made explicit: We need to know what the model (or the cyclist) was supposed to do in the first place. We can't judge some economic models as failures for not having predicted the crisis if that is not what they were meant or expected to do. Instead, a model such as Schelling's can and has been regarded as successful in offering explanations of urban segregation regardless of whether that is what Schelling expected the model to do[2].

In short, given that the criteria of success so far offered are most likely only sufficient conditions and that analysis of failure is expectation-dependent, an explicit analysis of failure, independent of that of success, is necessary for a comprehensive understanding of the modelling practice.

3. Modelling as a process

I have previously suggested that a way to understand models and the ways in which they might fail is to use a similar approach to how accidents are investigated in aviation. Part of the idea is thus to look at the 'chain of events' or the process of how models are conceived, built, and used. Such an approach and aspects inherent to modelling such as the role of expectations in judging failure that I just discussed are what I will call a pragmatic approach to modelling. This pragmatic approach not only traces the chain of events in modelling; it also considers different contexts in which models are used, how they are used by different agents, and how they might play in

[2] In the introduction of Schelling's Micromotives and Macrobehavior, Schelling describes how he was always very curious about how people arrange themselves. He gives a very vivid example of how he noticed how people sorted themselves out in a venue in which he was going to give a talk. This suggests, at least, that his model arose out of mere curiosity for phenomena which involves people's motives and the behaviour that arises as an emergent property and not because he intended to explain the process of racial segregation in particular.

arguments in which authority and power also play a role. Here I present an outline for a pragmatic approach to modelling, and focus in particular on elements that might be potential sources of model failure.

Let me start with the 'chain of events', or the process of how models are conceived, built and used. Obviously, there is a part of this that is not new. Some of the literature I discussed above presupposes this process and engages with a particular aspect of it. The branch of semantics, for instance, in particular the more recent accounts that are defined as pragmatic accounts of representation, recognise the importance of agency and purpose for representation to take place. However, in that literature the emphasis is given to the kind of relationship that is formed between a model and its target in the abstract. My interest here is that the process from conception to use is acknowledged and analysed as such, such that weak joints or spots, in which risk of failure is higher, can be identified. Another way to put it is that I'm interested in an approach that is accurately descriptive and at the same time capable of offering independent philosophical assessment.

There are other commentators who have addressed parts of the process explicitly. Mary Morgan (2012) discusses extensively, through a number of different models as examples, different aspects of models, their conception and their use. With respect to their conception, for instance, she discusses the Edgeworth Box (2012, Chapter 6), the role played by visualisation in its conception and how indifference and contract curves became standard in economics after this model, thanks to the possibility that visualisation of these curves offered. Marcel Boumans (1999) too has discussed a part of the process of modelling. Specifically, he has focussed on how models are built and has emphasised how this process brings together different ingredients. Boumans suggests that model building is like baking a cake without a recipe, because it is a process of trial and error, although there is already a good idea beforehand of what a cake should look and taste like. He emphasises that multiple ingredients are needed, which include functional forms or results of other models, as well as policy views and empirical facts.

The problem with the analyses of these commentators is that, like the others I have discussed above, they have focussed exclusively on the positive aspects of models, much at the expense of their weaknesses and possible sources of failure. Boumans's main claim is that the possibility to build a completely new model from different ingredients, without later being able to discern the individual ingredients is evidence of their built-in justification. This is in contrast to other commentators who argue that models are built and then tested against empirical data for their justification. Morgan, on the other hand, avoids making general claims about models; she prefers to engage with the details of each of the cases she treats. Sometimes she discusses how certain aspects of the modelling practice might be criticised or how they might have less epistemic import than other methods, but she is more focussed on highlighting the aspects that make each model interesting or significant for the discipline, rather than attempt to generalise why models might fail. Her aim, in her own words is "to present, as three-star tourist sites, some of the best known, and historically significant, models in economics" (2012, p. xv).

Mäki's (2017) account of models as representations can also be interpreted as focussing on (a part of) the modelling process. Mäki (2005, 2009) has generally defended a view of models as representations that allow modellers to isolate mechanisms of interest. He builds on the account of models by Giere (1990, 2004, 2006, 2010) as representations in terms of similarity. The basic tenet of this and other pragmatic accounts of representation is that representation is not seen as a dyadic relation between a model and its target, but as a many-placed relation that includes subjective elements such as agents' intentions and/or interpretation. The representational relation between a model and its target is determined by the agent who uses the model as a stand-in for the target[3]. Just like Giere, Mäki has offered his account of models as representational devices in which the emphasis is on the kind of the representational relation—Mäki defends resemblance, rather than similarity. However, over the years Mäki has modified his account, each time adding more elements that belong to the praxis of modelling. While Giere has defended the representational relation as a four-placed relation ("agent *A* uses model *M* to represent (part of the) world *W* for purpose *P*"), Mäki has continued to add elements to his account. In his (2009) the number of elements were six—agent, model, target, purpose, commentary and audience—and in his (2017) he had added another two—description and context. The fact that Mäki continues to add places to the representational relation[4] in order to be able to accommodate features of the modelling practice such as failure, demonstrates that the view that tries to frame the modelling practice purely in terms of representation is limited, ultimately misguided, and lends credibility to the need for a pragmatic account of models.

Another aspect of Mäki's (2017) account that requires attention is that it has dealt with model failure explicitly. I consider his approach to analyse model failure wanting in some respects, but some of the elements of his account are useful to characterise the process I'm interested in. Therefore, Mäki's account deserves a separate section.

[3] See Giere (2004) for a defence of the pragmatic aspects of representation in models, which suggests the emphasis ought to be in the act of representing rather than on representation.

[4] See Knuuttila (2005) for a critique on the emphasis on models as representational devices.

4. Mäki on failure

Let me introduce Mäki's account as it is presented in Mäki (2017). Mäki's departing point, as that of many other contemporary philosophers, is that the representational relation between a model and its target is not a simple two-placed relationship between the model and its target. A model doesn't represent its target unless there is some agent that intends this in the first place. In consequence, many philosophers now agree that for an object to represent something else, the relation between the object and what it is meant to represent is much more complex than the simple dyadic relation. Even though it is clear that the relation is not dyadic, different accounts have been offered about the kind of relation necessary for representation. Mäki's own account has itself changed, becoming more and more complex. His last version [ModRep] is as follows: (Mäki, 2017, p. 6):

Agent *A*
uses multi-component object *M* as
a representative of (actual or possible) target *R*
for purpose *P*,
addressing audience *E*,
at least potentially prompting genuine issues of relevant resemblance between *M* and *R* to arise;
describing *M* and drawing inferences about *M* and *R* in terms of one or more model descriptions *D*;
applies commentary *C* to identify and coordinate the other components;
and all this takes place within a context *X*.

Briefly, what this means is the following: An agent *A* stipulates that a model *M* will act as a stand-in for target *R*. This target can be an actual target, say the British labour market, or a possible one, such as an abstract process like the core-periphery concentration of economic activity. Agent *A* wants to use model *M* as a stand-in for the target for a particular purpose *P*. That purpose can be epistemic, such as explanation or non-epistemic, such as aiding in policy making[5]. The agent also has a particular audience in mind. Here Mäki talks about kinds of audiences, so academic, non-academic, experts, etc. He also talks about a purpose in addressing that specific audience, so to communicate, to persuade, to impress, to educate, etc. To prompt genuine issues of relevant

[5] Mäki only gives examples of these two purposes in passing and therefore doesn't say much about what exactly he means by "aiding in policy making". Considering that some models are meant to simulate the effects of a particular policy (and therefore learn about the possible effects), it is controversial that they are considered as non-epistemic by Mäki.

resemblance means that it is not enough that the agent merely stipulates a model as a stand-in for a target. It is necessary that i) the model has a likely capacity of resemblance with (or correspondence to) the target, "so resemblance must not be utopian" and ii) that irrelevant resemblances do not count. Furthermore, Mäki makes a distinction between the model and the model descriptions. The former is taken by Mäki to be imagined abstract objects, whereas the latter are the concrete items in which the model is typically expressed: mathematical symbols, narratives, diagrams, checkerboards, etc. The point here is that the modeller draws inferences about the model and the target in terms of these model descriptions. With the commentary, the idea is that the modeller makes explicit how the other components hang together: "Its task is to identify the various components of representation and to align them with one another". Finally, there is a context in which this all takes place. Mäki doesn't specify much more than this.

Now let me discuss Mäki's take on modelling failure. Let me start by his motivation. Mäki seems to be motivated mainly by a worry about the themes with which philosophy of science ought to be concerned. His point is thus broader in scope than mine when he argues that philosophy of science should pay more attention to the failures of science. He suggests that philosophy has been mostly concerned with those aspects of science that make it successful, and that in turn it has paid less attention to other aspects, like its failures, that constitute science and its practice just as much as the successes. He thus claims that "Developing accounts of the nature, conditions and dynamics of both failure and success should be on the philosophy of science's agenda. Ability to produce such accounts should be one of the criteria of success of philosophy of science itself" (p.2). Models here are thus an important but not the only element of science that should be analysed in terms failure. In this paper, Mäki focusses on models as a prominent style of scientific enquiry.

In his paper, Mäki identifies two types of modelling failure. On the one hand, he suggests, there is failure in modelling failures in the target system. By this he means that, there might be failure in modelling some targets to which we can ascribe a state of functioning properly but that may experience sudden breakdowns: heart failure in a properly functioning human body, for instance. Failure to model the heart failure in an otherwise healthy body is one kind of failure. This is analogous to the "failure in economic [models] to model the failures of the financial system" (2017, p. 3). The model captures correctly the properly functioning heart or economy, but can't account for the sudden breakdown. Mäki calls this "double failure": failure to account for the failure. The other kind of failure Mäki identifies is failure in modelling, which is when the model fails to model its target in its proper, "normal" functioning. Mäki focusses on the "double failure", since it better represents the failure of models in the context of the financial crisis[6].

Mäki recognises that there are currently many different accounts of modelling in the philosophical literature and claims that the capacity of any of these accounts to deal with model failure may be considered to be one of the criteria to judge their success as philosophical accounts. Since he has an account of his own [ModRep], Mäki's purpose is to 'test' it, by determining how much it has to say about model failure.

> There are many such accounts [of models] available in the literature, and the challenge is to compare them for their credentials. One obvious way to proceed is to check them against empirical evidence concerning actual models and actual modelling practices. And provided we take these practices to include failures, then the capacity of the philosophical accounts in dealing with such failures may be taken as a major criterion of the success of those accounts (p.2).

Mäki doesn't offer an argument for why his account is capable of accommodating modelling failure. Instead, once he has stated that he is interested in only one particular kind of modelling failure, the "double failure", he proceeds with a discussion of how each of the components of his account is allegedly capable of accommodating different sources of modelling failure. The test is thus, presumably, that if each of the components of his account is somehow able to deal with aspects of modelling failure, then the account can be regarded as having passed the test. It is not really clear, however, what being able to deal with aspects of the modelling failure is.

Since Mäki proposes this test and suggests that every account of modelling should be subject to testing, one could argue that he, at least implicitly, offers a model or framework to test other philosophical accounts of models for their capacity to say something about model failure. That is, in principle, we could use Mäki's way of proceeding to test his own account, namely to go through each of the components of the account, to do the same with other accounts. In fact, by claiming that a measure of the success of a philosophical account of models is its capacity to account for model failure, Mäki seems to be inviting proponents of other accounts to carry out such test of their own accounts. I could use such a procedure for my own purposes: if my interest is to explore what accounts of models have to say about model failure, I could follow his example and do what he does with other accounts.

[6] Dani Rodrik (2015, Chapter 5) also suggests something similar when he claims that the financial crisis can be interpreted as an 'error of omission' given that many of the causes have been analysed and understood by economists, but that they ignored these models for favouring others that support the idea that markets are efficient and don't need intervention.

The problem, however, is that Mäki says very little about what the test consists of. Except for suggesting that the accounts need to be checked against evidence related with actual models and modelling practices, it is unclear what the test is about. In fact, it seems that Mäki pursues two different projects in his paper. He first suggests that any account of models has to be tested for its capacity to accommodate failure as much as success. The test supposedly involves checking accounts against models or modelling practices. I can think of one way to do this, namely that one would take a model or a cluster of models and assess whether it is capable of accurately performing under normal circumstances and under a target failure, following Mäki's definition of failure. In order to do this, some kind of criterion would be necessary to determine what it is to perform well under normal circumstances and what it is to fail. Then, the test of the account would be to see whether the account is capable of accounting for or explaining the failure. Instead, what Mäki does is show, or rather attempt to show, that there might be sources of failure that could be categorised in each of the components of [ModRep]:

> The components [ModRep] and their relations will next be investigated as potential loci and sources of modelling failure. It appears that many existing critiques of economic modelling can be construed as focusing on some specific component in the structure of [ModRep] and that some other possible critiques can also be envisaged within this framework (Mäki, 2017, p. 6).

What he ultimately does is a classification exercise. To be more specific, Mäki starts with the first component of his account <<*agent A*>> and discusses some of the criticisms that have been made about models that could potentially be classified as a failure of the agent using the model. He continues with <<*uses multi-component object M*>>, and so on, until he reaches his last component <<*context X*>>. For instance, his discussion of <<*addressing audience E*>> is based on a criticism by Joseph Stiglitz and John Quiggin, stating that, prior to the crisis, there was the belief among leading academics and politicians that unregulated markets have self-stabilising capacities, and that their beliefs were mutually reinforced, leading them to ignore important aspects of the economy. This criticism signals the failure of overspecialised economic work and the prevalence given to ideas that are more likely to be accepted by the discipline as a whole. Mäki discusses this in terms of test partners as the first curators of the ideas that are disseminated to the larger audiences like academic journals and conference presentations. In general, the problem here seems to be that audiences determine to a great extent the ideas that are disseminated and, in this case, there was lack of a big picture because other aspects like technical details prevailed in the current audiences being addressed. Presumably, since the criticism is categorised in one of Mäki's components, this makes his account successful in dealing with this kind of model failure.

This way of proceeding can't be considered a test. If in order to test other accounts we were to proceed in the same way as Mäki does, namely by classifying criticisms into the different components in each of the accounts, we wouldn't really be able to appraise that account. Suppose that we proceed in the same way to test Giere's account. We would then classify the criticisms into the four components of Giere's account: agent *A*, uses model *M* to represent target *T* for purpose *P*. So, for instance, the problem I just mentioned that Mäki classifies as belonging to *audience*, in Giere's account it would have to be classified as of the agent *A*, or perhaps the purpose *P*. Giere's classification would just be coarser than Mäki's.

Perhaps one could make the case that Mäki's is a better account because it classifies potential failure more precisely, but that would require at least two extra steps. First, that the accounts be compared. In this case, the test would no longer be of the individual accounts with respect to models but among accounts. Second, that we come up with some criteria that allows us to judge which account accommodates failure better. It is far from obvious that the number of components is a useful criterion to compare accounts of models and to determine that the more components an account has, the better it is.

If Mäki's exercise does not constitute a test for the capacity to account for model failure, how does it help us to better understand model failure? Or, in other words, what do we learn from Mäki's account about model failure? Clearly, to be able to classify critiques into different categories might indeed signal different potential sources of model failure. To categorise criticisms as potential sources or kinds of failure is helpful in the same way that the causes of an aviation accidents are categorised into technical, human-communication, human-institutional, etc. If we identify a failure that may be prevented, we might avoid accidents or unnecessary model failure. So, to be able to say, for instance, that the shock therapy advocated for Russia by prominent economists to reintroduce a market economy after the collapse of the Soviet Union, which culminated in massive loss of industrial production, output and dramatic hyperinflation, might have been a failure in properly gauging the *context*—and not, say, the *purpose*—is certainly helpful. Identifying the failure as a lack of understanding of Russian culture, the underdevelopment of legal and social institutions and thereby misjudging that a market economy could suddenly emerge, is probably more accurate than to say that there was a failure in the purpose by attempting to reintroduce a market economy in a flawed command economy[7].

[7] I can see how this claim can be considered controversial for some, particularly in times when confidence in markets seems to be receding. The adverse effects of globalisation and the future of capitalism are a fascinating topics that for obvious reasons I do not address here.

Of course, to be able to judge whether it is a failure in *context* and not *purpose* or the *agent* and not the *model*, requires clear definition of each of the categories. Furthermore, some criteria, or at least some reasoning behind what precisely constitutes failure (and success) in each of the categories is necessary to determine actual failure. It is not sufficient to rely on criticisms such as Stiglitz's or Quiggins's as above, to establish failure. While the critique is a good place to start looking for possible sources of failure, the critique cannot be taken for granted. Independent criteria are needed. After all, these criticisms could be misguided.

Unfortunately, Mäki doesn't do any of this. Mäki offers a laundry list of aspects of the modelling process that come in handy in identifying potential sources of model failure but not a substantive account that tells us what are the criteria in each of the categories established that flag, at least potentially, failure.

Let me offer another example from Mäki's analysis to illustrate this. Take his discussion of the *agent* as a possible source of failure. Here Mäki discusses two aspects for which economists have been repeatedly criticised, namely for being too narrow in their interests, downplaying the importance of other social sciences and for being more self-seeking than other disciplines. He regards this aspect a possible source of model failure. It is useful to quote him at length (Mäki, 2017, p. 7):

> Economists are generally recognised as intelligent people. Yet the critics argue that this is not sufficient for successful modelling and that the failures regarding the 2008 crisis are one indication of this. They say economists are too narrowly educated (mainly just in contemporary economics, math and statistics), too ignorant about history (of the economy and of their own discipline), about the other social sciences, about culture and human psychology. Some say their competences and epistemic preferences are ill suited for modelling the complexities of social reality. Their mathematically inclined style of inquiry encourages them to streamline the nuances of the real world in epistemically harmful ways: they are extremely skilful in mathematical puzzle solving when reasoning about the model worlds, but relatively speaking clumsy and uninformed in connecting their formulas to the detailed complexities of real world economies. These capacities and their limitations may also nurture over- confidence, hubris, and arrogance – characteristics often attributed to the economics profession and conducive to the sorts of failure witnessed in connection to the 2008 crisis (see e.g. Posner 2009; Fourcade et al. 2015).[…] Regarding the contents of their worldviews, there is empirical literature suggesting that economists are more self-seeking than other professions, either due to economics education or self-selection (see e.g. Carter & Irons, 1991). This may be suggested to result in systematic biases in favour of models that put too much stress on self-seeking behaviour amongst the populace at large.

I can think of two ways in which these allegations can be interpreted. One is merely as pure rant, which is quite common in some circles[8]. Many people say that economists are narrow minded and poorly educated and that they are not even interested in the history of their own discipline. Blaug (2001) made this point distinctly. However, I doubt that rehearsing the rant is what Mäki intends. It is not what interests us, as philosophers, for an account of modelling that allows us to detect model failure. Another option is therefore that Mäki uses these criticisms as guidelines for establishing criteria for model failure. What we are interested in is criteria that help us judge when a model might be considered a failure. So "narrowly-educated/-minded model user" could be one, as per above. It could thus be that a certain level of narrowness in education increases the likelihood of model failure—e.g. for lack of a comprehensive view of the world. The problem with using this particular criticism as a guideline for a criterion of failure is that it can't really be used as such because, if being "narrowly-educated/-minded model user" is what explains the crisis and model failure in general, then it also makes model success a mystery. In other words, we can't say that economists' narrow education is what explains their model's failures because those same narrowly-educated economists also build successful models.

Furthermore, even if we put that aside and suppose that we should interpret the criticisms as guidelines for criteria, only identifying them is not sufficient. Critics may point to these aspects and we, as laypeople, may casually accept them because intuitively they make sense: J.M. Keynes, Irma Adelman or J.A. Schumpeter probably had a better sense of the world than, say, Ed Prescott. But the task of the philosopher, the task of an account of modelling, is to spell out *why* a narrowly-educated model user is likely to produce models that fail more often. It could well be the case that a critic comes up with a criterion and spells out why this is likely to be an aspect that renders models more fallible. Surely, as philosophers we can only but go along with such an appraisal. But as philosophers we also need to be able to provide independent reasons to support these as criteria. Mäki, unfortunately, doesn't do this. In fact, just as other philosophers, Mäki seems ready to take sides with respect to whether economists should be praised or condemned. In this case, they aren't praised.

In short, Maki's analysis is still incomplete with respect to the criteria that are likely to help us identify model failure. A substantive account of modelling would have to offer, at least, guidelines that indicate why a certain property of a model or the modelling process is likely to lead to failure. Maki's account doesn't provide that. In the remained of this chapter I will argue that there are at

[8] See Rapley (2017) for a recent example in popular media. See Csaba (2017) for an academic piece, and for a reply see Vergara-Fernández (2017).

least three further elements that have to be considered before a substantive account that accommodates failure can be offered. I will also offer some preliminary criteria that, though defeasible, is likely to encourage further work in analysis of model failure.

5. Three elements

I will focus on the categories of agent, purpose and context in particular. The elements I will offer in each of these categories, which are based on how some models are used in economics, will contribute to the analysis that is necessary to fully grasp the process of economic modelling and, specifically, to be able to systematise the understanding of failure. Only a substantial account of the modelling practice that is capable of explaining both the success and failure of models, is likely to help to prevent unnecessary model failures.

5.1. Incentives I

There's one aspect in the category of the agent that is quite significant and neither Mäki nor any other commentator of the philosophical accounts has so far discussed. Despite the prominence that recent accounts of models have given to the role of the agent in the establishment of the representational relationship, the agent or, more specifically, their 'identity', has been neglected. The agent has generally been assumed to be a person who uses a model in order to learn about the world; it is someone who attempts to uncover the truth about the target they are modelling. Often, the agent is assumed to be an academic who is trying to gain understanding of a particular phenomenon. Yet, as we know, model users are not only those who have exclusively epistemic interests. There are some who are interested in having an accurate picture of the world *in order to do something with that knowledge*. Therefore, they have the incentive to be as accurate as possible in the modelling exercise because there is a higher goal that they want to achieve. A good example are board members of the Fed (and their teams), whose ultimate aim is to carry out monetary policy. In some cases, the higher goal might be to advance some private gain, like investors or risk managers who benefit financially from their accurate assessments. These agents generally have the incentive to model their object of interest accurately, but sometimes they might have additional (and perhaps stronger) incentives that pull the modelling exercise in a different direction.

When all the incentives pull in the same direction, it is perhaps safe to assume that there's the sole interest to model the phenomenon of interest accurately and that there's thus a purely epistemic incentive. However, the analysis of some of the causes of the recent financial crisis suggests that when incentives are at odds with each other, non-epistemic incentives might trump the one to model the phenomenon of interest accurately. Take analysts at rating agencies such as Moody's prior to the crisis. While it was certainly in their interest to model the risks posed to each of the Collateralised Debt Obligation's (CDO) tranches accurately in order to give an accurate rating of these products, commentators have discussed how changes in corporate structures in the nineties changed these incentives, making accuracy in risk measurement less important.

Sam Jones (2008) has noted two important changes in the financial sector that might have created conflicts in the incentives analysts had. There is, first, the change in who the client of the rating agencies was. Since the 1920s, when it was already common practice to rate corporate bonds, ratings were paid by investors, in the form of subscriptions. Since investors are the ones interested in evaluating the risk involved in their potential investments, it is in their interest to have an expert rating agency do this work for them. However, there were two forces that generated tensions in this scheme. One was that the rated products increased in complexity—CDOs, for instance, emerged in the eighties—while the increasing size of the industry made the subscription scheme unsustainable for the rating agencies: the expertise needed was greater than what agencies could afford within such a system. The other was that, at some point, ratings became indispensable, with investors requiring two ratings for each financial product—one from Moody's and the other from Standard & Poor's, the only two players in the industry at the time[9]. Being the only two firms in the business, this gave the agencies power and independence, which in turn caused exasperation on the side of banks, the issuers, because the banks fully depended on these two companies to be able to issue their products. According to Jones (2008), these tensions led eventually to a change in the scheme, whereby ratings began to be paid by the issuers of the bonds: the banks. In addition, Jones (2008) suggests that the independence and stringency of judgement for which rating agencies were known changed due to the vision of an influential character, Brian Clarkson, who was in charge of Moody's mortgage bond division in 1997 and who changed the personality of the business. Clarkson viewed ratings as a service, which required cultivating their clients and establishing amiable working relationships with them, instead of the outcome of a serious and independent analysis by academics, as rating agencies were previously perceived.

Second, Jones (2008) points out that in the year 2000 Moody's was floated as a public company. This created a culture of being driven by profits within the company that no longer allowed it to have an independent judgement. Ann Rutledge, who worked at Moody's in the structured finance department in the mid-nineties, is quoted by Jones (2008) as saying that Moody's was "lily white"

[9] While Fitch, the other player in the rating industry, was founded in 1914, it only became a nationally recognised statistical rating organisation (NRSRO) by the U.S. Securities Exchange Commission in 1975 and only in the nineties became a significant player (Jones, 2008).

until the company went public and lost the oversight of the financial publisher Dun & Bradstreet. People like Clarkson deny that accuracy was ever jeopardised—it was never a policy to subordinate ratings to the company's market share. However, Jones adds that many do claim that there was fear of losing business, in particular because employees were often rewarded with stock options. This clearly creates the incentive to act in ways that maximise the immediate gains of the company but not necessarily its sustainability, which is best secured by offering the most possibly accurate results and thereby maintaining its reputation.

This example shows the tension that was created in the incentives of analysts when a conjunction of conditions was in place: investors handed over the role of client of rating agencies to the banks, the issuers of the products to be rated; a culture of service to the clients, rather than of expertise and judgement took over the industry; and Moody's in particular, floated as a public company changed the reward schemes of its employees. The tension in incentives arises because satisfying clients doesn't translate into making accurate assessments necessarily. Instead, it translates into subjecting assessments to the expectations of their clients—e.g. a junk CDO with a triple A rated tranche. In turn, satisfying the client brings in more business and more business translates into higher corporate profits and, in a public company, this translates often into higher share prices, with which employees are compensated.

How are we supposed to analyse a situation like this one, when trying to understand model failure? If we ignore the clash of incentives, and all we observe is a sudden downgrade to 'Junk' of billions of dollars in CDOs that were originally rated as triple A, without an account that is capable of accounting for the influence of the incentives of the agent, we might be inclined to attribute the miscalculation to the technical performance of the model alone[10]. Instead, if there is awareness of the clash of incentives of the model user, the failure will accurately be regarded as of the agent and not exclusively of the model. The point here is therefore that since some incentives might clash with the purely epistemic ones, it is necessary to consider this in the analysis of models for an accurate assessment of failure. In the example just provided the clash is between what I've called epistemic and private incentives, but there might be other cases in which the incentives that clash are of a different nature. Failure to acknowledge these clashes will inevitably lead to bias the judgement of model failure, for instance by attributing it to a lack of resemblance or similarity with

[10] Some commentators–-e.g.Barnett-Hart (2009); Benmelech & Dlugosz (2010)–-have suggested that the miscalculation of CDOs was indeed caused to some extent by the failure to include default correlations at a macro level. But they also acknowledge the prevalence of 'rating shopping' by issuers in the build-up to the crisis.

the target, ultimately generating the type of accusations that many commentators have made, namely that the models—and not their users—failed.

Considering that economic models are used not only within the academic domain, but also, and increasingly, in many other domains such as government administration (see den Butter & Morgan (2000) for a range of contributions about the role of economic models in policy making) and the private sector—Cherrier & Backhouse (2016) refer to the frontier between academic economics and private enterprises like Google as 'increasingly porous'—competing, and often contradictory, incentives are more likely to emerge. This situation should compel us to consider the identity of the agent, or the elements that allow us to judge what their incentives might be. Mäki's account could thus be amended with the incentives of the agent such that we obtain:

Agent *A,* who has the identified incentives *I*,

uses multi-component object *M* as […].

In addition, preliminarily, we can offer the following guideline for identifying model failure: models whose users have (identified) clashing incentives are more likely to fail. Obviously, before we take it seriously such a guideline we have to, at least: i) test it in other contexts; ii) provide a threshold or definition of what counts as clashing incentives. I'll leave this research for a future project. Here I'm interested in signalling the potential different sources of failure and the categories that we'd have to consider before we start delving into each of them.

5.2. Goals G

One could argue that the need to account for the incentives of the agent that I have just discussed is already taken care of by the inclusion of the purposes of the agent in the account of models offered by Mäki. That is, the purposes of the agent have been considered precisely because it is acknowledged that agents have a variety of aims when they use a model. So, in the example that I provided above with rating agencies, it could be said that the purposes of agents at Moody's differ from those of an academic economist: the former's are to maximise private gains whereas the latter's are to understand the phenomenon of investigation. While I don't think there's anything in Mäki's interpretation of purposes that precludes this way of accounting for the different incentives that agents have, there are at least two problems with this proposal.

First, that purposes and incentives are not the same thing and thus a single place for both is insufficient. Above I did not suggest that agents have a multiplicity of purposes. Instead, I suggested that, in general, agents probably have the same epistemic purpose, like correctly representing the phenomenon of interest and understanding it, but that there might be a discrepancy between the incentives they have. The purpose of analysts at Moody's was to use their risk models to accurately assess the risk associated for each of the tranches of a CDO. They also had the incentive to do this properly. However, they also had incentives that, as I showed above, conflicted with the main purpose of using the model. This suggests that in order to properly be able to account for the possibility of conflicting purposes and incentives, a single place in the account for purposes is not sufficient.

Second, even though Mäki doesn't explicitly impose any restrictions on what belongs in the category of purposes, other elements in his account and the way in which he defines model failure suggests that the purposes considered are rather limited. For there are some purposes that are not necessarily advanced by an improved resemblance between the model and the target. In the discussion of one of the elements, the required resemblance between model and target, <<. . . *at least potentially prompting genuine issues of relevant resemblance between M and R to arise*>>, Mäki suggests that the model has to resemble the target in relevant ways in order for the agent to be able to use the model as surrogate of the target. He says as much in Mäki (2009), where he makes explicit that it is not sufficient to stipulate the representational relation between the model and its target; learning about the target requires resemblance: "Secondly, one could only hope to learn about target *R* by examining model *M* if *M* represented *R* in the second sense: *M* resembles, or corresponds to, the target system R in suitable respects and sufficient degrees" (p.32). On this basis, failure is defined as not resembling the target either because the modeller tried and failed—issues of resemblance didn't arise—or because the modeller didn't intend the model to resemble the target in the first place—the modeller engaged in "substitute modelling" rather than "surrogate modelling".

There are, however, situations in which *less* resemblance between model properties and the target are what advance modellers' purposes. For instance, in a paper in which Northcott (2017) tries to make a case for giving more attention to the purpose of prediction than what scientists and philosophers have given, he describes the case of weather forecast models in which less resemblance between the model and the target has proved beneficial for the predictive purposes of the modellers. Northcott's point of departure is that scientists and philosophers have generally favoured the development of theory and mechanisms, with the aim of being able to explain phenomena. He then suggests that empirical success is a necessary condition for explanation, and that in social and field sciences this success is sometimes only achieved by using purely predictive models. In order to advance his case, he discusses weather forecast models. The core of these models is some differential equations that represent the laws of thermodynamics, which govern dynamics of air in the atmosphere and how these are affected by temperature, pressure, and other factors. Northcott suggests that these equations are known as the "fundamental theory that remains a true description of the weather system" (p. 21). But, according to Northcott, in the attempt by weather forecasters to improve their forecasts, not only is it not sufficient for these equations to produce accurate forecasts, but the refinements that have been made of the models based on the fundamental theory hasn't proved successful. Instead, in a first version of a model in which the effect of mountains on atmospheric circulation, air flow, and precipitation was included, a physically realistic cut-off mountain height proved to decrease the predictive accuracy of the model. A following version of the model, still based on the fundamental theory, but deviating from the physically realistic, causally explanatory feature of mountains, their cut-off height, improved the forecasts produced by model. Thus, a model that resembled the target less, proved better at predicting: "Notice the sequence here: the *less* physically realistic formulation was the one eventually adopted, because it generated more accurate forecasts" (Northcott, 2017 p. 24). Reiss (2007) has made a similar point.

Another case are the tests that were carried out at the beginning of the fifties by the Cowles Commission of Lawrence Klein's structural models of the US economy. Klein's 16-equation model was compared in its predictive performance against two 'naïve models': one predicted that tomorrow's income would be like today's plus a random error—Y*(t+1)=Y(t)+e(t) and the other included the difference in output between today and yesterday plus a random error—Y*(t+1)=Y(t)+(Y(t)-Y(t-1))+e(t). In the end, the naïve models proved to offer better predictions than Klein's models, despite the fact that the latter were allegedly capturing the structure of the economy (see Christ (1951) for details; Maas (2014, Chapter 6) and Boumans et al. (2010, pp. 42–46) for discussions).

The point is therefore that if Mäki's account defines the element of <<. . . *at least potentially prompting genuine issues of relevant resemblance between M and R to arise*>> as I described above, then purposes like prediction, or at least some cases thereof, for which it is not necessary that issues of relevant resemblance arise, can't be properly accounted for by Mäki's account.

Of course, with respect to the weather forecast models Mäki could reply that *both* models resemble the target because the fundamental equations are still part of the core of both models. So there are some issues of resemblance that arise for both models. The ontological and pragmatic aspects of resemblance imposed by [ModRep] are being fulfilled by both models. If this is the case, however, it is unclear how relevant resemblance of a model with its target (or lack of it) is determined. Is there some kind of threshold that determines whether relevant resemblance takes place? Can we say that the naïve models resemble the US economy? Mäki could still bite the bullet and reply that this ought to be determined by the purposes the agent has, and that for this reason [ModRep] leaves this unspecified. However, it should be noted that in both cases, modellers regarded the cut-off mountain height as properties of the model and target as relevant for their purpose and therefore they were meant to be represented as realistically as possible. Only later they discovered that less resemblance than they had originally intended returned better predictions.

The problem here for Mäki's account is that it fails to accommodate cases like the above. While the account is meant to have a placeholder for any purpose, it cannot deal with those cases in which less resemblance delivers better results. If the account is to be saved by insisting that both models do resemble their targets, then it is unclear what different degrees of resemblance do or whether resemblance is at all necessary, which also creates problems for the account.

A way to give some room in Mäki's account to these purposes is to make a distinction between the epistemic purpose *P* of a model and the practical goals *G*. The idea is that the purpose *P* is related to the representational relation—or the one without which the representational relation would not obtain—and the goals are of a practical nature. I use purposes *P* for the epistemic purpose and goals *G* for practical purposes for mere convenience.

The distinction is important because it allows us to take into account the fact that models are not used for a single purpose. Often, models are used for goals that piggyback on the epistemic purpose. So, in the example of the weather forecast models above, the epistemic purpose was to learn about the causal determinants of the weather. The aim of improving the weather forecast predictions piggybacks on this epistemic purpose.

Accounting for multiple aims and distinguishing them between epistemic and practical has a few advantages. First, it is a more accurate description of how models are used and thus helps with my attempt to understand modelling as a process. The advantages of the process view of modelling were discussed in the previous chapter. Second, in an account of models in which representation and epistemic purposes are prominent, such as Mäki's, the introduction of goals as a different category allows for taking into account other, practical aims that are not necessarily related to representation and that may often be more easily observable than the purely epistemic purpose. Third, it allows us to disentangle two categories for which we would, at least sometimes, want to have different standards. In the example discussed above of the weather forecasts, modellers were interested in being as realistic as possible in their models in order to offer better weather forecasts. We could thus say that their goal was to have better forecasts by means of first achieving their purpose of learning about the determinants of the weather—by accurately representing them in the model. Ultimately, they gave up accurate representation of the phenomena for better predictions. So, we can say that they succeeded in their goal but not in their purpose.

Let me illustrate these advantages with another example. Take again the Fed-Penn-MIT model that I briefly mentioned in chapter four. Recall that this large-scale model[11] was the outcome of a collaboration between the Federal Reserve, the University of Pennsylvania, and the MIT. So, a collaboration between an institution in charge of carrying out policy and academia. According to Backhouse & Cherrier (2017), the model had several aims. First, the model was intended to show quantitatively the structure and dynamics of the economy. There was a specific interest in representing the monetary sector, since existing models at the time did not have this feature. This is why the Fed commissioned this model in the first place. Model simulations determined whether this goal was satisfactorily achieved. This can be regarded as their epistemic purpose: to represent the structure and dynamics of the economy, including the monetary sector, in order to gain understanding of those dynamics and, specifically, learn about the potential effects of different monetary policy scenarios. Second, Modigliani and Ando were interested in resolving a theoretical controversy that emerged with the rise of monetarism. In 1963, before the model was commissioned by the Fed, Friedman, trying to revive the quantity theory of money, had published a paper with Meiselman in which they showed that the correlation between money and private consumption was higher and more stable than the Keynesian multiplier, implying that monetarism, rather than Keynesianism, was the correct macroeconomic theory. Ando and Modigliani's response had been to argue that they had failed to specify correctly the variables involved (Backhouse & Cherrier, 2017) and attempted to solve the controversy using the commissioned model. Third, the Fed aimed to carry out a monetary policy that was consistent with the objective to have low and stable unemployment (Rancan, 2017). Fourth, the Fed had a political goal, namely to demonstrate that independent, monetary policy could well counteract unemployment. In 1951, the independence of the Fed from the government had been declared, but the new theory of finance, as well as monetarism undermined the role that monetary policy could have in that front (Cherrier, 2017; Rancan, 2017). These three latter aims can be regarded as practical goals since they are not (directly) concerned with learning about the phenomenon. As such, these goals can—and should—be independently appraised. Each requires different standards of appraisal. Treating them

[11] It is difficult to determine whether this model can be considered a single model. During its construction, which lasted from 1966 until 1970, different groups with different specialties worked on independent sub-models that would later be put together. In addition, the Fed worked on an aspect of the model that was to remain secret and was therefore not shared with the rest of the group (See Backhouse & Cherrier (2017) for details).

independently offers a more comprehensive understanding of the model and the contributions it made in different fronts.

There are some important lessons in terms of model failure that we can draw from this example. Above I mentioned that the model pursued a number of goals in addition to the epistemic purpose. This is by itself an important aspect to consider in terms of model failure because it shows that a model does not simply fail or succeed with respect to a single purpose or goal. Insofar as there are different goals that are somehow independent from the epistemic goal, these are more or less observable, and they have different measures of appraisal, a single model might simultaneously fail and succeed. The Fed-Penn-MIT model could thus be a success for the simulations it made and the monetary policy that followed based on these simulations, while perhaps it was considered a failure in the theoretical contributions it made.

In fact, Backhouse & Cherrier (2017), argue that some of the goals of the model were contradictory, suggesting that they were not all simultaneously attainable. They mention that there was a point in the process in which the dynamic simulations, estimates of GDP and unemployment based on forecasts of the model of previous periods, improved if current income was dropped from the consumption equation. While the economists at the Fed were happy to trade predictive accuracy for theoretical consistency, the academic economists were not as satisfied: "I am surprised to find that in these equations you have dropped completely current income. Originally this variable had been introduced to account for investment of transient income in durables. This still seems a reasonable hypothesis" (Modigliani, quoted in Backhouse & Cherrier (2017)).

An implication is thus that to talk about model failure simpliciter is misguided and of model failure with respect to its epistemic purpose is incomplete, if considered as an appraisal of the model as a whole. Even when a model might be thought to have failed epistemically—or, for that matter in its capacity to accurately represent a target—it is misguided to regard the model a failure. That models pursue epistemic purposes and practical goals and that they perform differently at each of them is probably a good explanation for why some models have been regarded as failures and yet continue to be used by some economists—e.g. DSGE models. Furthermore, a single model is used many times in relatively different contexts and by different model users. Economists such as Gilboa, Postlewaite, Samuelson, & Schmeidler (2014) and Rodrik (2015), who have discussed the way in which models are used have defended this view, as much as philosophers such as Kuorikoski & Ylikoski (2015) and Ylikoski & Aydinonat (2014). This can count as well as a model with a specific epistemic purpose which satisfies several goals at different times.

Obviously, this is not to suggest that models seldom fail or that they always succeed at something. The point is rather that understanding how models are used by economists within academia as well as in many other institutions, implies acknowledging the multiple goals they are sometimes meant to accomplish as well as the nature of these multiple goals.

Multiple goals being simultaneously pursued, together with conflicting incentives of the model user, surely renders models and the modelling practice complex and, perhaps, some might argue, unnecessarily so. It seems to me that insofar as these complexities are likely to be the sources of failure, there's no other way than to try to analyse them and make sense of them.

5.3. An expanded Context X

Mäki recognises the importance of a context; it is the last addition to his account. The elements that he suggests should be accounted for in this category are "lots of various further ingredients that make a difference for models and modelling practices" (2017, p. 16), which include items such as "intra-disciplinary conventions and practices, standards and incentives, arrangements of education, research and publishing, and so on" (2017, p.16). These are all elements that indeed make a difference to the models produced by the discipline. But these elements that pertain primarily to the academic environment. In the examples that he offers as possible sources of model failure in the domain of the context, Mäki mentions the deficiency of offering adequate commentary of models, which in turn is possibly caused by the narrow way in which economists are allegedly educated; a disciplinary fracture between macroeconomics and finance, which in turn leads to a failure to account for financial aspects in macroeconomic models, and a specific set of epistemic values and conventions underlying the economics practice that favour the use of some controversial kinds of idealisations or techniques at the expense of others.

However, as the examples that I discussed above illustrate, model users are not confined to academic institutions. Above I discussed the role of analysts at rating agencies in prioritising private gains over model performance and economists at the Fed who had a political interest aside from developing monetary policy. This demonstrates that the elements of the context that are relevant for a philosophical account of models are much broader than its presupposed by philosophical accounts. The category of context should be able to accommodate a larger context than the purely academic one.

Obviously, these elements that belong to the context might be of very different kinds and therefore can't be all determined a priori. How do we know how broad the category has to be? Where to begin? My proposal is to start by exploring the historical context in which models have been developed. The distinct historical circumstances in which models have been developed will be able to uncover those aspects that have shaped the model, its epistemic purpose, and its practical goals, and which give the model its epistemic and practical significance.

Biddle & Winsberg (2010) have argued something similar with respect to climate models. Specifically, they argue that climate models, and particularly that at which they are good at, such as forecasting global surface temperature, are determined by (extra-epistemic) historical circumstances under which they are built. Climate models consist of a number of modules that, according to a process of trial and error, are assembled together. Some parts of the process are principled, but others are "kluges", which means that they are assembled in a specific way because they are functional, but not because they follow a specific theoretical rationale. Therefore, two models that begin with the same basis, but have been enlarged with a slightly different order of modules, will perform differently with respect to say, predicting the global mean surface temperature. Model performance is therefore path-dependent on how that model has been developed[12]. And, Biddle & Winsberg argue, how a model is developed is determined by decisions by modellers to emphasise certain prediction tasks over others.

Let me now offer three brief concrete examples of important aspects of models that can only be known if the historical circumstances of the model are considered. This, in turn, has implications for how the model is assessed. First, take again the risk models at rating agencies. Until 2004, Moody's used a "diversity score" as part of its ratings procedure. This score prevented structured financial products like CDOs from repackaging the same kind of collateral if they were to get the highest rating. This means that a CDO would not get a triple A rating if it consisted only of mortgages. Moody's was the only one of the three rating agencies that used this score and scraped it when it became clear that Fitch and S&P were getting more clients because it was easier to get higher ratings with them (Jones, 2008). Second, take again the Fed-MIT-Penn model. As I suggested above, this model began to be built in 1964 and only until 1970 there was something that could be called *the* model. Before then, different groups of individuals were working on parts of the model depending on their expertise. Each of these groups worked according to certain restrictions, like the way in which the data should be treated and the way in which parameters were to be estimated. The idea was to later assemble all the working parts into a large model and the

[12] Instead of using the term path-dependence, Biddle & Winsberg use Wimsatt's notion of "generative entrenchment" employed by Wimsatt to explain the relationship between biological development and evolution. A "generatively entrenched feature of a structure is one that has many other things depending on it because it has played a role in generating them" (Wimsatt, 2007, p. 133, quoted in Biddle & Winsberg 2010).

sub-models had thus to have a common ground with the other parts[13]. Third, take the famous Solow growth model. In this model, as it is well known, it is technological progress and, specifically the labour-augmenting technological progress, that allows for per capita growth. Without technological progress, the model states, per capita growth declines given that population grows at a constant rate, for any given savings rate, which is also exogenous. The key in this model is, therefore, technological progress: a constant savings rate is not sufficient to have per capita growth. An interesting feature about this model is that although the individual decision-making process that determines the savings rate is not modelled, the model result is consistent with those models in which the savings rate is determined within the model. More specifically, insofar as individual behaviour is assumed to lead to a path of savings that is consistent with smooth consumption over time, formally it can be shown that the results obtained from the Solow model are substantially similar to those models that do model individual saving and investment decisions (Athreya, 2013, Chapter 5).

The general point is the following. The three kinds of models that I just discussed, in isolation, only as an end-product, fail to convey information about the model that is crucial to determine the epistemic purpose of the model, its practical goals, and some of the incentives of agents using them. In other words, the model in its 'final stage' and whatever its relation to the target doesn't convey enough information to accurately appraise the model. A look at the history of the model will determine which elements are important for the performance of the model and therefore belong in a particular category. In the case of Solow's model, for instance, to know that its result is substantially similar to models in which the individual decisions are endogenous, suggests that the model holds under a wider set of conditions. We know that the model result holds when savings rates are determined exogenously and also when they correspond to a certain individual decision making process. This wouldn't be the case if the model alone were analysed. This information, as part of the context of the model, is very likely to be helpful in judging its epistemic import. It is also likely to help in judging the conditions in which the model can be 'safely' used as a quick shortcut—without having to model explicitly household decisions—given that we know that its results are consistent with certain household decisions.

Likewise, an exploration of the historical context of the model will often shed light on the elements that belong in the other categories. The Fed-MIT-Penn model shows that there were different groups and agents that had different goals with respect to what the model was meant to

[13] For an example, see a discussion by Duesenberry & Klein (1965) about how different parts of the Brookings model, a predecessor of the Fed-MIT-Penn were to be integrated.

accomplish. Likewise, in the case of the rating agencies, knowing the fact that a diversity score was considered important by Moody's at some point, suggests that private incentives probably did have an effect on how models were used.

Conclusions

In this chapter I argue that analysis of model success doesn't serve also as an analysis of failure and that, in consequence, there is a need for an explicit analysis of model failure. Furthermore, I argue that an analysis of model failure in particular demands a pragmatic approach that views modelling as a process and concerns with how models are used. Commentators have recognised, even if to a limited extent, the importance of this process. The problem, however, is that they have generally focussed only on a part of the process, treating it in the abstract, and have favoured the study of successful modelling: those cases that are considered exemplary, generally because the model has epistemic import, in one way or another.

In order to better understand the potential sources of failure, my proposal here has been to amend Mäki's account of models. I argue for the inclusion of three elements. First, I the identity of the model user. While Mäki and other commentators acknowledge the importance of the agent in the representational exercise, the underlying assumption is that this agent has only an epistemic purpose when using models. I argue that, especially nowadays, model users are not only confined to academic institutions, but are part of other organisations whose business is other than purely understanding a phenomenon. Examples I provide above are model users at the Federal Reserve and at rating agencies. These, in turn, may cause the agent to have conflicting incentives that affect the modelling exercise. For this reason, it is important to account for the incentives that might be driving the modelling exercise. I call this the identity of the model user.

Second, practical goals in addition to the epistemic purpose. Maki's, as well as other accounts of models do recognise that models are used for a specific purpose. Accordingly, the representational relation of the model with the target is established. However, models generally have, in addition to its epistemic purpose, practical goals. These also drive the modelling exercise and help determine the standards according to which a model is appraised.

Third, the historical context of a model. Investigating the history of a model is useful in two fronts. It helps to shed light on elements that might be relevant to add in the other categories and, simultaneously, offer clues about what are the elements of the context of the model that are helpful for its appraisal and its possible sources of failure.

This attempt has some loose ends. I haven't specified, for instance, whether the elements I have discussed are sufficient or necessary for an analysis of model failure. Neither have I specified what precisely is model failure and how to recognise it. Is it even possible to make sense of model failure, with capital m and capital f? While I can see how this might be disappointing for some readers, I think that such an analysis brings us far already in grasping the extent of other dimensions of the modelling practice that until now have been underestimated by philosophers concerned with models.

References

Athreya, K. B. (2013). *Big ideas in macroeconomics: a nontechnical view*. Cambridge, Mass.: MIT Press.

Backhouse, R., & Cherrier, B. (2017). The Ordinary Business of Macroeconomic Modelling: Working on the MIT-Fed-Penn model (1964 - 1974). Presented at the The History of Macroeconometric Modelling, Utrecht University.

Barnett-Hart, A. K. (2009). *The story of the CDO market meltdown: An empirical analysis*. Retrieved from http://www.valueplays.net/wp-content/uploads/2009-CDOmeltdown.pdf

Benmelech, E., & Dlugosz, J. (2010). The Credit Rating Crisis. In D. Acemoglu, M. Woodford, & K. S. Rogoff (Eds.), *NBER macroeconomics annual* (Vol. 24, pp. 161–208). University Of Chicago Press. Retrieved from http://www.nber.org/chapters/c11794

Biddle, J., & Winsberg, E. (2010). Value Judgements and the Estimation of Uncertainty in Climate Modeling. In P. D. Magnus & J. Busch (Eds.), *New Waves in Philosophy of Science* (pp. 172–197). Palgrave-Macmillan.

Blaug, M. (2001). No history of ideas, please, we're economists. *The Journal of Economic Perspectives*, *15*(1), 145–164.

Boumans, M. (1999). Built-in Justification. In M. S. Morgan & M. Morrison (Eds.), *Models as Mediators: Perspectives on Natural and Social Science* (pp. 66–96). Cambridge University Press.

Boumans, M., & Davis, J. B. (2010). *Economic Methodology: Understanding Economics as a Science.* Basingstoke [England]; New York: Palgrave Macmillan.

Cherrier, B. (2017, March 15). The ordinary business of macroeconometric modeling: working on the MIT-Fed-Penn model (1964-1974). Retrieved 13 September 2017, from https://beatricecherrier.wordpress.com/2017/03/15/the-ordinary-business-of-macroeconometric-modeling-working-on-the-mit-fed-penn-model-1964-1974/

Cherrier, B., & Backhouse, R. (2016). The age of the applied economist: the transformation of economics since the 1970s. *Open Science Framework*. https://doi.org/10.17605/OSF.IO/FGRJF

Christ, C. (1951). A test of an econometric model for the United States, 1921-1947. In *Conference on business cycles* (pp. 35–130). NBER. Retrieved from http://www.nber.org/chapters/c4760.pdf

Claveau, F., & Vergara Fernández, M. (2015). Epistemic Contributions of Models: Conditions for Propositional Learning. *Perspectives on Science*, 405–423. https://doi.org/10.1162/POSC_a_00181

Csaba, L. (2017). Comparative economics and the mainstream. *Economics and Business Review*, *3 (17)*(3), 32–51. https://doi.org/10.18559/ebr.2017.3.3

den Butter, F. A. G., & Morgan, M. S. (Eds.). (2000). *Empirical Models and Policy-Making: Interaction and Institutions*. London: Routledge.

Duesenberry, J., & Klein, L. (1965). Introduction: The Research Strategy and its Application. In J. Duesenberry, G. Fromm, L. Klein, & E. Kuh (Eds.), *The Brookings Quarterly Econometric Model of the United States*. Rand McNally & Company.

Fotheringham, W. (2017, June 24). Chris Froome's Tour de France rivals? Porte, Quintana, Contador and Bardet | William Fotheringham. *The Guardian*. Retrieved from http://www.theguardian.com/sport/blog/2017/jun/24/chris-froome-tour-de-france-rivals-porte-quintana-contador-bardet

Giere, R. N. (1990). *Explaining Science: A Cognitive Approach*. University Of Chicago Press.

Giere, R. N. (2004). How Models Are Used to Represent Reality. *Philosophy of Science*, *71*(5), 742–752. https://doi.org/10.1086/425063

Giere, R. N. (2006). *Scientific perspectivism*. Chicago, Ill.; Bristol: University of Chicago Press.

Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese*, *172*(2), 269–281. https://doi.org/10.1007/s11229-009-9506-z

Gilboa, I., Postlewaite, A., Samuelson, L., & Schmeidler, D. (2014). Economic models as analogies. *The Economic Journal*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/ecoj.12128/abstract

Jones, S. (2008, October 17). How Moody's faltered. Retrieved 27 September 2017, from https://www.ft.com/content/65892340-9b1a-11dd-a653-000077b07658

Knuuttila, T. (2005). *Models as epistemic artefacts toward a non-representationalist account of scientific representation*. Helsinki: Department of Philosophy, Univ. of Helsinki. Retrieved from http://urn.fi/URN:ISBN:952-10-2798-3

Kuorikoski, J., & Ylikoski, P. (2015). External representations and scientific understanding. *Synthese*, *192*(12), 3817–3837. https://doi.org/10.1007/s11229-014-0591-2

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton University Press.

Maas, H. (2014). *Economic methodology: an historical introduction.*

Mäki, U. (2005). Models are experiments, experiments are models. *Journal of Economic Methodology*, *12*(2), 303–315. https://doi.org/10.1080/13501780500086255

Mäki, U. (2009). MISSing the World. Models as Isolations and Credible Surrogate Systems. *Erkenntnis*, *70*(1), 29–43. https://doi.org/10.1007/s10670-008-9135-9

Mäki, U. (2017). Modelling Failure. In Hannes Leitgeb, I. Niiniluoto, P. Seppälä, & E. Sober (Eds.), *Logic, Methodology, and Philosophy of Science: Proceedings of the Fifteenth International Congress*. College Publications. Retrieved from https://pdfs.semanticscholar.org/5332/6e9790dc24be8d3597ec98b8a2fdda541bde.pdf

Morgan, M. S. (2000). Experiments Without Material Intervention: Model experiments, virtual experiments and virtually.

Morgan, M. S. (2005). Experiments versus models: New phenomena, inference and surprise. *Journal of Economic Methodology*, *12*(2), 317–329. https://doi.org/10.1080/13501780500086313

Morgan, M. S. (2012). *The World in the Model.* Cambridge University Press.

Northcott, R. (2017). When are purely predictive models best? *Disputatio*. Retrieved from http://eprints.bbk.ac.uk/18061/

Rancan, A. (2017). Notes on the Theoretical and Political Meaning of the MPS Model (1966 - 1970). Presented at the The History of Macroeconometric Modelling.

Rapley, J. (2017, July 11). How economics became a religion | John Rapley. *The Guardian*. Retrieved from http://www.theguardian.com/news/2017/jul/11/how-economics-became-a-religion

Reason, J. (1990). *Human Error*. Cambridge University Press. https://doi.org/10.1017/CBO9781139062367

Reiss, J. (2007). Do we need mechanisms in the social sciences? *Philosophy of the Social Sciences*, *37*(2), 163–184.

Rodrik, D. (2015). *Economics rules: why economics works, when it fails, and how to tell the difference* (First edition). Oxford ; New York: Oxford University Press.

Ross, D. (2014). *Philosophy of economics*. Palgrave Macmillan.

Vergara-Fernández, M. (2017). Rejoinder to Comparative economics and the mainstream by László Csaba. *Economics and Business Review*, *3 (17)*(3), 138–142. https://doi.org/10.18559/ebr.2017.3.9

Ylikoski, P., & Aydinonat, N. E. (2014). Understanding with theoretical models. *Journal of Economic Methodology*, *21*(1), 19–36. https://doi.org/10.1080/1350178X.2014.886470

6

Conclusions

There's little doubt that philosophers of science have become attentive to scientific practice in the last decades. Philosophical claims are generally based on case studies in one or more sciences that speak for the descriptive accuracy of the enterprise. Furthermore, philosophers welcome the participation of practitioners at conferences, in the attempt to learn from them and establish links with the practice. In this respect, in comparison with logical positivism, philosophy of science has become much more practice-oriented.

In this dissertation I have argued, however, that this attention to practice has come with what I've called an optimistic bias. In the philosophical accounts of models in particular, this bias has been reflected in the incessant attempt to explain the mystery of models: their success as vehicles of scientific knowledge, despite their patent falsities. This bias, in turn, limits the questions that we ask and influences the conclusions we draw. I argue for a look at scientific practice that leaves behind the baggage of the optimistic bias and takes the study of scientific practice as an end in itself. This spirit is already present in some areas in philosophy of science—e.g. philosophy of science in practice, though it remains heterogeneous—and my claim is that more of this approach should be brought to the extant philosophical accounts of models. Two main aspects that I highlight in this respect are the need to shift the unit of analysis of models from individual models to clusters of models that reflect research questions and the need to address model failure explicitly in our philosophical accounts of models.

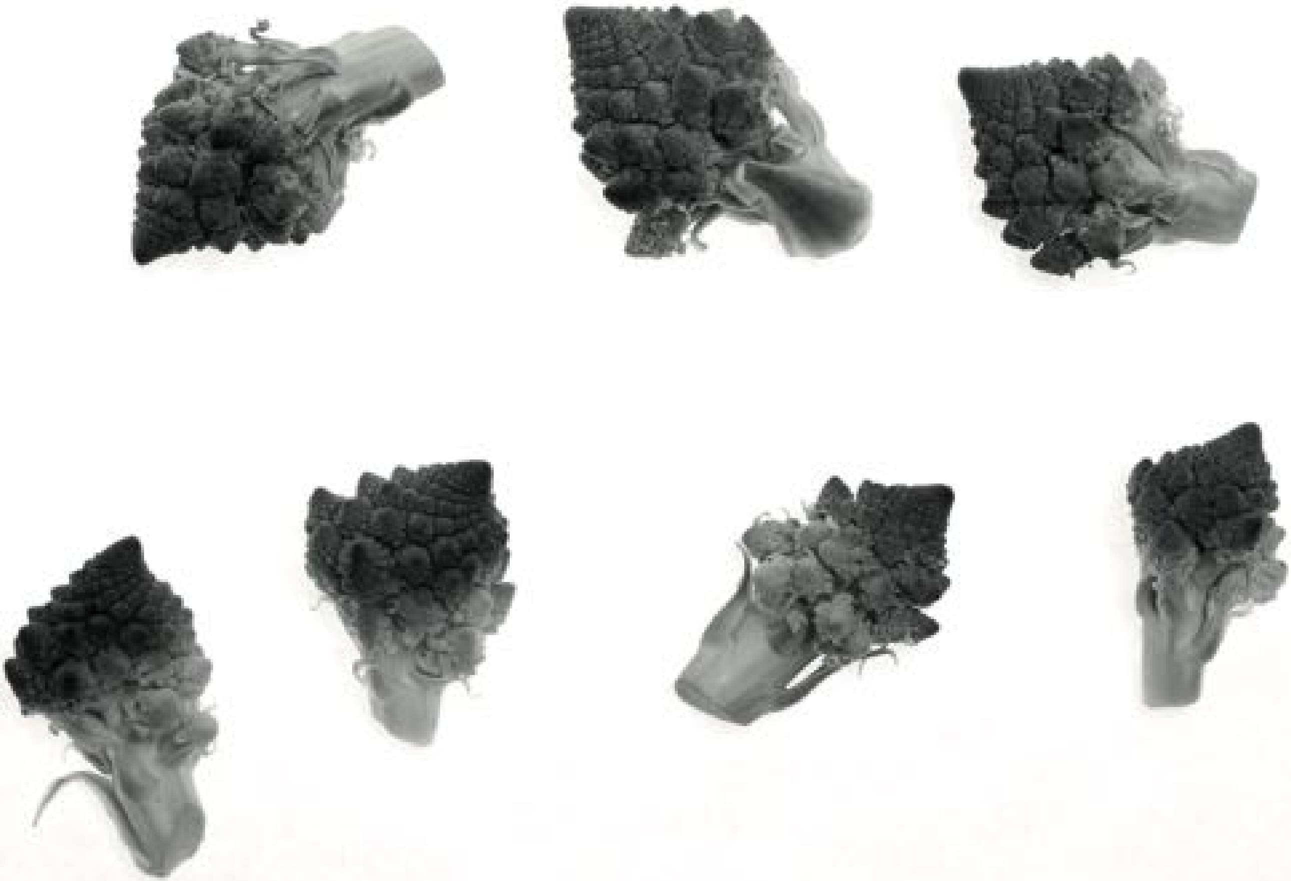
There are at least two ways in which my project might be considered to be incomplete. On the one hand, I haven't offered a coherent account of models that incorporates each of the elements that I discuss on each of the chapters. In other words, I've mostly flagged elements of the modelling practice that have not been picked up by extant philosophical accounts of models and that I argue a relevant for an accurate and comprehensive understanding of the practice. For instance, I have not said much about how a pragmatic account of models that is capable of accommodating the potential sources of failure that I highlight in the last chapter, concretely relates to the view advocated in chapters two and three that models ought to be analysed as clusters. In fact, Mäki's account, the one I offer amendments for, is an account that has been thought for an analysis of a single model. One could thus argue that, by amending Maki's [ModRep] account of models, I endorse the analysis of single models that at the outset I reject. My answer to this is still what I claim in chapter two: we need more analyses of more models. Only a more extensive survey of the use of models will be able to uncover whether a general philosophical account (that is capable of accommodating failure) of models is actually possible, as some commentators presume, or whether such an approach is a misguided enterprise, as some others have at least implicitly suggested, by endorsing an approach that refrains from making general claims about models, but instead try to uncover the nuances of individual cases. . I think there is merit in both approaches. But to defend either requires a more extensive survey. It can't be yet claimed that philosophical accounts of models, as currently offered, cover all the important aspects there are to modelling or even that they are helpful to understand current scientific practice. Above I argued how limited they are. But it can't be claimed either that such an attempt is a futile exercise and therefore shouldn't attempt to establish general accounts of models. My contribution here, perhaps more negative than positive, has been to argue why the general accounts, as they currently are, are unsatisfactory.

On the other hand, to accept my claim that the modelling practice ought to be studied as it is, means that there are many more aspects that need to be studied than I what could possibly engage with in this dissertation. One that has become more pressing in the last years, for instance, is how big data may affect modelling practices. In this dissertation I have tried to emphasise that theoretical and statistical models are usually used together and that analysing theoretical models in isolation is a misguided enterprise, particularly as a reflection of the discipline. A pressing question is whether statistical models will continue to have the significance they have in conveying information about populations if and ever big data can identify the sort of patterns economists and other social scientists are interested in. This is not a question that philosophical accounts of models, if focussed only on theoretical models can answer. Another is whether, given current trends in which potentially significant data is owned by private companies such as Google or Apple, national statistical departments, on which many social scientists rely to get their data, are likely to become redundant or, worse, totally unrepresentative. Or, in relation to what I argued before about the incentives of modellers, assuming that private companies gain more prominence in the social scientific domain as the rightful owners of data, whether this is likely to generate incentives that conflict more often. Most likely some of these questions are being addressed in other areas of philosophy and science and technology studies. The philosophical accounts of models in economics could definitely profit from those analyses in order to guide its explorations of the modelling practice.

I introduced this dissertation claiming that philosophers of economics didn't take advantage of the greatest recession since 1929 to sell some philosophy of models. Clearly though, that claim is just partly true. Philosophers of economics haven't produced much philosophy of models *for economists* or about model failure, but the philosophy of economics, within general philosophy of science, has never been in a better state. The philosophy of economics has become increasingly recognised within philosophy of science, attracting more and more students, becoming more present in the field with participation in journals and conferences, and even job openings in philosophy departments. This is clearly a good thing.

My attempt here has been to suggest that, particularly if one endorses a view of a socially responsible philosophy of science, a more comprehensive look at the modelling practice, one that is interested in all philosophical issues that arise from it, might not have to wait for the next economic crisis to become attractive for economists and other social scientists. That, I think, is a good thing too.

Additional Material



Summary

Since the beginning of the twentieth century, in the search for objectivity in the social sciences, scientists have given models an increasingly important role. Economics is nowadays acknowledged as a model-based science and other social sciences such as sociology, which used to lean more towards interpretive methodologies, have also become more model-oriented. Similarly, the attempt to remove discretion from public policy and enhance instead objective, evidence-based policy[1], has also driven policy makers and governmental institutions to rely on models, empirical and theoretical.

Up until the seventies, philosophers of science had given little attention to models. These were mostly thought of as the result of imperfect, ongoing thinking that would become redundant once a proper theory was established. Later, thanks to an explicit attempt to offer philosophical accounts of the practice that were descriptively accurate, philosophers recognised the ubiquity that models have in scientific practice and acknowledged their significance: models, instead of theories, are what ultimately represent phenomena and carry scientific knowledge.

In this dissertation I start from the premiss that, since the significance of models in the scientific enterprise was acknowledged, the main purpose of the literature has been to solve a mystery that models pose: they are often capable of yielding understanding about phenomena despite their evidently false assumptions. I argue that the attempt to solve this mystery has generated philosophical accounts of models that suffer from an *optimistic bias*. That is, models are presupposed to have epistemic import and the attempt has been to uncover how this is possible. In general, three perspectives have been taken. Accounts that attempt to explain model success by the kind of entities they are (ontology), by the kind of representational relation they have with their targets (semantics), and relatedly, explore the kind of epistemic import models afford and how they accomplish this (epistemology).

Furthermore, I argue that the attempt to solve the mystery of models has narrowed the lens through which the modelling practice is observed. In other words, accounts of models are descriptively accurate only insofar this conduces to solving the mystery. But important aspects of the modelling practice have been understated at best or ignored at worst. I argue for a look at the modelling practice as an end in itself, without the baggage imposed by the mystery of models. This

[1] Here I'm not only referring to the evidence-based policy (EBP) movement that has emerged in the last couple of decades whose main attempt is to test causality claims, but to the much longer tradition that became quite prominent in the USA with the reformers at the end of the nineteenth century.

fresh look, in turn, raises philosophical questions that are both important and relevant for a socially responsible philosophy of science, given the many different realms beyond the purely scientific, in which models are used.

In chapter two, I address the question of what unit of analysis of models philosophers should investigate. I argue that the philosophical investigation of models should be focussed on clusters or research questions, rather than on single models and their components, as has generally been done. I suggest that two specific philosophical questions, which attempt to solve the mystery posed by models, may have guided the interest of philosophers towards individual models and model components. One of the questions is the kind of representational relation that exists between models and their targets. The other is the role of models within the debate on realism. The modelling practice, as well as philosophical arguments that maintain that our models are incapable of fulfilling at once all the purposes we might have for them, are, nevertheless, compelling reasons to explore how models among themselves are related. Using models in the New Economic Geography (NEG) as example, I argue that models are used such that they establish what I call vertical and horizontal complementarities with other models. These are important to determine the epistemic import of models.

In chapter three, François Claveau and I discuss three epistemic roles that models might play and offer sufficient conditions for a model to actually play each of those roles. We use the traditional definition of knowledge as true justified belief (KATJB) as a basis to define learning and thus establish these conditions. The motivation of the chapter is that, while there has been a long interest from philosophers to defend the epistemic success of models, there has seldom in the literature been a clear definition of what precisely this epistemic benefit is. Having defined learning as 'coming to know', the attempt here is thus to establish the sufficient conditions that a model would have to fulfil in order to determine whether it can be said to have epistemic benefit. The three epistemic roles we discuss are, evidential, which states that models can count as evidence for a claim about the world; stimulating, which states that models can be a stimulus for carrying out empirical research, and revealing, which states that models can generate new hypotheses about the world.

Although this chapter might be considered to fall prey to the attempt to solve the mystery of models that I have criticised, it has three important features that correspond to the fresh look at the praxis that I defend. First, it is an exploration of a cluster of models or, rather, of a research question, instead of a single model. In this sense, it can be taken as an example of what was argued in chapter two. We use the Diamond-Mortensen-Pissarides (DMP) model[2], which is part of the search and matching theory that has been developed in economics since the 70s, as a case from which we pick out the potential epistemic roles models can have. Second, since our analysis is on a research question rather than on a single model, we are able to rely on the relations that exist between purely theoretical models, statistical models, and data. So, in our example, justification for believing a particular proposition about the world coming from the model is possible thanks to a network of beliefs that agents derive from other sources such as empirical data. In this sense, models are not mysterious at all but rather another tool, among many, that are used to understand the world. Third, it is an enquiry in the field of labour economics, which has important relations with and implications for macroeconomics. This is a field that has seldom been investigated by philosophers of economics concerned with modelling.

In chapter four I move to a different subject. In an attempt to make sense of what it means for macroeconomic models to have failed, which many commentators have argued in light of the financial crisis, I survey the philosophical literature for guides as to how the accounts that have been offered so far, can elucidate these claims of failure. My conclusion is that there is little that has been explored by philosophers with respect to model failure and little in their accounts that can be used for making sense of this aspect of modelling. This exercise suggests that, despite the great interest that models have received from philosophers, especially due to the acknowledgement that models play a major role in scientific practice, the reach of the literature has been constrained to comprise only three aspects. One of them is the almost exclusive focus on theoretical models (as individual units). This is particularly remarkable in economics, given the transformation that the discipline has gone through in the last decades to a more empirical (or applied) science. This transformation is not reflected in our philosophical accounts of models and raises important questions about how close philosophical enquiry actually is to practice. The other two are the focus on explanation and understanding, and on the identification of causal mechanisms. Surely these aspects are important for science and the use of models more generally. But they are not the only aspects driving science and the use of models. In studying only these, other aspects of the practice are underestimated.

Finally, in chapter five, I continue with the subject of model failure. I argue for the need of an explicit analysis of model failure and, specifically, for a pragmatic account of models. Such a pragmatic account, I argue, is capable of accommodating aspects that determine the outcomes of

[2] The DMP model is known as a model (singular). However, strictly speaking it is a class of models that were developed throughout the years by the three economists mentioned above.

the modelling activity, and that cannot be accommodated by extant accounts of models. I discuss Uskali Mäki's (2017) account of models "[ModRep]" for two reasons. First, because, though introduced as an account of representation, it has been extended over the years to include pragmatic elements. Second, and more importantly, because Mäki (2017) suggests that [ModRep] is sufficient to accommodate model failure. I argue, based on a few examples of the practice, that some potential sources of failure cannot be accounted for by [ModRep] and therefore suggest three additional elements. First, the incentives that the model user might have. The idea here is that depending on the identity of the model user, they will have different (private) incentives. Second, the goals that agents might have with their models. Here I make a distinction between epistemic purposes, which is part of [Mod Rep], and practical goals. Third, the historical context of the model. This element recognises the importance of previous uses and purposes in shaping current use and potential failure.

Samenvatting

Sinds het begin van de twintigste eeuw hebben sociale wetenschappers, in hun poging de objectiviteit van sociale wetenschap op een of andere manier te garanderen, steeds intensiever gebruik gemaakt van modellen. De economie wordt tegenwoordig gezien als een modelgebaseerde wetenschap en andere sociale wetenschappen, zoals de sociologie die toentertijd meer een interpretatieve inrichting insloeg, bedienen zich nu ook toenemend van modellen. Evenzo heeft de ontwikkeling van discretionair beleid naar een voorkeur voor objectief 'evidence based' beleid bij beleidsmakers geleid tot een vergroot vertrouwen in modellen, of die nou empirisch of theoretisch zijn.

Tot de zeventiger jaren was er weinig aandacht voor het modelgebruik in de wetenschappen. Deze werden gezien als het resultaat van imperfect en tentatief denken, overbodig zodra een adequate theorie zou worden ontwikkeld. Pas later – toen filosofen expliciete belangstelling begonnen te ontwikkelen voor een descriptief accurate weergave van wat er feitelijk gebeurde in de wetenschap – kwam de erkenning van de alomtegenwoordigheid van modellen in de wetenschappelijke praktijk: modellen en niet theorieën representeren de sociale verschijnselen en zij zijn de dragers van wetenschappelijk inzicht.

In dit proefschrift kies ik als startpunt de premisse dat, sinds het belang van modellen in de wetenschappelijke praktijk werd erkend, de filosofische literatuur vooral als doel heeft gehad om een mysterie op te lossen dat modelgebruik schijnt te genereren: dat modellen blijkbaar wetenschappelijk inzicht helpen geven in verschijnselen ondanks dat zij uiteraard berusten op onware aannames. Ik beweer dat de poging om dit mysterie op te lossen filosofische analyses heeft voortgebracht die onderhevig zijn aan een *optimistische bias*. Dat wil zeggen, modellen worden verondersteld ware kennis te genereren en filosofen hebben geprobeerd om te ontdekken hoe dit eigenlijk mogelijk is. Daartoe zijn drie perspectieven gehanteerd. Er zijn verhandelingen die het succes van het gebruik van modellen in de sociale wetenschap verklaren door wat modellen zijn (de ontologische benadering), door welk verband zij onderhouden met hun bedoelde referent (de semantische benadering) en, ten slotte, door de wetenschappelijke kennis die modellen genereren en de manier waarop zij deze genereren (de epistemologische benadering). Bovendien beweer ik deze pogingen om het mysterie op te lossen de lens, waardoor de praktijk van modelleren wordt bezien, vernauwen. Anders gezegd, beschrijvende adequaatheid van een studie van modelgebruik kan alleen komen als deze het mysterie oplost.

Maar heel belangrijke aspecten van het gebruik van modellen zijn in het beste geval onderschat en in het slechtste geval compleet over het hoofd gezien. Daarom kies ik voor een benadering die het onderzoek naar modelgebruik als een doel op zichzelf stelt., zonder de bagage van het hierboven genoemde mysterie van modellen. Zo'n frisse blik levert namelijk nieuwe filosofische vragen op die er toe doen als het gaat om een sociaal verantwoordelijke wetenschapsfilosofie. Zeker gezien de vele gebieden buiten de academische wetenschap waarin modellen in gebruik zijn.

In hoofdstuk 2 ga ik in op de vraag wat het object van filosofische analyse behelst. Mijn visie is dat zo'n analyse gericht moet zijn op clusters, of op onderzoeksvragen, in plaats van op de gebruikelijke enkelvoudige modellen en hun componenten. Ik denk dat twee heel specifieke filosofische vragen, gericht op het oplossen van het mysterie, hebben geleid tot een beperking tot enkelvoudige modellen en modelbouwstenen. De eerste vraag is welke afbeeldingsrelatie er bestaat tussen modellen en hun referenten. De tweede betreft de rol van modellen in het realismedebat. Toch geven zowel de wetenschappelijke praktijk van modelgebruik als allerlei filosofische overwegingen – namelijk dat modellen nooit alles kunnen doen dat we van ze verwachten – reden genoeg om eens uit te zoeken hoe meerdere modellen zich *tot elkaar* verhouden. Met behulp van een voorbeeld uit de Nieuwe Economische Geografie (NEG) laat ik zien dat modellen gebruikt worden om 'verticale en horizontale complementariteiten' tot stand te late komen met andere modellen. Dat is cruciaal om het epistemische belang van modellen te bepalen.

In hoofdstuk 3 gaan François Claveau en ik verder in op de vraag welke epistemische rol modellen spelen. We formuleren de voldoende voorwaarden voor zo'n epistemische rol. Daarvoor kiezen we de filosofisch traditionele definitie van kennis als *gerechtvaardigd waar geloof* als basis om *leren* te definiëren. De aanleiding voor dit hoofdstuk is dat, terwijl filosofen veel aandacht hebben gehad voor het epistemische succes van modellen, er zelden een scherpe omschrijving is gegeven van wat dit succes dan wel pleegt te zijn. Na 'leren' te hebben omschreven als 'te weten komen' proberen we de voldoende voorwaarden op te stellen voor het epistemische nut van een model. De drie epistemische rollen van modellen die we aan de orde stellen zijn: die van bewijsvoering, om bewijs te leveren voor de waarheid van een bewering over de wereld; die van stimulans, om empirisch onderzoek te entameren; en ontsluiering, de rol van modellen om onverwachte hypotheses over de wereld voort te brengen.

Nu mag het lijken dat dit hoofdstuk lijdt aan precies dezelfde kwaal van pogingen om het mysterie op te lossen die ik eerder al bekritiseerd had, het levert wel een benadering met drie eigenschappen die corresponderen met de frisse blik op de praxis die ik verdedig.

In de eerste plaats rapporteert dit hoofdstuk over een zoektocht naar de aard van modellen in clusters, d.i. naar aanleiding van een onderzoeksvraag, in plaats van naar enkelvoudige modellen. Daarom is het een voorbeeld van wat in hoofdstuk 2 werd bepleit. We gebruiken het Diamond-Mortensen-Pissarides (DMP) model, onderdeel van de *search and matching* theorie die sinds de zeventiger jaren werd ontwikkeld, als een casus waaruit we de potentiële epistemische rol, die modellen kunnen spelen, afleiden.

In de tweede plaats kunnen we gebruik maken van de verbanden tussen theoretische modellen, statistische modellen, en data doordat we onze analyse richten op onderzoeksvragen in plaats van op een enkelvoudig model. Bijgevolg is, in ons voorbeeld, de rechtvaardiging voor een van het model afkomstige propositie, die nochtans over de werkelijkheid gaat, alleen mogelijk doordat de gebruikers van het model binnen een netwerk van wetenschappelijke overtuigingen denken; en dat netwerk is weer voortgebracht door andere bronnen, zoals empirische gegevens. Zo gezien is er helemaal niets mysterieus aan modellen. Het zijn gewoon instrumenten als alle andere die ingezet worden om de werkelijkheid beter te begrijpen.

In de derde plaats betreft het DMP model een onderzoek arbeidsmarkteconomie en dit heeft implicaties voor de macro-economie waarvan we eerder opmerkten dat deze nogal onderbedeeld is in de filosofische aandacht voor economische modellen.

Hoofdstuk 4 gaat over een ander onderwerp. Het levert een onderzoek naar wat de filosofische literatuur ons leert als we willen weten hoe macro-economische modellen precies kunnen *falen*. Deze vraag doet er nogal toe sinds veel critici beweren dat economen er niet in slaagden de financiële crisis van 2008 te voorspellen of te voorkomen. Mijn onderzoek laat zien dat er bijzonder weinig door filosofen is nagedacht over dit aspect van modelgebruik. Ondanks al die aandacht voor en erkenning van het belang van modelgebruik belicht de literatuur niet meer dan drie aspecten. Eén komt voort uit de exclusieve focus op *theoretische* modellen (als op zichzelf staande eenheden). Dat is opvallend daar het de economie betreft, gegeven de transformatie die deze discipline heeft ondergaan in de richting van meer empirie en van een meer toegepaste wetenschap. In onze filosofische aandacht voor modelgebruik is weinig van deze transformatie merkbaar en dit doet de vraag ontstaan over de descriptieve adequaatheid van wetenschapsfilosofisch onderzoek. De andere twee aspecten komen voort uit, respectievelijk, een focus op verklaren en begrijpen en een focus op oorzakelijke mechanismen. Dit zijn in het algemeen belangrijke aspecten om te onderzoeken als het om wetenschap en het gebruik van modellen gaat. Maar zij zijn lang niet de enige aspecten die de machinerie van de wetenschap gaande houden en deze exclusieve focus brengt een onderschatting van het belang van andere aspecten met zich mee.

In hoofdstuk 5, ten slotte, vervolg ik mijn onderzoek naar het falen van modellen. Een veel explicietere analyse van het falen van modellen is hard nodig. Meer specifiek verdedig ik de bewering dat daarbij een meer pragmatische benadering van modellen nodig is. Zo'n benadering kan inzicht geven in wat de uitkomst van modelleren nou precies bepaalt; een resultaat dat we niet mogen verwachten van de gangbare benaderingen. Er zijn twee redenen om daarbij Uskali Mäki's (2017) verhandeling over modellen "[ModRep]" aan de orde te stellen. De eerste is dat er, al begon deze met de vraag naar representatie, gaandeweg steeds meer elementen van pragmatiek in werden opgenomen. Maar belangrijker is de tweede reden, namelijk dat Mäki zijn [ModRep] voldoende acht om het falen van modellen in kaart te brengen. Ik gebruik enkele voorbeelden uit de praktijk van economisch onderzoek om te laten zien dat potentiële bronnen van falen helemaal niet door Mäki's verhandeling kunnen worden begrepen en ik stel drie aanvullende elementen voor. Het eerste element betreft de individuele prikkels die de respectievelijke gebruikers van een model ondervinden. Deze zijn mede afhankelijk van de identiteit van deze modelgebruikers. Ten tweede zijn de doelen, die zij met het model hebben, van belang. Ik onderscheid daartoe epistemische doelen - onderdeel van [ModRep] - van praktische doelen. Ten slotte is er het element van de historische context. Hiermee kan informatie over vroeger gebruik en voormalige doelen ingezet worden om het tegenwoordige gebruik en het potentieel falen van een model te verklaren.

May 2018

MELISSA VERGARA FERNÁNDEZ

Curriculum vitae

Lecturer
University College Groningen (UCG)
University of Groningen
PO Box 1022
9701 BA Groningen
The Netherlands

melissa@mvergarafernandez.nl
s.m.vergara.fernandez@rug.nl
+31 (0) 624978161

EDUCATION

2012 - 2018 **Doctoral Researcher**
Faculty of Philosophy, Erasmus University Rotterdam – The Netherlands
- Dissertation: "The Use of Models in Economics"
- Supervisors: Prof Dr Julian Reiss (Durham University, UK) and Prof Dr Jack Vromen (EIPE, Erasmus University Rotterdam)

2009 – 2012 **MA Philosophy and Economics, cum laude**
Erasmus University Rotterdam – The Netherlands

2005 – 2008 **Art (with emphasis on electronic media)**
Universidad de los Andes, Bogotá - Colombia

2002 – 2007 **BSc Economics**
Universidad de los Andes, Bogotá - Colombia

PUBLICATIONS

- "Rejoinder to 'Comparative Economics and the Mainstream' by László Csaba" *Economics and Business Review.* 2017 3(17):3, 138-142
- "Epistemic Contributions of Models: Conditions for Propositional Learning" *Perspectives on Science.* 2015 23:4, 405-423 (With F. Claveau)
- "International Network for Economic Method, 13-15 June" 2013. *The reasoner* 7 (8). (With T. Wells)

TEACHING

2018 **University of Groningen - University College Groningen (UCG)**

- *Philosophy of the Social Sciences,* second year undergraduate course for students of the social sciences.

2016 - 2018 **University of Amsterdam (UvA) – Faculty of Economics and Business**

- *Economic Methodology,* philosophy of science, third year undergraduate course that reflects on what makes economics a science (with Dr Dirk Damsma)

2015 **Erasmus University Rotterdam – Faculty of Philosophy**

- *Advanced Philosophy and Methodology of Economics*, research master course with topics on evidence-based public policy and empirical methods in economics (with Prof Dr Julian Reiss)

2014 **Utrecht University – Utrecht School of Economics**

- *Economic Methodology*, third year undergraduate course on the historical development of methods in economics and the philosophical conundrums they raise (lectured and taught tutorials)

2012 & 2013 **Erasmus University Rotterdam – Faculty of Philosophy/School of Economics**

- *Philosophy of Economics,* third year undergraduate course on theoretical, methodological and ethical aspects of economics (with Dr Julian Reiss (2012) and Dr Conrad Heilmann (2013))

2008 **Universidad de los Andes – Alberto Lleras Camargo School of Government**

- *Ethics, Justice and Public Policy*, interdisciplinary undergraduate course on theories of justice and ethical aspects of public policy (with Daniel Castellanos)
- *The Politics of Public Policy,* master course on political economy (with Daniel Castellanos)

2008 **Universidad de los Andes – Faculty of Economics**

- *Game Theory,* second year undergraduate course (With Dr. Oskar Nupia)

TALKS (Selection)

- "On the Goals of Economics: Past and Future". VI ALAHPE Conference, Universidad de los Andes, Bogotá, Colombia. November 2017 (abstract submitted)
- "Macroeconomic Models, Context, and History". Workshop *What to Make of Highly Unrealistic Models?,* TINT, University of Helsinki, Finland. October 2017 (abstract submitted)
- "What do Philosophical Theories Say about Model Failure?". XIII INEM Conference, University of the Basque Country, Donostia, Spain. August 2017 (abstract submitted).
- "What's the Use of Philosophical Theories of Models?".EIPE20 Conference. Erasmus University Rotterdam. The Netherlands. March, 2017 (abstract submitted).
- "What do Philosophical Theories of Models Say about Model Failure?". Economics in a Post-factual Democracy Conference. Centre for Information and Bubble Studies, University of Copenhagen. Denmark. February 2017 (abstract submitted)
- "More Models". Models and Simulations 7. University of Barcelona. Spain. April 2016 (abstract submitted)

- "More Models". XII INEM Conference, University of Cape Town, South Africa. November 2015 (abstract submitted)
- "Bridging the Gap between Philosophers, Economists, and Economic Practitioners" YSI-INET Pre-conference workshop. INET annual conference *Liberté, Égalité, Fragilité.* OECD, Paris, France. April 2015 (invited).
- "The Relevance of Philosophy for Social Science and Policy". OZSW Winter School *Philosophy, Policy and Social Science.* Erasmus University Rotterdam, The Netherlands. December 2014 (invited).
- "Economic Explanations: Towards a Shift in the Unit of Analysis". Seminario CEDE. Universidad de los Andes, Bogotá, Colombia. November 2014. (invited)
- "Economic Explanations come from Clusters, not from Models". Symposium on interdisciplinary explanations in economics. PSA Biannual meeting, Chicago, Ill., USA. November 2014 (abstract submitted)
- "The Future of the Philosophy of Economics". Panel Discussion XI INEM Conference, Erasmus University Rotterdam, The Netherlands. June 2013 (invited)
- "Model Inconsistency in Economics". XI INEM Conference, Erasmus University Rotterdam, The Netherlands. June 2013 (abstract submitted)
- "Dealing with Inconsistency in Theories and Models". Graduate conference, University of Groningen, The Netherlands. April 2013 (abstract submitted)
- "On the Evidential and Heuristic Roles of Models". Models and Simulations 5, University of Helsinki, Finland. June 2012 (abstract submitted)
- "Towards More Rationalised Criticisms". IX INEM Conference, University of Helsinki, Finland. September 2011 (abstract submitted)

SERVICE TO THE PROFESSION

- Referee for: European Journal of Philosophy of Science (EJPS), International Studies in the Philosophy of Science (ISPS), Journal of Economic Methodology (JEM), and Oeconomia.
- Elected Member of the Community Report Committee (CRC). Young Scholars Initiative, Institute for New Economic Thinking (YSI - INET). April 2017 to date.
- Coordinator of the Philosophy of Economics working group. Young Scholars Initiative, Institute for New Economic Thinking (YSI – INET). May 2015 to date.
- Co-organiser of the ALAHPE pre-conference workshop "Trends and Current Methods in the History of Economics". Universidad de los Andes, Bogotá, Colombia. November 28, 2017.
- Organiser of the YSI-INEM pre-conference workshop: "Capitalism, Technology, and Scientism: Threats to Democracy?" University of the Basque Country, Donostia, Spain. August 27, 2017.
- Co-organiser of the YSI Plenary "Piecing together a paradigm". Central European University. Budapest, Hungary, October 2016.
- Co-organiser of the YSI workshop on basic income "The Future of Work". Zurich, Switzerland. 4 May, 2016
- Organiser of the workshop "Bridging the Gap between Economists and Philosophers". University of Cape Town. November 17, 2015
- Co-organiser of the INEM/CHESS/EIPE Summer School in Philosophy and Economics "Macroeconomics, Microfoundations, and Evidence-Based Social Policy". San Sebastián (Donostia), Spain. 6-8 July, 2015.

LANGUAGES

Spanish: Native
English: Near native
Dutch: Upper-Intermediate (B2 – Staatsexamen NT2-II Diploma)

HONOURS

2012 - 2015	Erasmus Institute for Philosophy and Economics (EIPE). Doctoral scholarship.
2012	Erasmus Institute for Philosophy and Economics (EIPE). MA, cum laude.
2010	Scholarship to attend the XIII summer school of the Urrutia Elejalde Foundation "Learning from the Great Recession: Failures and New Directions in Economic Theory and Policy". Donostia, Spain.
2009 – 2011	Colfuturo. Loan-Scholarship awarded to pursue Research Master at Erasmus University Rotterdam.
2007	BSc thesis "Economics Between Science and Discourse: the Scientistic Methodology of Economics" (in Spanish) nominated to the *Ulpiano Ayala Award* to the best undergraduate thesis in the year 2007.

OTHER ACTIVITIES

Uit je eigen stad (Urban farm), Rotterdam

2013	Volunteer work

Universidad de los Andes – Alberto Lleras Camargo School of Government

2008 – 2009	Editorial/Communications Coordinator
2007 – 2008	Administrative Assistant

Universidad de los Andes – Faculty of Economics

2007	Research assistant for Dr Jimena Hurtado Prieto and Dr Christian Jaramillo