London School of Economics and Political Science

How Models Represent

James Nguyen

A thesis submitted to the Department of Philosophy, Logic and Scientific Method of the London School of Economics and Political Science for the degree of Doctor of Philosophy, March 2016.

DECLARATION

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I confirm that material from Section 6.1 is to be published in *Philosophy of Science* (Nguyen, 2016). I confirm that material from Chapter 9 is my contribution to a working paper co-authored with Roman Frigg.

I declare that my thesis consists of 72,972 words.

ABSTRACT

Scientific models are important, if not the sole, units of science. This thesis addresses the following question: in virtue of what do scientific models represent their target systems? In Part i I motivate the question, and lay out some important desiderata that any successful answer must meet. This provides a novel conceptual framework in which to think about the question (or questions) of scientific representation. I then argue against Callender and Cohen's (2006) attempt to diffuse the question.

In Part ii I investigate the ideas that scientific models are 'similar', or structurally (iso)morphic, to their target systems. I argue that these approaches are misguided, and that at best these relationships concern the accuracy of a pre-existing representational relationship. I also pay particular attention to the sense in which target systems can be appropriately taken to exhibit a 'structure', and van Fraassen's (2008) recent argument concerning the pragmatic equivalence between representing phenomena and data. My next target is the idea that models should not be seen as objects in their own right, but rather what look like descriptions of them are actually direct descriptions of target systems, albeit not ones that should be understood literally. I argue that these approaches fail to do justice to the practice of scientific modelling. Finally I turn to the idea that how models represent is grounded, in some sense, in their inferential capacity. I compare this approach to anti-representationalism in the philosophy of language and argue that analogous issues arise in the context of scientific representation.

Part iii contains my positive proposal. I provide an account of scientific representation based on Goodman and Elgin's notion of representation-as. The result is a highly conventional account which is the appropriate level of generality to capture all of its instances, whilst remaining informative about the notion. I illustrate it with reference to the Phillips-Newlyn machine, models of proteins, and the Lotka-Volterra model of predator-prey systems. These examples demonstrate how the account must be understood, and how it sheds light on our understanding of how models are used. I finally demonstrate how the account meets the desiderata laid out at the beginning of the thesis, and outline its implications for further questions from the philosophy of science; not limited to issues surrounding the applicability of mathematics, idealisation, and what it takes for a model to be 'true'.

ACKNOWLEDGEMENTS

Thanks to my supervisor Roman Frigg for his constant support and encouragement whilst writing this thesis. He and I have spent many hours talking about the topic, and others, and I couldn't have asked for a better supervisor. Thanks also to other members of the department's faculty, including but not limited to, Bryan Roberts and Miklós Rédei for stimulating discussions on material related to the thesis. The LSE has been a wonderful place to work, and being part of the philosophy of science community in the University of London more generally has afforded me the opportunity to learn a lot (including how much is left to learn). Thanks also to my fellow PhD students who have contributed to a fantastic working environment and have taught me much.

I have presented material from this thesis at various conferences over the last 3 years and discussions with numerous people during, and after, them have greatly improved this thesis. Thanks to my supervisors and peers at the University of Cambridge and KCL for stimulating my interest in philosophy, and thanks to the British Arts and Humanities Research Council for supporting me throughout my post-graduate education.

Finally, thanks to my family for their limitless support. And most importantly thanks to Alice and Oscar for making me happy, every day.

CONTENTS

LIST OF FIGURES

'... In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography.'

Jorge Luis Borges, *On Exactitude in Science*.

Part I

WHAT'S THE PROBLEM OF SCIENTIFIC REPRESENTATION?

1

INTRODUCTION

It's 1953 and economists in the Central Bank of Guatemala are topping up the water tank in their recently purchased Phillips-Newlyn machine, a system of pipes and reservoirs with water flowing through it. The land reform act (Decree 900) passed in Guatemala the previous year had redistributed unused land to local farmers. US corporation Wrigley's, one of the largest buyers of Guatemalan chicle gum, had announced that it would stop imports from Guatemala in protest against the land reform. The economists working in the Central Bank are worried about a politically motivated decrease in foreign demand for Guatemalan goods, and want to know what effect such a decrease would have on the national economy, given that their currency is pegged to the US dollar. They adjust the machine to account for the macroeconomic conditions in Guatemala and proceed to let the machine reach equilibrium. They close a valve marked 'exports' and watch what happens. The effect of this is that the flow marked 'income' starts falling, and in order to keep the level of the tank marked 'foreign held balances' constant, water flows out of the machine into the spare tank. This provides the sought-after indication of the effects of falling exports on the Guatemalan economy.[1]

Wait. What the economists were manipulating were the valves and tanks of a hydraulic machine that pumped water from reservoir to reservoir. How could manipulating such a machine tell them anything about the Guatemalan economy? The answer, I think, is that the machine is (and was) a *model* that *represents* the Guatemalan economy.

It's 1957 and John Kendrew is sitting at his desk threading a 'sausage' of plasticine through a system of vertical rods. The shape of the sausage is determined by the electron density data that he and his team at Medical Research Council Unit for Molecular Biology at the University of Cambridge had recently collected from crystallised samples of whale myoglobin. Once he's done, Kendrew looks at the shape of the sausage

1 See Phillips (1950) for the original presentation of the model; Morgan and Boumans (2004) and Morgan (2012, Chapter 5) for a useful philosophical introduction; and Aldana (2011) and Stevenson (2011) for discussions about the the Guatemalan Central Bank's use of their Phillips-Newlyn machine. I discuss the example in more detail in Chapter 9.

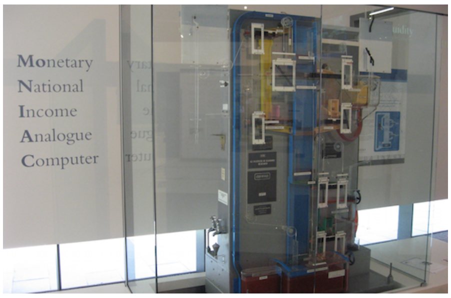

Figure 1: The Phillips-Newlyn machine on display (public domain)[3]

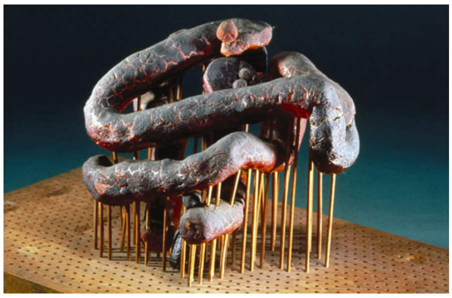

Figure 2: Kendrew's 'sausage model' of myoglobin (Science Museum)[4]

and observes that it twists and turns back on itself; it's folded in a particular pattern that is highly difficult to explain in English. He concludes that myoglobin has a particular tertiary structure, and remarks that 'the arrangement seems to be almost totally lacking in the kind of regularities which one instinctively anticipates, and is more complicated than has been predicated by any theory of protein structure' (1958, 665).[2] He goes on to win the Nobel Prize in chemistry for his work in determining this structure.

Wait. Kendrew was looking at a lump of plasticine. How could investigating the shape of such an object tell him anything about the tertiary structure of a microscopic protein molecule? The answer again is that the object is a model that represents myoglobin.

It's 1926 and Vito Volterra wants to understand why the First World War had the impact it did on the relative proportion of predator and prey fish in the Adriatic Sea. More precisely, before the First World War there was a certain proportion of predator to prey fish sold in the Italian fish markets. During the war fishing decreased significantly, which led to a higher proportion of predator to prey (i.e. light fishing favoured the predators). After the War, where fishing increased to pre-War levels, the proportion of predator to prey returned to its pre-War level. Volterra wanted to know why the change in fishing levels, which, at least *prima facie* affected predator and prey

2 See Kendrew et al. (1958) for his original discussion of the model, and de Chadarevian (2004) for a useful philosophical discussion. The example is discussed in more detail in Chapter 10.

4 Available here https://commons.wikimedia.org/wiki/File:MONIAC.jpg#/media/File:MONIAC.jpg.

4 Available here http://www.sciencemuseum.org.uk/HoMImages/Components/799/79987_3.png.

at the same rate, led to the change in proportions in the fish markets.[5] He writes down the following coupled non-linear differential equations:

$$\frac{dV}{dt} = \alpha V - (\beta V)P \tag{1}$$

$$\frac{dP}{dt} = \gamma(\beta V)P - \delta P \tag{2}$$

He writes t = time; V = size of the prey population; P = size of the predator population; α = the intrinsic growth rate of the prey; δ = intrinsic death rate of the predators; and β and γ are prey capture rate and rate at which each predator converts captured prey into more predator births respectively (the 'phase space' of these equations is illustrated in figure 4 and explained in more detail in Chapter 10). He notices that the equations have two solutions where $\frac{dV}{dt} = \frac{dP}{dt} = 0$. First where $P = V = 0$, and second where $\alpha V = (\beta V)P$ and $\delta P = \gamma(\beta V)P$. The second is his primary interest. He calculates that although it is unstable, it corresponds to the mean values of P and V over an indefinitely large amounts of t. Letting $\hat{P}$ and $\hat{V}$ denote these values and $\rho = \frac{\hat{P}}{\hat{V}}$ he derives the following equation:

$$\rho = \frac{\alpha\gamma}{\delta} \tag{3}$$

He argues that heavy fishing corresponds to lower values of α and higher values of δ, which means that heavy fishing corresponds to lower values of ρ, and thus higher relative size of prey population with respect to predator population. He thus concludes that pre-First World War fishing activity led to higher prey to predator ratios, and thus more prey in the Adriatic fish markets, and the decrease in fishing activity associated with the First World War led to higher predator to prey ratios.

Wait. Volerra was writing down mathematical equations and then finding solutions to them. He was investigating how changing real valued parameters in mathematical functions affected the values of such functions. How could this tell him anything about fish populations in the Adriatic Sea? Again, this is because he constructed a model that represents the fish populations.

It's 1987 and Per Bak, Chao Tang, and Kurt Weisenfeld are interested in systems that exhibit what they call 'self-organised criticality' (SOC): systems that have the tendency to evolve to complex states without the need to fine-tune their initial conditions.[6]

5 See Volterra (1926, 1928) for his original discussion. See Weisberg (2007); Weisberg and Reisman (2008); Weisberg (2013) for useful philosophical discussions that I draw upon here. The example is discussed in more detail in Chapter 10. It's worth noting that equations 1 and 2 below were proposed independently by Alfred Lotka (1925).

6 See Bak et al. (1987) for their original discussion and Bak (1996) for an accessible introduction to SOC and the uses of models like the one discussed below. My discussion is greatly informed by Frigg (2003),

They consider the following scenario: take a flat surface and add sand (grain by grain) randomly across it in an attempt to build a a sandpile. As the pile grows, the addition of extra grains causes local motion. As it continues to grow, a state is reached where the pile is a critical size and adding more sand somewhere on the surface leads to sand falling off the edge elsewhere. In this state, the addition of a sand grain at a point can cause an avalanche that affects the entire pile and avalanches of all sizes can occur throughout the pile. The state is 'critical' in the sense that local events can affect the system in its entirety, and 'self-organised' in the sense that no deliberate effort is made to ensure the system arrives in such a state, and it will arrive in such a state regardless of the specific ways in which the grains are added.

To understand such a system, Bak, Tang and Weisenfeld reasoned as follows: consider a two dimensional lattice with a coordinate system such that tuples of the form (n_x, n_y) denote a cell in the lattice, where $n_x, n_y \in N$, where N is some finite subset of $\mathbb{N}$. There are N^2 of these cells. Each cell (n_x, n_y), is assigned an integer $Z(n_x, n_y)$ (at the 'edges', i.e. wherever $n_x = 0, n_y = 0, n_x = N$, or $n_y = N$, they stipulate that $Z(n_x, n_y) = 0$ is constant). This is illustrated in figure 3. At $t = 0$ let $Z(n_x, n_y) = 0$ for all (n_x, n_y). As t increases in discrete values, randomly chose a cell (n_x, n_y) and increase $Z(n_x, n_y)$ by $1 : Z_{t+1}(n_x, n_y) = Z_t(n_x, n_y) + 1$. Call this the 'addition of a sand grain' to the cell. Define a threshold K such that for all cells (n_x, n_y) and values of t, if $Z_t(n_x, n_y) \geq K$ the following occurs (unless the event occurs at the edge of the lattice, in which case the 'grain' is lost):

$$\begin{aligned} Z_{t+1}(n_x, n_y) &= Z_t(n_x, n_y) - 4 \\ Z_{t+1}(n_x \pm 1, n_y) &= Z_t(n_x \pm 1, n_y) + 1 \\ Z_{t+1}(n_x, n_y \pm 1) &= Z_t(n_x, n_y \pm 1) + 1 \end{aligned}$$

Call this a 'toppling event'. If such an event yields an additional toppling event on any of $(n_x \pm 1, n_y)$ or $(n_x, n_y \pm 1)$, then the process of choosing cells at random and increasing their Z value by 1 is paused until the toppling events have yielded a stable lattice (i.e. a lattice where $Z(n_x, n_y) < K$ for all (n_x, n_y)). The gap between a toppling event and reaching a stable state where 'grains' are added to the system again is called an 'avalanche'.

Let the system run for enough time, and it turns out that once the lattice is crowded enough (there are enough cells whose Z value is close to K), increasing the value of Z for some cell (n_x, n_y) can have dramatic effects throughout the lattice. Indeed

who provides a philosophical introduction to SOC and a careful consideration of how models based on the idea work.

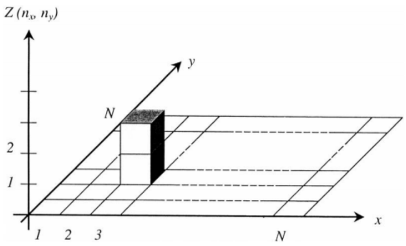

Figure 3: A two dimensional lattice representing two grains of sand at (3, 3) (Frigg, 2003, 616)

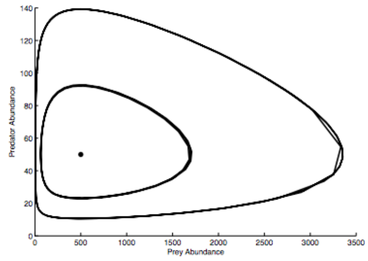

Figure 4: Phase space of the Lokta-Volterra model (Weisberg and Reisman, 2008, 133)

increasing the value of Z for one cell can yield every other cell in the lattice to be affected. For each avalanche, denote its size by s and its duration by d. It turns out that the number of avalanches whose size is s (denoted by $N(s)$), and the number of avalanches whose duration is d (denoted by $N(d)$) obey a power law:[7]

$$N(s) = s^{-a} \text{ and } N(d) = d^{-b} \text{ where } a \text{ and } b \text{ are real valued constants.}$$

Bak, Tang and Weisenfeld conclude that, when trying to build a sandpile, they will observe analogous behaviour. Once the actual sand reaches a certain height, the addition of one grain will cause the pile to topple, causing avalanches whose effects may be felt throughout the sandbox. And if they were to count the number of avalanches at a certain size (or of a certain duration), then they would observe that they obeyed the same power law.[8]

Wait. Bak, Tang and Weisenfeld were reasoning about two dimensional lattices, where integers were assigned to cells, and they defined a simple rule of distributing integers across adjacent cells when K was reached. How could that tell them anything about sandpiles? Sandboxes are not discrete lattices, and grains of sand are not integers assigned to cells. It should be of no surprise that, again, I think this is because they constructed a model that represents a sandpile.

In each of these examples scientists used a model to represent a system in the world, what I call a 'target system'. These are not isolated cases. In general, models play a central role in contemporary science. Scientists construct models of elementary particles, molecules, biological populations, individual decision makers, and economies. These models are used to build particle colliders and bridges, to synthesise medicines,

7 The fact that they obey a power law turns out to be a hallmark of SOC states.
8 See Bak (1996, Chapters 3,4) for discussions of actual experiments inspired by the model.

and set government budgets. Models are studied in order to gain knowledge and understanding about the parts, or features, of the world that they represent. If we are to understand how models are used to learn about the world, then we need to understand how they come to represent. This is the central question that I address in this thesis: in virtue of what do scientific models represent their target systems?

1.1 MOTIVATIONS

1.1.1 *Why models matter*

Philosophers of science didn't use to think that models were important units of science. The so-called 'received', or 'syntactic', view of scientific theories took theories to be linguistic entities, usually formalised in a predicate logic, which represent the world linguistically. Predicates refer to physical properties (or sets of entities in their extensions), particular entities are named, and sentences are constructed which can be either true or false depending on the nature of the world. The positivists who subscribed to this view under-emphasised the importance of models, and where they were discussed, assigned them a heuristic role at best. Carnap claimed that 'the discovery of a model has no more than an aesthetic or didactic or at best heuristic value, but it is not at all essential for a successful application of the physical theory' (1938, 210), and Hempel remarked that 'all reference to analogies or analogical models can be dispensed with in the systematic statement of scientific explanations' (1965, 440).

Things started changing in the 1960s. Starting with the work of Suppes (1969a,b, 2002), and followed by van Fraassen (1980, 1989, 2008) and Suppe (1989), along with Giere (1988), Lloyd (1994), Da Costa and French (1990, 2003), French and Ladyman (1997), and others, the so-called 'semantic' view of theories emerged as the dominant position in the philosophy of science. Rather than construing scientific theories as sets of sentences, adherents to the semantic view urge us to treat them as collections of models instead. And although the supporters of the semantic view held, and continue to hold, various different accounts of the precise nature of the models (many of which are detailed in Part ii of this thesis), what they have in common is a rejection of the idea that scientific representation is a linguistic matter. At their heart, scientific theories are taken to be collections of objects – mathematical structures; state or phase spaces; abstract entities; and so on – that represent the world, rather than names, predicates, words, and sentences, in formal (or informal) languages.

Independently of the semantic view, philosophers have emphasised the role that models play in scientific practice. Originating from the likes of Black (1962), Hesse

(1963), and Achinstein (1968), through Cartwright (1983, 1999), Cartwright et al. (1995), and delivering the 'models as mediators' view subscribed to, in various forms, by the contributors to Morgan and Morrison (1999), philosophers of science who find 'rational reconstructions' distasteful have pointed out that the practice of science is full of scientists using models to represent specific parts of the world. Linguistic theories might provide general laws, or principles, that govern the natural world, but in order to represent anything specific, be it the fall of a banknote from the top of St Stephen's Cathedral or the behaviour of quarks inside a Baryon, scientists construct models. If philosophers of science are to take the practice of science seriously, then they best account for how models represent. On the models as mediators view, theories are not to be identified with collections of these models, nor eschewed altogether in favour of them. Rather, models are taken to 'mediate' between abstract theories and target systems.

The models as mediators view is the most prominent alternative to the semantic view of theories. But despite their differences, both accounts share an commitment to models being, if not the only, then at least a central, way in which science represents the world. And again, despite the differences in what models are taken to be, they are primarily seen as non-linguistic. This raises a question. Whereas the positivists could appeal to work in the philosophy of language (and logic) to provide an account of how sentences represent – in terms of denotation, reference, satisfaction, truth, and so on – these do not straightforwardly carry over to explain the representational capacity of models. This is the primary motivation for this thesis. If science represents via models, then how do models represent? An answer to this question has to be given if either of the views outlined above are to be at all viable.

1.1.2 *Realism vs. anti-realism*

I think that philosophers in general should be interested in how models represent. Philosophers of language have spent decades arguing about how words represent; philosophers of mind focused on mental states for at least as long; and aestheticians have investigated the nature of representation in art for thousands of years. Scientific models are just another kind of thing that we use to represent the external world. They, along with related representations like maps, diagrams, and graphs, deserve the same attention philosophers have given to names, sentences, mental states, and works of art.

For those not convinced, there are also reasons internal to the philosophy of science that motivate why the question of scientific representation needs answering. The standard definition of scientific realism is split into three separate theses (Psillos, 1999; Chakravartty, 2015):

METAPHYSICAL: The world exists, organised in a certain way, mind independently.
SEMANTIC: Our scientific theories are to be interpreted at face value as truth conditioned descriptions of the world.
EPISTEMIC: Mature and predicatively successful theories are approximately true.

The metaphysical thesis makes no reference to what science provides as representations of the world. According to the semantic thesis, scientific theories are 'truth-conditioned descriptions', and if the epistemic thesis is correct, these descriptions are true. Such a definition of scientific realism inherits from the syntactic view the idea that scientific theories are linguistic entities, since it is linguistic entities that have truth conditions, and are the sorts of things that can be true. Of course what it means for a sentence to be true is a question that deserves addressing.[9] But by and large, philosophers of science can push this question off to philosophers of language, or engage in philosophy of language themselves if they prefer to provide an account of scientific truth that diverges from the everyday notion of truth applied to non-scientific statements.

But models are not descriptions, at least in any straightforward manner. Nor are they obviously the sorts of things that can be true or false. So as it stands, the semantic and epistemological theses of scientific realism are simply inapplicable to models. This leaves us with three options. Option one is to define realism solely in terms of scientific theories, and construe theories linguistically, ignoring what it would mean to adopt a realist or anti-realist stance towards models. Option two is to provide an account of scientific models according to which they are truth-apt descriptions when interpreted at face value. Option three is to reinterpret scientific realism in such a way as to make it applicable to models, without taking them as truth-apt.

Given the discussions in the previous section, I don't think the first approach is satisfactory. Even if theories should be construed linguistically, models remain important representational units of science, and so a definition of realism that remains quiet about them is not a fully general account. Both of the latter two approaches are viable options. The account of scientific representation I provide in Part iii does not take models themselves to be truth-apt, and as such I don't think that the second option is the right way to think about scientific realism. Indeed the majority of philosophers working on modelling accept that, strictly speaking, models are not the sorts of things

9 Should we subscribe to a correspondence theory of truth for scientific statements? A deflationary theory? Something else entirely? Related questions are discussed in Chapter 8.

that can be true or false, even though they might 'tell' us things about the world.[10] So my preference is for the third option: to provide a general account of realism that does not make reference to the notion of truth, or of a description.

What would such an account look like? The following is a naïve first attempt:

METAPHYSICAL: The world exists, organised in a certain way, mind independently.
SEMANTIC: Our scientific models are to be interpreted at face value as more or less accurate representations of the world.
EPISTEMIC: Mature and predicatively successful models are approximately accurate.

With the exception of some remarks about the 'structure' of the world in Chapter 6, I have little to say about the metaphysical thesis. Similarly, although I discuss the question of when models are accurate representations at various points throughout this thesis, as I argue in Section 1.3, before investigating what it means for a model to be an accurate representation of its target system, we have to understand how they represent their target systems in the first place. So in order to understand the semantic and epistemic thesis of realism in the context of model-based science, we need to get clear about in virtue of what scientific models represent their target systems. This is a primary motivation of this thesis. For want of space I cannot develop a definition of realism as applied to scientific models here, but my hope is that the account I provide will clear the ground for further research in this area. What is important to remember is that, without an account of scientific representation, we do not have a clear grasp of what the scientific realism/anti-realism distinction amounts to in the context of model-based science. And as such, I would expect philosophers of science to see the motivation for addressing the question.

1.2 OUTLINE AND METHODOLOGY

1.2.1 *Outline*

The thesis is split into three parts. Part i is composed of two chapters. In Chapter 1 I situate the question of scientific representation in the broader philosophical landscape. I also disentangle a number of issues that are twisted together in the literature on scientific representation. The result is a number of questions and key desiderata that any successful account should answer and meet. These provide a novel lens through which to investigate scientific representation, broadly construed, and serve to provide

10 An exception is Mäki (2011) who argues that models can be truth-valued. I discuss his views in further detail in Chapter 4.

the conceptual framework on which this thesis rests. In Chapter 2 I argue that, contra Callender and Cohen (2006), there is a special problem of scientific representation, and that it cannot be easily answered by a blunt appeal to the intentions of model users.

Part ii of this thesis is a detailed investigation into the accounts of scientific representation currently available in the literature. In Chapter 4 I address the time-honoured view that models represent in virtue of being similar to their targets (a view associated with Giere (1988, 2004, 2010), Mäki (2009), and Weisberg (2013) amongst others). In Chapter 5 I consider a popular version of this view that specifies the similarity in terms of structural properties of models and their targets (a view associated with Da Costa and French (1990), French (2003), French and Ladyman (1997, 1999), Suppe (1989), Suppes (1969a), and van Fraassen (1980, 2008) amongst others). In Chapter 6 I address the metaphysical question concerning 'the structure' of the world. Both the similarity and structuralist views take seriously the idea that scientific models are 'model-systems' in their own right, in the sense that they mediate our understanding of our target systems. In Chapter 7 I investigate an opposing view that takes the descriptions and equations that are used to present scientific models to directly represent their target systems instead (Levy, 2015; Toon, 2010a, 2012). Finally, in Chapter 8 I address the family of accounts that are closest to my own (Contessa, 2007; Frigg, 2010a; Hughes, 1997; Suárez, 2004, 2015). Such accounts emphasise the inferential role that models play in generating predictions and hypotheses about their target systems. But in order to understand how this role relates to their representational status, it is necessary to address much broader philosophical questions concerning the relationship between representation and inference in general.

Part iii is devoted to a positive proposal concerning the nature of scientific representation. In Chapter 9 I develop Goodman's (1976) and Elgin's (1983; 1996; 2009; 2010) account of pictorial representation in the scientific framework. The result is what I call the DEKI account of scientific representation. Models are *interpreted* in the appropriate manner. They *denote* their target systems and *exemplify* certain features which may then be translated by a model *key* into features to be *imputed* onto their target systems. The result is a highly conventional account of scientific representation which operates at the appropriate level of generality to capture all of its instances, whilst remaining informative about the notion. In Chapter 10 I further illustrate how the account works and analyse it through the conceptual lens provided earlier in the thesis. Finally, in Chapter 11 I offer some concluding remarks.

1.2.2 *Methodology*

To set the tone for the rest of this thesis, it's useful to outline the philosophical methodology I use to approach the question of scientific representation. The way I see it, scientific representation is a sub-species of representation in general. Just as sentences, propositions, names, mental states, works of art, photographs, and maps represent states of affairs, name-bearers, subjects, or terrains, scientific models represent their target systems (although perhaps they do so in a somewhat different manner). And if philosophers of science are to get to grips with scientific representation, then they should draw on the rich tradition of investigating the nature of representation in other domains.

Such a view is, perhaps, not very fashionable these days. Compared to the 20th Century, contemporary philosophy of science has distanced itself from other areas of philosophy, the philosophy of language in particular. This is, in some sense, a good thing. It has allowed philosophers of science to get to grips with the complex philosophical issues facing practising scientists without worrying too much about what some might call 'philosophical navel gazing'. And it has instilled a healthy respect in its practitioners for an understanding of science that goes beyond using toy examples to illustrate seemingly substantial claims about the nature of the natural world.

But at the same time there is the danger of getting so focused on the messy details of science that the philosophical questions are lost. This is especially pertinent here, given that the questions I am addressing in this thesis are stated at such a level of generality that they cross scientific fields. As such, the examples I use to illustrate various positions are drawn from diverse areas of science, and are simple enough to be accessible to an educated layperson. This is not work on the philosophy of any particular science, but rather one in the general philosophy of science.

Moreover, it is a firmly philosophical work rather than a case-study driven investigation of scientific practice. And as such I draw heavily throughout on works in aesthetics investigating the nature of pictorial representation and, to a lesser extent, work in the philosophy of language investigating the nature of denotation (or reference) and truth. I come at the question of scientific representation in the same way that others have come at the question of 'how does a picture represent its subject?' or 'how does a name denote its bearer?'. My hope is that the result is a piece of work that addresses a fundamental question about how we represent the world in a philosophically sophisticated manner, and at the same time pays due respect to practising science.

In addition to this broad philosophical approach, throughout this thesis I take a specific attitude on how to address questions surrounding scientific models. Many discussions of models focus primarily on their ontology (see, for example Frigg (2010b) and the views I discuss in Chapter 7): what is a harmonic oscillator? What is an infinitely large population of interbreeding rabbits? In general, what are these models that, as Hacking (1983, 216) elegantly puts it, we hold 'in our heads rather than our hands'? And then only once an answer to this question has been provided does the question of scientific representation get addressed.

My approach is the opposite. My primary interest is in the semantic, or epistemological questions, that arise for scientific models. I want to figure out in virtue of what they represent, and how they can be used to learn about the world. This is not to say that the ontological question is ignored entirely. Indeed as I argue in the next section, various accounts of scientific representation constrain, at least to some extent, what sorts of entities models must be. For example, if one thinks that models represent their targets in virtue of being isomorphic to them, then it better be the case that scientific models are the sorts of things that can enter into isomorphisms in the first place. However, it is not the case that an account of scientific representation needs to fully determine the ontology of models. As I suggest in Chapter 10, the account that I provide in this thesis is compatible with multiple different ontological accounts of scientific models. Although I think the ontological questions are interesting in their own right, they are not my primary focus here.

Finally, I do not wish to claim that all models are representations – in fact the account of scientific representation I offer in Part iii is explicitly designed to account for models without target systems – nor that representation is the only function of models. Various authors working on the topic of scientific modelling have emphasised that models play a variety of different roles. Knuuttila (2005) points out that the epistemic value of models is not limited to their representational function and develops an account that views models as epistemic artifacts which allow us to gather knowledge in diverse ways; Hartmann (1995) discusses the role that models play in the development of theories; Peschard (2011) investigates the way in which models may be used to construct other models and generate new target systems; and Bokulich (2009) and Kennedy (2012) present non-representational accounts of model explanation (although see Reiss (2012) and Woody (2004) for more general discussions about the relationship between representation and explanation). My premise is the more modest one: at least some models represent their target systems, and these representational relationships play some role in how we learn about the world. I do not think that either of these claims are controversial, nor that they conflict with the aforementioned discussions.

Moreover, I am not addressing everything that could be called 'scientific representation'. Scientists represent the world with models, but they might also do so with graphs and diagrams (see Perini (2005a,b, 2010) for discussions of this sort of representation); with tree rings and disease symptoms, or more generally things like thermometer readings and litmus paper colours which are usually described as measurements rather than models (see Tal (2015) for an introduction to measurement); and indeed, if one does not want to identify theories with models, then with scientific theories as well (see Winther (2015) for an introduction to the structure of scientific theories). With the exception of the discussion of scientific data in Chapter 6, I have little to say about these other forms of representation. And thus my use of the term 'scientific representation' is something of a misnomer. 'Model representation' would be more accurate, but I sacrifice accuracy for style and follow the literature in using the former phrase instead. As long as the reader keeps in mind that my focus is on how models represent, I do not think any confusion will arise.

1.3 CONCEPTUAL FRAMEWORK

Models represent their target systems, a part, or feature, of the world. My primary interest in this thesis is in what makes this the case. In virtue of what do scientific models represent their target systems? This is what Frigg calls 'the enigma of representation' (2006, 50) and Suárez calls the 'constitutional question' (2003, 230; 2010, 93; 2015, 38) concerning scientific representation. To get to grips with this problem it's useful to consider the same problem in the context of pictorial representation. When seeing Obama's 2008 presidential campaign poster, designed by Shepard Fairey, we immediately recognise that it depicts a man in a suit and tie. Why is this? *Per se* the poster is just ink distributed across a rectangular sheet of paper, or, if viewed on a computer screen, a collection of pixels. How does this come to represent something else, something external to itself?

Similarly, consider the Phillips-Newlyn machine, a collection of pipes, reservoirs, and valves through which water is pumped. Guatemalan economists were able to manipulate the machine in order to generate a prediction about what would happen to the economy if there was a decrease in foreign investment into the country. But how could a change in water levels in one of the machine's reservoirs represent what was going on in the Guatemalan economy? How did the machine come to represent the economy, something external to itself?

Likewise, before being representations of particles, molecules, chemical reactions, and populations, models are equations, mathematical structures, fictional scenarios, or material objects in their own right. The problem then, is what turns these entities into

11 Available here `https://en.wikipedia.org/wiki/File:Barack_Obama_Hope_poster.jpg#/media/File:Barack_Obama_Hope_poster.jpg`.

Figure 5: Obama's 2008 presidential campaign poster (Fairey)[11]

representations of something beyond themselves? It's customary to phrase questions like this in terms of necessary and sufficient conditions, and I adopt this approach in this thesis. However, such a philosophical approach to analysing concepts has had some bad press in recent years (see Laurence and Margolis (1999) for an overview), so it's worth briefly defending it here. I phrase the question of scientific representation in terms of necessary and sufficient conditions for three reasons.

Firstly, with the notable exceptions of van Fraassen (2008), Hughes (1997) and Suárez (2004, 2015), whose views are discussed with reference to this approach in Chapters 6 and 8, the majority of philosophers working on the topic have adopted this practice. Thus, for ease of presentation of their views in Part ii, adopting the same framework proves useful.

Secondly, the standard arguments against such an approach is that concepts are, in some sense 'fuzzy'. Should we classify the offspring of a tiger and a lion as falling under the extension of our concept of tigers, or lions, or both? It's not clear. But if these concepts can be strictly delineated in terms of necessary and sufficient conditions, then there is a fact of the matter about this, and this seems to conflict with how we think about the world. However, at least when it comes to scientific representation, I am not too interested in the boundary cases. Even with respect to clear instances of things that fall under the concept of scientific representation – a Newtonian model of the celestial orbits, the Phillips-Newlyn machine, ball-and-stick models of molecules – we do not have a clear answer to in virtue of what they represent their target systems. These central instances are the focus of my analysis here. Once we understand how they represent, then we can start worrying about whether the necessary and sufficient condition approach is too strict.

Thirdly, and relatedly, I think framing the question in this way is pragmatically justified. Even if it turns out that analysing the concept of scientific representation like this is misguided, this does not imply that it has no value whatsoever. Perhaps the concept of knowledge does not admit necessary and sufficient conditions. But this is not to say that attempts to provide them have told us nothing about it. We have learnt much about the nature of belief justification from exploring Gettier cases and the like. Moreover, the approach has the significant benefit of adding precision to the way we think about the concept, at least in the first instance.

With this in mind, the central question concerning scientific representation is the following: what fills in the blank on the right-hand-side of the following biconditional: a model M represents a target system T if and only if ________ (*cf.* Frigg, 2003, 14)? For reasons that will become clear soon, I call this the *Epistemic Representation Question*, or *ER-Question* for short. By and large I remain silent about the scope of T. I understand the phrase 'target system' liberally to include parts of the world that models represent, including, but not limited to, particular objects, features, events, mechanisms, processes, states, and states of affairs. However, in answering the ER-question I also provide an account of what turns an object (mathematical structure, fictional scenario, material object and so on) into a model.

What conditions of adequacy must any answer to this question meet? I think the most important one is the following. Models allow us to (attempt to) learn about their target systems; they are, in some sense, informative about them. This is not to say that everything we learn from a model must be true, indeed as discussed below, models frequently provide us with false hypotheses about their targets, but they at least allow for the possibility of learning. They provide us with hypotheses, which may be right or wrong, about what their target systems are like, what features their target systems do, or do not possess.[12]

Swoyer (1991) introduced the phrase 'surrogative reasoning' to describe this aspect of using scientific models, and – with the notable exception of Callender and Cohen (2006) whose views are discussed in the next section – there is widespread agreement on this point. Bailer-Jones emphasises that models '*tell us* something about certain features of the world' (2003, 59, original emphasis); Bolinska (2013) and Contessa (2007) both call models 'epistemic representations'; (Frigg, 2006, 51) sees the potential for learning as an essential explanandum for any theory of representation; Liu (2013) emphasises that the main role for models in science and technology is epistemic; Morgan

12 By 'hypotheses', I mean any claim about the target system, not just observable predictions.

and Morrison take models to be 'investigative tools' (1999, 11); Suárez (2003, 2004, 2015) submits that models licence specific inferences about their targets; and Weisberg (2013, 150) observes 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'. As such, since this aspect of modelling is so tightly linked to their representational ability, I take it that any acceptable answer to the ER-question must account for the possibility of using models to learn about their targets. I call this the *Surrogative Reasoning Condition*.

This immediately poses an additional problem. Although models can be used to learn about their targets, they are far from unique in this respect. A picture provides us with information about its subject: Barack Obama wears a suit, for example. Similarly a map allows us to generate hypotheses about its target terrain: the distance between Brixton and King's Cross is less than 6 miles, the A1 connects Newcastle and London, and so on. In general many kinds of representation provide us epistemic access to their targets or subjects, despite not being reasonably classed as 'models', and hence it would not be surprising if conditions which satisfied the learning condition apply to these sorts of representations as well. This is a problem for an analysis of scientific representation in terms of necessary and sufficient conditions because if something that is not *prima facie* a model (for instance a map or a photograph) satisfies the conditions, then one either has to conclude that the account fails because it does not provide sufficient conditions, or that first impressions are wrong and other representations (such as maps or photographs) are in fact scientific representations.

Neither of these options are particularly appealing, and for this reason I broaden my scope from the representation relationship that holds between models (proper) and their targets, to what Contessa (2007) calls 'epistemic representation'. The ER-problem then, is what are the necessary and sufficient conditions on things like models, maps, photographs, and pictures, representing their targets and subjects. In my thesis I primarily restrict my focus to scientific models, but it is an interesting further question to what extent the accounts discussed and proposed apply to these other kinds of epistemic representations.

My view, which I don't argue for here, is that by and large they do. But this invites a further question: can scientific models be demarcated from other epistemic representations, and if so, how? For obvious reasons, I call this the *Demarcation Problem*. Callender and Cohen (2006, 69, 83) raise this problem and then immediately raise sceptical doubts about our ability to answer it. By and large, in the literature on scientific representation, authors commonly provide examples, such as maps (Contessa, 2007; Frigg, 2010b) and pictures (Elgin, 2010; French, 2003; Frigg, 2006; Suárez, 2004; van Fraassen, 2008) to illustrate their approach. This implicitly suggests that they either haven't considered the question, or that they share Callender and Cohen's scepticism that it can be answered. But this is by no means a neutral position, and as such, it deserves to be made explicit whether such accounts entail that scientific models can be demarcated from other epistemic representations. I do this throughout Part ii. And, following the aforementioned authors, I draw on similar examples of non-scientific epistemic representations to illustrate my own preferred account in Part iii.

Even if there is no sharp line between scientific and non-scientific epistemic representations, it is clear that even for paradigmatic examples of scientific models, not all of them represent their target systems in the same way. In the case of visual representations this is clear: an Egyptian mural, a two-point perspective ink drawing, a stylised stencil portrait, an architectural plan, and a tube map represent their respective targets in different ways. This pluralism is not limited to visual representations. Not all model representations seem to be of the same kind either. Woody (2004) argues that chemistry as a discipline has its own ways to represent molecules, and this is very different to the way that physicists represent them (using Schrödinger's equation or approximations thereof). Differences in style can also appear in models from the same discipline. As Frigg (2006, 50) points out, Weizsäcker's liquid drop model represents the nucleus of an atom in a different manner to the Bohr model. A scale model of a ship in a tank seems to represents an actual ship in a way that is different from how a mathematical fluid dynamical model does. Or the Phillips-Newlyn machine and Hicks' IS-LM equations both represent economies, but they do so in very different ways. Following Frigg (2006, 50) I call this the *Problem of Style*.[13] I take it that categorising scientific models by style, like categorising pictures, is rather fluid. Different styles may come into existence as scientific fields grow and develop. As such, I doubt that an exhaustive list of clearly defined styles can be given. However, this is not to say that nothing can be said about how models that clearly represent their respective targets in diverse ways differ with respect to their style. The aforementioned discussion demands some important caveats about how I have phrased the ER-problem. I want to be clear that the phrasing in terms of necessary and sufficient conditions does not demand that epistemic representation is, in any sense, a homogeneous concept. The conditions given can be disjunctive, which would allow for multiple styles to have their own particular conditions. Or stated at a level of generality that models in different styles can satisfy them in different ways (by satisfying different, more concrete, conditions for example). Or, if one were to take a strong stance on the demarcation condition, the ER-problem would be answered by providing different sets of conditions for scientific and non-scientific representation, without discussing what they have in common (essentially

13 Suárez (2003, 231-232) uses this observation as an argument – which he calls the 'argument from variety' – against various answers to the ER-problem.

separating the ER-problem into two distinct problems). I would count all of these as satisfactory answers to the problem, and as such asking it in this way does not build in any substantial requirements.

The next important problem associated with the topic of scientific representation concerns in virtue of what they are *accurate* representations of their target systems. Even if it is established that a model represents its target, there is an additional question regarding whether it does so accurately.[14] I call this the problem of specifying the *Standards of Accuracy*. In principle, answering the ER-problem might not entail anything about how these are to be specified. However, many of the accounts of scientific representation discussed in Part ii are highly ambiguous between answering the ER-problem, and providing standards of accuracy on preexisting representational relationships between models and their targets. As is shown there, it is important to untangle these two questions, if we are to provide a clear account of scientific representation.

The problem goes hand in hand with another desideratum on a successful answer to the ER-problem. Any specification of in virtue of what scientific models represent their target systems should allow for the *Possibility of Misrepresentation*. One motivation for this condition is clear: look at the history of science for examples of models that *mis*represent their target systems. Thomson's plum pudding model of the atom represents atoms, but does so inaccurately: it represents the positive charge of an atom as uniformly 'smeared out', rather than concentrated in the atomic nucleus. But in order for the model to inaccurately represent the atom, it has to represent it in the first place. In general, for a model to be a misrepresentation – i.e. to represent its target inaccurately – it has to represent it in the first place.

A second motivation for this condition is more conceptual. In discussions of representation in other fields of philosophy, it is often made explicit that any account of, say, linguistic, mental, or even genetic representation has to allow for the possibility of misrepresentation. Fodor's (1990) 'disjunction problem' charged at informational or teleosemantic accounts of mental content explicitly relies on the fact that any account of belief has to allow for the possibility of *false* beliefs. Plausibly, the same requirement is used by the later Wittgenstein when he argues against the possibly of a private language (Wittgenstein 1953; *cf*. Kenny, 1973/2006, 141-159). In order for a term 'S' to name a sensation, there must be some standards under which it is appropriate for the speaker to use the term 'S'. Since these standards don't exist in the case of a private lan-

14 To avoid confusion further down the line, it's worth noting that an answer to this question may make reference to the purposes of the model user in the modelling context, and that, at least as I use the notion, accuracy of scientific representation, unlike representation itself, might be the sort of thing that admits a degree.

guage – in order to ascertain whether they are met they are compared to the memory of the sensation, i.e. the memory of the meaning of 'S' which is 'as if someone were to buy several copies of the morning paper to assure himself that what it said was true' (Wittgenstein, 1953, 265) – Wittgenstein concludes that such a language cannot exist. But in invoking these standards the possibility of a term being applied inappropriately is clearly relied upon. Again, in order to represent the sensation, it must be the case that 'S' can *mis*represent it.

For reasons related to these arguments it is now common to explicitly require that any account of representation allow for the possibility of misrepresentation. For example, Stich and Warfield (1994, 6-7) demand that any account of mental representation allow for this, and in their discussion of the 'genetic code' Sterelny and Griffiths (1999, 104) require that any account of the representational capacity of genes allow for the possibility that they misrepresent the phenotype that they eventually generate. Since I see no reason why these considerations do not apply in the case of scientific representation as well, I take it that any answer to the ER-problem must allow for the possibility of a model misrepresenting its target.[15]

A further issue tied up with the question of scientific representation stems from the highly mathematical nature of many (if not most) scientific models. Even if they are not purely mathematical entities – whether that means they should not be identified with syntactic equations or mathematical structures – their mathematical aspects seem to directly contribute to their representational capacity. As such, when investigating how they represent their target systems, we run up against what Wigner called the 'unreasonable effectiveness of mathematics in the natural sciences'; namely the questions of how mathematics applies to the physical world, and why it is so successful. These questions, in the context of this thesis, amount to the questions of how mathematical models represent their target systems, and how they do so accurately.

Surprisingly, with the exception of Bueno and Colyvan (2011) and Pincock (2012), there has been little attempt to bridge the gap between the literature on the applicability of mathematics and discussions of scientific representation. This should be rectified, and in this thesis I hope to make at least a modest attempt to do so. In Part ii I explain how the existing accounts of the latter explain the applicability of mathematics, and in Part iii I offer some suggestions about how my account could be used to help understand how mathematics applies to the world. Of course, I do not expect that an account to the ER-problem will fully solve Wigner's puzzle. Indeed as I outline in Part iii, it turns out that there are various ways of accommodating the

15 This approach of sharply distinguishing between the question of in virtue of what models represent their targets, and in virtue of what they do so accurately, is also found in Suárez (2004); Contessa (2007); and Frigg (2006, 2010a). For a dissenting voice see Bolinska (2015).

mathematical nature of scientific models in my account of scientific representation, and narrowing down which of these ways serves to provide the 'correct' account of how mathematics applies to the real world is not something I can do in this thesis. Nevertheless, I do expect that any successful account of scientific representation will, at the very least, account for mathematical models in such a way as to help explain how their mathematical aspects contribute to their representational capacities. I call this the *Applicability of Mathematics* condition.

Finally, and relatedly, in answering all of the above questions, it is clear that any account of scientific representation will run up against the question of what models *are*, ontologically speaking. Are they equations, structures (and if so, set theoretical, group theoretical, category theoretical, or what?), fictional scenarios, abstract objects, descriptions, or something else entirely? I call this the *Problem of Ontology*. While some authors develop an ontology of models, the accounts discussed, and ultimately rejected, in Chapter 7 abandon an understanding of models as 'things' and push a programme that can be summed up in the slogan 'modelling without models' (Levy, 2015). As noted in the previous subsection, many philosophers of science working on the question begin by investigating their ontology before moving onto the problems discussed in this section. My approach is the opposite. But this is not to say that I remain silent on the ontological questions entirely. In Part ii I point out that many of the existing answers to the ER-problem place heavy constraints on the ontology of (specifically non-physical) scientific models. They don't, however, uniquely determine this ontology. This is also the case for my own positive proposal, and in Part iii I explicate how it is compatible with what I take to be the most plausible answers to the ontological problem.

To sum up, I have argued that the following questions surrounding the question of scientific representation should be carefully distinguished:

ER-PROBLEM: Provide necessary and sufficient conditions of a model representing its target.

DEMARCATION PROBLEM: Demarcate scientific models from other epistemic representations (or motivate why they should not be so demarcated).

PROBLEM OF STYLE: Account for the fact that different scientific models seem to represent their targets in different ways.

STANDARDS OF ACCURACY: Provide standards of accurate epistemic representation.

PROBLEM OF ONTOLOGY: Help us understand what (specifically non-physical) models are, ontologically speaking.

Moreover, any successful account of scientific representation (in particular, an answer to the ER-problem) should meet the following conditions:

SURROGATIVE REASONING CONDITION: Account for the fact that models, and epistemic representations more generally, can be used to attempt to learn about their targets.

POSSIBILITY OF MISREPRESENTATION: Allow for the fact that a model may misrepresent its target system, i.e. represent it inaccurately.

APPLICABILITY OF MATHEMATICS: Help us understand how mathematical models represent their target systems.

I do not claim that these problems and conditions should be addressed in order, nor that they are independent of one another. Rather, I think that by distinguishing between them, we arrive at a standpoint from which 'the question' of scientific representation can be considered more carefully. Together, these problems and conditions provide my theoretical framework with which to critically address the existing accounts of scientific representation (Part ii), and moreover, I take as a significant attribute of my own account that it clearly distinguishes between each of these problems, and meets each of the conditions (Part iii).

2

IS THERE A SPECIAL PROBLEM OF SCIENTIFIC REPRESENTATION?

Before addressing the more substantial accounts of scientific representation, it's worth first discussing Callender and Cohen's claim that 'there is no special problem about scientific representation' (2006, 67). After all, if this is the case there would be little point in this thesis, or at the very least, it would not be a work in the philosophy of science. Their argument comes in two parts. Firstly, they claim that scientific representation is not different from 'artistic, linguistic, and culinary representation' (2006, 67), and that the representational capacities of all of these types of representation, including scientific models, is 'derivative from the representational status of a privileged core of representations' (2006, 70). They call this position 'General Griceanism'. The second part of their argument concerns how to account for the capacities of the derivative representations in terms of the core representations. The answer that they suggest is that it suffices to stipulate that a derivative representation represents its target. I address each of these claims in order.

2.1 GENERAL GRICEANISM

Callender and Cohen characterise General Griceanism as follows:

> 'The General Gricean holds that, among the many sorts of representational entities (cars, cakes, equations, etc.), the representational status of most of them is derivative from the representational status of a privileged core of representations. The advertised benefit of this General Gricean approach to representation is that we won't need separate theories to account for artistic, linguistic, representation [sic], and culinary representation; instead, the General Gricean proposes that all these types of representation can be explained (in a unified way) as deriving from some more fundamental sorts of representations, which are typically taken to be mental states. (Of course,

> this view requires an independently constituted theory of representation for the fundamental entities.)' (2006, 70).

And since they group scientific models in the class of non-fundamental representations, this amounts to explaining scientific representation in two stages. Firstly, to explain how scientific representation can be reduced (in some sense) to the more fundamental type of representation, and secondly, to provide an account of how the more fundamental type works. According to them:

> 'of these stages [...] most of the philosophical action lies at the second. The first stage amounts to a *relatively trivial trade* of one problem for another: you thought you had a problem of representation for linguistic tokens (or whatever you take to be derivative representations)? Exchange it for a problem of representation for mental states (of whatever you take to be fundamental representations). This trade, in effect, just pushes back the problem of representation by a single step. The second stage, in contrast, amounts to a *fairly deep metaphysical mystery*.' (2006, 73, emphasis added).

The outcome of this, at least from the perspective of philosophers of science interested in the ER-problem, is that they can push any substantial answer to the question off onto philosophers of mind (or philosophers, or scientists, working on whatever the fundamental type of representation is taken to be).

I think there is an element of truth in General Griceanism. But at least as Callender and Cohen state it, it does not suffice to either answer the ER-problem, or to dismiss it as a significant problem. Let's start with what we agree on. For the sake of argument, I grant that, fundamentally, there is a privileged class of representations – mental states perhaps – and that things like scientific models and epistemic representations more generally gain their representational status at least in part in virtue of their relationship with elements in this class. As I discuss in Part ii many of the more developed answers to the ER-problem invoke the intentions of model users as playing a vital role in establishing that models represent their targets.

But, even if one accepts that in order to explain how models represent their targets we must invoke more fundamental representational entities, it does not follow, as Callender and Cohen assert, that this explanation will be the same as the explanation of how artistic, linguistic, or culinary representation derives from the privileged core of representation. In fact, if one thought that the demarcation problem could be answered in the positive – that scientific epistemic representation is fundamentally different from non-scientific epistemic representation – then they may be tempted to account for this difference precisely in what establishes their representational capacities.

With respect to the demarcation problem, Callender and Cohen state:

> 'We are not optimistic about solving [the demarcation] problem. And we think it a virtue of our account that it allows one to see clearly that the demarcation problem for representation just is an instance of the general demarcation problem concerning the difference between science and non-science' (2006, 83).

Although this might be a virtue in one sense, it is a vice in the sense that it rules out the approach outlined above. I'm not saying that this approach is the correct one; rather, it bears noting that Callender and Cohen's point is not entailed by a more modestly stated version of General Griceanism. Just because the representational capacity of non-fundamental representations depends on a privileged class of representations, nothing is entailed about the nature of this reliance, nor that it is the same across different types of non-fundamental representation.

Most importantly, the modest version of General Griceanism I accept certainly does not entail that explaining the representational capacities of models in terms of more fundamental representations amounts to a 'relatively trivial trade of one problem for another'. That there has to be *some* relationship between fundamental and non-fundamental representations does not mean that this relationship can be trivially identified. The only reason that Callender and Cohen supply for why this might be the case is that their preferred explanation of the relationship is a simple one: in order for a non-fundamental representation to represent its subject or target, it suffices that someone simply stipulate that this is the case. I call this position 'Stipulative Fiat'.

2.2 STIPULATIVE FIAT

Callender and Cohen claim that:

> 'Can the salt shaker on the dinner table represent Madagascar? Of course it can, so long as you stipulate that the former represents the latter. [...] Can your left hand represent the Platonic form of beauty? Of course, so long as you stipulate that the former represents the latter. [...] On the story we are telling, then, virtually anything can be stipulated to be a representational vehicle for the representation of virtually anything [...]; the representational powers of mental states are so wide-ranging that they can bring about other representational relations between arbitrary relata by dint of mere stipulation. The upshot is that, once one has paid the admittedly hefty one-time

> fee of supplying a metaphysics of representation for mental states, further instances of representation become extremely cheap' (2006, 73-74).

The above quotation contains two important assertions. Firstly, (virtually) anything can represent (virtually) anything, and secondly, that whenever something does represent something else, at least in the case of non-fundamental representations, this is established in virtue of someone stipulating that this is the case. With respect to the first claim, I agree. Things that function as models don't do so in virtue of their ontological status, and it would be a mistake to think that that some objects are, intrinsically, representations and other are not. This point has been made by others too (including Frigg (2010a, 99), Giere (2010, 269), Suárez (2004, 773), Swoyer (1991, 452), and Teller (2001, 397)). With respect to the second, this is a significant claim about how scientific models come to represent their targets. This is their answer to the ER-problem:

STIPULATIVE FIAT: A model M represents a target T if and only if a model user stipulates that M represents T.

If correct, then Callender and Cohen would be justified in their claim that reducing scientific representation to a more fundamental type of representation amounts to the relatively trivial trade of one problem for another. But there are at least two significant problems with this account. Firstly, it is unclear whether an act of stipulation in general suffices to establish a representational relationship. Secondly, even if it did, it's certainly not the case that the sort of relationship thereby created is the sort of relationship that holds between epistemic representations and their targets or subjects. I address both of these concerns in order.

In order to investigate the relationship between stipulation and representation in general, it is useful to turn to the philosophy of language. The pertinent question there, is whether stipulation is sufficient to establish that a word or phrase denotes whatever it denotes, or means whatever it means. In some cases, this seems to be the case. In order to introduce a name into a language, it can suffice to use it to 'baptise' its bearer (Kripke, 1980), or to provide a stipulative definition for what it means. But the position that stipulation, or the bare intentions of a user, suffice to establish denotation or meaning in general is a much more controversial claim. It faces the so-called 'Humpty Dumpty' problem, named in reference to Lewis Carroll's discussion of Humpty using the word 'glory' to mean 'a nice knockdown argument' (MacKay 1968 *cf.* Donnellan 1968). If stipulation is all that matters, then as long as Humpty simply stipulates that 'glory' means 'a nice knockdown argument', then it does so. And this doesn't seem to be the case. Even if the utterance 'glory' *could* mean 'a nice knockdown argument' – if, for example, Humpty was speaking a different language – in the case in question it doesn't, irrespective of Humpty's stipulation.

In the contemporary philosophy of language the discussion of this problem focuses more on the denotation of demonstratives and indexicals rather than proper names, and there are various attempts to prop up existing accounts so as to ensure that a speaker's intentions successfully establish the denotation of demonstratives and indexicals uttered by the speaker (Michaelson, 2013). Whatever the success of these endeavours, their mere existence shows that successfully establishing denotation requires moving beyond a bare appeal to stipulation, or brute intention. But if brute intention fails to establish the appropriate representational relationships for demonstratives and indexicals – the sorts of things that they would seem most applicable to – then it would suggest that they cannot do the work STIPULATIVE FIAT requires of them for establishing representational relationships in general. At the very least, this suggests it needs to be developed if it is to explain how derivative representations represent in virtue of more fundamental representations like intentions and mental states.

Even supposing that stipulation were sufficient to establish that a model represents its target, as an answer to the ER-problem it remains unsatisfactory. In particular, it fails to meet the surrogative reasoning condition: even if it were the case that the salt-shaker represented Madagascar in virtue of Callender or Cohen stipulating that this is the case, this does not explain how the salt-shaker could be used to attempt to learn about Madagascar. And if the salt-shaker was an epistemic representation of Madagascar, then this would have to be the case. As Toon (2010a, 78-79) points out, at best STIPULATIVE FIAT would establish denotational relationships, and these are weaker than the sort of representational relationship that holds between epistemic representations and their subjects. The proper name 'Napoleon' represents Napoleon in virtue of denoting him. But it represents Napoleon in a very different way to the way in which a picture of Napoleon represents Napoleon.[16] In the latter case, we can use the picture to attempt to learn about the French general. In the former case, no matter how we investigate the word 'Napoleon' we will not learn anything we didn't already know.

Supporters of STIPULATIVE FIAT could try to mitigate the force of this objection in two ways. First, they could appeal to additional facts about the object, as well as its relation to other items, in order to account for surrogative reasoning. For instance, the salt shaker being to the right of the pepper mill might allow us to infer that Madagas-

16 Bueno and French (2011) gesture in the same direction when they point to Peirce's distinction between icon, index and symbol and dismiss Callender and Cohen's views on grounds that they cannot explain the obvious differences between different kinds of representations.

car is to the east of Mozambique. Moves of this sort, however, invoke (at least tacitly) a specifiable relation between features of the model and features of the target (similarity, isomorphism, or otherwise), and an invocation of this kind goes beyond mere stipulation.

Callender and Cohen's preferred strategy is to push the epistemic aspect of scientific models into the realm of 'pragmatics'. It's clear that STIPULATIVE FIAT makes scientific representation cheap, so in order to explain why scientists go to so much effort in constructing scientific models – why bother when they could simply stipulate that the clock on their desk represents the target system of interest? – Callender and Cohen accept that different representations can be more or less useful in generating inferences about their targets. But according to them:

> 'the questions about the utility of these representational vehicles are questions about the pragmatics of things that are representational vehicles, not questions about their representational status *per se*' (2006, 75).

This, in effect, amounts to removing the surrogative reasoning condition from the desiderata of an account of scientific representation, and as I have argued previously, this aspect of models is one of the hallmarks of scientific representation, and divorcing the two should not be done lightly. Models don't just 'inform us' that their targets have such and such features, they do this precisely by representing their targets as having those features. It is by representing their targets in a particular manner that they allow us to reason about their targets.

In order for their argument to go through, Callender and Cohen would first need to motivate separating surrogative reasoning from the notion of representation – no such argument is forthcoming in their discussion – and secondly, even if this were the case, an explanation of surrogative reasoning would still be required. Even if it is related to 'pragmatic' features of models, what these features are, and how they licence inferences from models to targets, remains to be spelt out. Given my previous discussions about the tight connection between representation and surrogative reasoning, I'm not optimistic that this could be done without invoking, at least tacitly, the way in which scientific models represent their targets.

So, to briefly summarise, I have argued that General Griceanism does not entail that a relatively trivial trade of the question of scientific representation with the question of how the more fundamental kinds of representations gain their representational capacity. The answer to the ER-problem that they offer – STIPULATIVE FIAT – would make this trivial, but even if it accounted for some sort of representational relationship between models and their targets, it is ultimately unsuccessful in explaining the specific kind of relationship that holds between models, and epistemic representations more generally, and their targets.

Further, assuming that General Griceanism is correct, allowing for the fact that things like the mental states and intentions of model users play a role in establishing that a scientific model represents its target, developing an account of representation that specifies precisely how they do this, and how they do this in such a way that allows us to attempt to learn about targets from their models, is a difficult philosophical task. And in light of Callender and Cohen's discussion of the demarcation problem, even if there is not a special problem of scientific representation, there is a special problem of epistemic representation that remains to be addressed.

Part II

THE FAILURE OF CURRENT ACCOUNTS

3

OVERVIEW

The task for this part of my thesis is to critically evaluate the answers to the ER-problem, and, where applicable, the other problems laid out in Chapter 1 that are currently available in the literature.[17] Articulated attempts to specifically address the ER-problem only arose in the last 20 years or so. So from a philosophical point of view the topic is relatively novel. However, many of the views discussed here have ancestors found in two distinct discussions: aesthetics; and the original presentations of the semantic view of theories. For this reason the following two chapters, which discuss the role of similarity and structural similarity in establishing that models represent their targets, begin with brief simple presentations of their respective answers to the ER-problem.[18] These simple presentations are the primary targets of Frigg (2006) and Suárez (2003), and I do not spend too much time rehearsing their arguments here. Rather, I concentrate on developing more refined versions of answers to the ER-problem that appeal to these similarities. The results, I think, are answers that are more persuasive than even some adherents to those positions have made clear. The confusion arises from thinking that similarity, of structure or otherwise, could have anything to do with representation *simpliciter*. At best, they seem to be notions that establish the accuracy of pre-existing representational relationships by supplying standards of accuracy. However, I argue that they fail even there.

Just as answers to the ER-problem place constraints on the ontological status of models, the structuralist account of scientific representation places constraints on the ontological status of target systems. As such, in Chapter 6 I consider the metaphysi-

17 Although my arguments in this part of the thesis are, by and large, towards negative conclusions, it is instructive to see precisely how the existing accounts of scientific representation fail. The lessons learnt from these failures motivate my own answer to the ER-problem presented in Part iii.

18 Full references will be given in the relevant chapters; but the accounts I have in mind are associated with Giere, Mäki, and Weisberg with respect to similarity, and Suppe, Suppes, van Fraassen, and French and his collaborators with respect to structural similarity. I don't mean to imply here that any of these authors currently subscribe to the naïve version of their accounts. Their specific views will be discussed in more detail in the coming chapters.

cal question concerning 'the structure' of the world. I pay particular attention to van Fraassen's (2008) recent claim that, from the perspective of a model user, accurately representing a target system and a data model extracted from such a system are 'pragmatically equivalent'.

The accounts discussed in Chapters 7 and 8, what I call the 'Direct View' (associated with Toon and Levy, references to come) and 'Inferentialism' (primarily associated with Suárez, but related views are proposed by Hughes, Contessa, and Frigg), are much more recent additions to the debate, and as such have yet to be properly scrutinised. This is as good a time as any to rectify this. So let's get started.

4

MODELS AND SIMILARITY

In this chapter I critically evaluate accounts of scientific representation that rely on the idea that, in some sense, scientific models are similar to their targets. The structure of the chapter is as follows. I begin by outlining the similarity conception in its most basic guise, and review the criticisms to which it has been subjected. I then move onto the more developed versions, defended more recently by Giere (2004, 2010); Mäki (2009, 2011); and Weisberg (2007, 2013), that, rather than relying on similarity *tout court*, rely on the idea that targets and their models are 'similar in the relevant respects, to the appropriate degree'. I argue that the most plausible versions of these accounts – plausible in the sense that they allow for the possibility of misrepresentation – relegate similarity to a condition on accurate representation, rather than representation itself. Moreover, I argue by means of examples, that we should be suspicious about whether 'similarity' is the appropriate notion, even there.

4.1 NAÏVE SIMILARITY

In its simplest guise, the similarity-based account of scientific representation provides the following answer to the ER-problem:

SIMILARITY 1: A scientific model *M* represents a target system *T* if and only if *M* and *T* are similar to one another.

This answer to the ER-problem is beset with difficulties. But it is worth briefly outlining how it meets at least one of the conditions of adequacy laid out in Chapter 1. The view offers an elegant account of surrogative reasoning. Similarities between a model and its target can be exploited to carry over insights gained in the model to the target. If the similarity between *M* and *T* is based on shared properties, then a property found in *M* would also have to be present in *T*; and if the similarity holds between properties themselves, then *T* would have to instantiate properties similar to

M. Thus, from investigating the properties of M, scientists would be able to infer that T had certain properties as well.

The account is also associated with a rejection of the demarcation problem. Given how prominent the notion of similarity is in discussions of representations in non-scientific contexts, it's reasonable to suppose that this conception has universal aspirations in that it is taken to account for representation across a broad range of different domains. Paintings, statues, and drawings are said to represent by being similar to their subjects.[19] And recently Giere, one of the leading contemporary proponents of the similarity view (although by no means currently committed to Similarity 1), claimed that it covers scientific models alongside 'words, equations, diagrams, graphs, photographs, and, increasingly, computer-generated images' (2004, 243 *cf.* 1996, 272, and for further discussion Toon (2012, 249-250)). So the similarity view repudiates the demarcation problem and submits that the same mechanism, namely similarity, underpins different kinds of representation in a broad variety of contexts.

Despite this initial appeal, Similarity 1, is a non-starter as an answer to the ER-problem. Its first major flaw is that it is unclear how models without targets fit into the account. If models represent in virtue of being similar to their targets, then what are we to say about target-less models, like Maxwell's model of the ether (because ether doesn't exist), or populations with more than two sexes?[20] With the exception of Weisberg (2013, Chapter 7) adherents of the similarity account have had little to say about such models. For the purposes of this chapter I do not pursue this argument any further, since it applies to many accounts discussed in the coming chapters, and I pay particular attention to it in Chapter 7.

Secondly, as Goodman (1972) pointed out, similarity has the wrong logical properties to ground representation. Similarity is reflexive and symmetric, yet representation is neither.[21] In general, epistemic representations do not represent themselves, and that X represents Y does not in general ensure that Y represents X.

19 See Abell (2009) and Lopes (2004) for relatively current discussions of similarity – usually called 'resemblance' – in the context of pictorial representation. I take it that 'similarity' and 'resemblance' are synonyms

20 Interestingly, Weisberg (2013, Chapter 7) points out that, despite what Fisher thought when constructing such models, there are some populations that could appropriately be deemed to consist of three sexes. But the point remains that one could construct a five-sex population model, and the similarity account would have little to say about it.

21 Note that in claiming that epistemic representation is not a symmetric relation ($\neg\forall x\forall y(Rxy \rightarrow Ryx)$), I am not claiming that epistemic representation is either antisymmetric ($\forall x\forall y((Rxy \wedge Ryx) \rightarrow x = y)$) or asymmetric ($\forall x\forall y(Rxy \rightarrow \neg Ryx)$). Similarly, I am claiming that epistemic representation is not reflexive ($\neg\forall x Rxx$) rather than anti-reflexive ($\forall x\neg Rxx$).

A partial answer to this objection is to claim that similarity is not symmetric. In empirical studies, Tversky (1977) found that participants were far more likely to agree with the claim that 'North Korea is similar to China' than the claim 'China is similar to North Korea', and when moving to his preferred real valued measure of similarity (discussed in more detail below), participants rated North Korea more similar to China than they did China to North Korea. Or consider Poznic's (2015) example of the Rosemary's baby in the Polanski film. The baby, fathered by the Devil, could be said to be similar to its father. But we are less inclined to judge the Devil as similar to the baby. And if similarity is not symmetric, then this allows for the possibility of models being similar to their targets but not *vice versa*. Therefore, under this understanding of similarity, Similarity 1 allows for the fact that model represent their targets, but not the other way around.

Notice that even if correct, these considerations only concern symmetry, not reflexivity. It remains (e.g. for Weisberg (2013) who uses Tversky's account to answer the ER-problem, and others) that similarity is a reflexive relation. Moreover, however the issue of the logical properties is resolved, serious problems for Similarity 1 remain. The first basic problem is that the notion of 'similarity' needs to be qualified. Intuitively, what is required for two objects X and Y to be similar to one another is that they share some properties. But if this is all that is required then we are left with the result that everything is similar to everything else, since everything shares *some* property with anything else (Goodman, 1972). This can be easily seen under a extensional understanding of properties – where a property P is identified with its extension (and a two-place relation identified with the set of ordered pairs thus related, and so on) – since with no restriction on what sets we consider we can easily construct shared properties for an arbitrary pair of objects X and Y. Simply consider the property of being $\in \{X, Y\}$. In fact, we can construct an uncountably infinite number of such properties. Simply consider the properties of being $\in \{X, Y, i\}$ for all $i \in \mathbb{R}$.

Even if we deny that membership in these gerrymandered sets genuinely constitutes a shared property, it remains to be seen how we can distinguish between 'genuine' properties and gerrymandered ones in such a way that avoids the conclusion that any pair of objects are similar in some sense. For example any pair of objects presented to me as potentially similar to one another would share the property of being presented to James at a certain time t, being thought about by James at t, and so on. Alternatively, any pair of earthbound objects share the property of being earthbound (which would seem like a genuine property). Any pair of non-earthbound objects share the property of not being earthbound. And any pair of objects such that one is earthbound and the other is not, presumably share properties like being a certain distance from the midpoint of a 'straight-line' between them. For Similarity 1, this has the unfortunate consequence that anything represents anything else because any two objects are similar in some respect. Unqualified similarity is just too easy to come by to account for representation.

An obvious response to this problem is to delineate a set of relevant respects and degrees to which *M* and *T* have to be similar in order for the former to represent the latter. The 'relevant respects' aspect allows us to isolate a particular set of properties to be shared by similar objects, and the 'relevant degrees' aspect allows for objects to share some, but not all of these properties.[22]

Taking this into account, we arrive at the following answer to the ER-problem:

SIMILARITY 2: A scientific model *M* represents a target system *T* if and only if *M* and *T* are similar to one another in the relevant respects and to the appropriate degree.

I address the question of how these respects and degrees are to be specified shortly. But I suspect that, however this is done, SIMILARITY 2 is going to face the following objection. Even similarity suitably restricted remains insufficient on epistemic representation. Relying on an intuitive understanding of the relevant respects and degrees, it would seem that two copies of the same newspaper are relevantly similar to one another, but neither represents the other. Two identical twins are relevantly similar to one another, yet neither represents the other. There are numerous cases where two items are similar with no representational relationship between them. And this won't go away even if similarity turns out to be non-symmetric. That North Korea is similar to China (to some degree) does not imply that North Korea represents China, and that China is not similar to North Korea to the same degree does not alter this conclusion.

The problem is highlighted in Putnam's (1981) thought experiment involving an ant tracing a shape through the sand. We can suppose that the shape of the trace bears a relevant similarity (or resemblance) to Churchill: the shape includes a cigar shape, joined to a mouth shape, above a two-fingered salute shape. Has the ant produced a picture of Churchill? Putnam claims, and I agree, that it hasn't because the ant has never seen Churchill, is causally isolated from Churchill, and had no intention of producing an image of him. Although someone else might see the trace as a depiction of Churchill, the trace itself does not represent Churchill. This, Putnam concludes,

22 Or alternatively moves us from the idea that objects have to co-instantiate properties in order to be similar to one another to the idea that they instantiate appropriately similar properties. I discuss the distinction between these two ways of thinking about similarity below, but for an intuitive example, consider how we understand a London bus and a British telephone box to be similar to each other with respect to their colour. We might do so because they co-instantiate redness, or because they instantiate particular shades of red, which are in turn appropriately similar to each other.

shows that 'similarity [...] to the features of Winston Churchill is not sufficient to make something represent or refer to Churchill' (Putnam, 1981, 1). And what is true of the trace and Churchill is true of every other pair of similar items: similarity, relevant or not, on its own does not suffice to establish representation.

A subtly different observation is made by Suárez (2003, 233-234). Using the example of Velázquez's portrait of Pope Innocent X, we can consider someone cleverly disguising themselves as the Pope; dressing up in a red shawl and hat, growing their facial hair, seating themselves on a gilded chair, and so on. Now if it is the relevant similarity between the painting and the Pope that grounded the fact that the former represented the latter, then the same similarity which holds between the painting and the cleverly disguised individual (in virtue of the disguise) should suffice to establish that it represents him too. And it doesn't seem that this is the case. In cases like these, which Suárez calls 'mistargeting', a model represents one target rather than another, despite the fact that both targets are relevantly similar to the model. As in the case of Putnam's ant, the root cause of the problem is that the similarity is, in some sense, accidental. In the case of the ant, the accident occurs at the representation-end of the relation, whereas in the case of the disguised man, the accidental similarity occurs at the target-end. Both cases demonstrate that SIMILARITY 2 cannot rule out accidental representation.

I think this is enough to demonstrate that SIMILARITY 2 won't work as an answer to the ER-problem, but it is worth emphasising that it faces another crucial objection: it cannot handle models that misrepresent their targets. Recall that M misrepresents T if and only if M represents T, but does so inaccurately or unfaithfully. Phrasing this in terms of similarity, it would be reasonable to suggest that M misrepresents T by portraying it as sharing certain relevant properties (to the appropriate degree) with M that, it turns out, are *not* shared by T. Consider again Thomson's plum pudding model, which represents atomic structure as negative charge bearing electrons surrounded by positive charge in a way analogous to raisins being surrounded by suet in a plum pudding. One would think that the relevant similarity between the model and the target here would be the way in which positive and negative charge are distributed. But we know that in the case of atoms positive charge is concentrated in the atomic nucleus, and thus the model is not similar to the target in the relevant respect. But then, on SIMILARITY 2, the model does not represent its target at all, since the relevant similarity doesn't hold.[23] But as I argued in Chapter 1, the possibility of misrepresentation is a condition of adequacy for any acceptable account of scientific representation, and

23 Ducheyne (2008) embraces this conclusion when he offers a variant of a similarity account which explicitly takes the success of the hypothesised similarity between a model and its target to be a necessary condition on the model representing the target.

so misrepresentation should not be conflated with non-representation in the way that SIMILARITY 2 would seem to entail. Thomson's model still represents atomic structure, it just does so inaccurately.

Finally, notice that even restricted to relevant similarities, it remains that every object is maximally similar to itself. So similarity remains a reflexive relation, and since representation is not, SIMILARITY 2 still provides the wrong logical properties to the representation relation. This, combined with the above objections, shows that it fails as an answer to the ER-problem. It cannot deal with accidental similarities between models and systems in the world that they don't represent, and it doesn't allow for the possibility of models representing their targets. This is notwithstanding the fact that the notion of 'relevantly similar to the appropriate degree' has yet to be spelt out.

One aspect of this phrase that I have ignored so far is its pragmatic, context sensitive character. In order for a similarity to be 'relevant', it needs to be relevant *for something*. And in the case of epistemic representation, it would seem that this notion of relevance will be sensitive to the purposes, goals, and intentions of the model user. It's plausible that in order to account for this, an adherent to the similarity approach will need to shift from what has been called a 'naturalistic' account of representation to a non-naturalistic one.[24] The question then is whether non-naturalistic versions of the similarity account fare any better than the ones considered in this section.

4.2 NUANCED SIMILARITY

One way to specify the relevant similarities that need to hold to the appropriate degrees in order for a model to represent its target is to utilise what Giere (1988, 81) calls 'theoretical hypotheses': statements asserting that model and target are similar in relevant respects and to certain degrees. This idea can be moulded into the following answer to the ER-problem.

SIMILARITY 3: A scientific model M represents a target system T if and only if a model user provides a theoretical hypothesis H specifying that M and T are similar to one another in the relevant respects and to the appropriate degree.

This avoids many of the problems discussed in the previous section. The requirement of theoretical hypotheses linking models and their targets means that even if similarity is reflexive (and possibly symmetric), it needn't be the case that scientific

24 Suárez, drawing on van Fraassen (2002) and Putnam (2002), defines 'naturalistic' accounts of representation as ones where 'whether or not representation obtains depends on facts about the world and does not in any way answer to the purposes, views or interests of enquirers' (2003, 226-227).

representation is. Just because everything is similar to itself doesn't mean that someone provides a theoretical hypothesis specifying that an object is similar to itself. Moreover, accidental similarities, like those involved in Putnam's ant and the Pope Innocent impostor, are circumvented since there is no theoretical hypothesis to specify that the representational vehicles are similar to their respective targets. Allowing for the possibility of misrepresentation requires that the theoretical hypotheses are fallible; they could be false. The idea being that the theoretical hypothesis establishes the representation relation, and then if the theoretical hypothesis is true (or 'true enough'), then this establishes that the representation relation is accurate or faithful.

SIMILARITY 3 thus looks like a promising answer to the ER-problem. And once we pay attention to the fact that the theoretical hypotheses are provided by a model user, and so are sensitive to their intentions and purposes, it comes close to Giere (2004, 2010) and Mäki's (2009; 2011) considered views on scientific representation. The former adopts an agent-based account of scientific representation which focuses on 'the activity of representing' (2004, 743). Analysing representation in these terms amounts to analysing schemes like '*S* uses *X* to represent *W* for purposes *P*' (2004, 743), or in more detail:

> 'Agents (1) intend; (2) to use model, *M*; (3) to represent a part of the world *W*; (4) for purposes, *P*. So agents specify which similarities are intended and for what purpose' (2010, 274).

To account for this in SIMILARITY 3, the purposes of the model users guide the content of the theoretical hypothesis: they specify which similarities are relevant, and to what degree. The intentions of the model agent come in via the act of offering the theoretical hypothesis: a model user intends to use a model to represent a target by providing such a hypothesis.

Mäki offers an extension of Giere's view, which adds two conditions to Giere's: the agent uses the model to address an audience *E* and adds a commentary *C* (Mäki 2011, 55-57 *cf.* Mäki 2009). The addition of an audience makes the activity of representing a communal one. It is unclear to me why Mäki demands this, since it rules out cases where scientists develop, and use, models in isolation; which, at least *prima facie*, would still seem to be cases of scientific representation. The role of the commentary is to specify the nature of the similarity. This is needed because

> 'representation does not require that all parts of the model resemble the target in all or just any arbitrary respects, or that the issue of resemblance legitimately arises in regard to all parts. The relevant model parts and

> the relevant respects and degrees of resemblance must be delimited' (Mäki, 2011, 57).

What these relevant respects and degrees of resemblance are depends on the purposes of the instance of scientific representation in question. These are not determined 'in the model' as it were, but are pragmatic features of the contexts in which it is used. From this it transpires that in effect *C* plays the same role as that played by theoretical hypotheses in Giere's account. Certain aspects of *M* are chosen as those relevant to the representational relationship between *M* and *T*.

So, does SIMILARITY 3 provide a successful similarity-based account of epistemic representation? Unfortunately not. A closer look reveals that the role of similarity has shifted. As far as offering a solution to the ER-Problem is concerned, all the heavy lifting in SIMILARITY 3 is done by the appeal to agents and their specification of the theoretical hypothesis. Giere implicitly admits this when he writes:

> 'How do scientists use models to represent aspects of the world? What is it about models that makes it possible to use them in this way? *One way, perhaps the most important way, but probably not the only way,* is by exploiting similarities between a model and that aspect of the world it is being used to represent. Note that I am not saying that the model itself represents an aspect of the world because it is similar to that aspect. There is no such representational relationship. [footnote omitted] Anything is similar to anything else in countless respects, but not anything represents anything else. *It is not the model that is doing the representing; it is the scientist using the model who is doing the representing*' (2004, 747, emphasis added).

But if similarity is not the only way in which a model can be used as a representation, and if it is the use by a scientist that turns a model into a representation (rather than any mind-independent relationship the model bears to the target), then similarity has become otiose in a reply to the ER-problem. A scientist could invoke any relation between *M* and *T* and *M* would still represent *T*.

Moreover, on Mäki's account, he is keen to stress that:

> 'Naturally, an account of scientific representation must accommodate failures. The notion of 'prompting issues of resemblance to arise' starts taking care of this. Failing to prompt those issues is a major failure in representation (and it is here that 'mis-representation' may be appropriately applied), while failing to resemble is a lesser failure. Respectively, prompting issues of resemblance gives us weak success, while succeeding to resemble is a matter of stronger success' (2011, 57).

For sure, this move ensures that SIMILARITY 3 can account for misrepresentations, but it's the 'prompting' issues of similarity, or resemblance, that establishes the prior representation relationships, rather than similarity itself. Being similar in the relevant respects to the relevant degree now provides the standards of accuracy of representation, rather than grounding it *per se*. But before investigating the role it plays there, it's worth getting more specific about what it means for two objects to be similar to one another.

4.3 WHAT IS SIMILARITY ANYWAY?

Unfortunately the philosophical literature contains surprisingly little explicit discussion about what it means for something to be similar to something else. In many cases similarity is taken to be primitive, possible worlds semantics being a prime example. The problem is then compounded by the fact that the focus is on comparative overall similarity, rather than on similarity in respects and degrees (for a critical discussion see Morreau (2010)). Where the issue is discussed explicitly, the standard way of cashing out what it means for an object to be similar to another object is to require that they co-instantiate properties. This is the idea that Quine (1969, 117-118) and Goodman (1972, 443) had in mind in their influential critiques of the notion. They note that if all that is required for two things to be similar is that they co-instantiate some property, then everything is similar to everything else, since any pair of objects have at least one property in common.

The issue of similarity seems to have attracted more attention in psychology. In fact, the psychological literature provides attempts to capture it directly in more fully worked out formal frameworks. The two most prominent suggestions are the geometric and contrast accounts (see Decock and Douven (2011) for an up-to-date discussion). It's worth briefly outlining both of these accounts so we can get clear about what it takes for two objects to be similar to one another before investigating whether this notion can bear the weight of providing standards of accuracy.

4.3.1 *The geometric account*

The geometric account takes similarity relations to be captured by a metric space. It was associated originally with Shepard (1980). Formally, a metric space is a pair $\langle X, \delta \rangle$, where X is a set of points, and δ a function from pairs of those points to the positive real numbers, including 0, that satisfies three conditions required to be a distance function. That is $\delta : X \times X \to \mathbb{R}_0^+$ such that for all $x, y, z \in X$:

MINIMALITY: $\delta(x,y) \geq 0$ and $\delta(x,y) = 0 \Leftrightarrow x = y$,
SYMMETRY: $\delta(x,y) = \delta(y,x)$, and
TRIANGLE INEQUALITY: $\delta(x,y) + \delta(y,z) \geq \delta(x,z)$.

This allows us to compare similarity relations across pairs of objects located in the space in a straightforward manner. x is more similar to y than z is to w if and only if $\delta(x,y) \leq \delta(z,w)$. Moreover, if we want to generate an absolute measure of similarity we can find a threshold $t \in \mathbb{R}_0^+$ such that x is similar to y if and only if $\delta(x,y) \leq t$. This would allow us to account for how models and targets are similar to one another: each of them are located at a point in X based on the properties they instantiate, and then δ can be used as a measure of the dissimilarity between the two (possibly using a function of δ as a measure of similarity between the two, to ensure that the lower the value of δ the higher the measure of similarity).

Although at first glance it looks as though the geometric account of similarity is committed to the idea that similarity is a symmetric relation (in virtue of δ having to meet the symmetry condition), and is insensitive to the context and purposes of model users, Gärdenfors (2000) provides a development in the context of his 'conceptual spaces' framework that could be utilised by a defender of the geometric account of similarity. Rather than assuming there is only one metric space under which to measure how similar two objects are, we can consider multiple spaces. When we're interested in whether or not two objects are similar with respect to their colour, we can use a colour space; when we are interested in whether or not two objects are located in similar positions in space, we can use a three dimensional Euclidean space; and so on. Thus the purposes and context of using a particular model to represent a target system will provide a set of relevant similarities, which in turn determine a metric space. Moreover, there is nothing to say that the same space has to be used when measuring how similar an object a is to an object b and when measuring how similar b is to a. And if a different space is used in each of these situations, then it needn't be the case the similarity is a symmetric relation.[25]

These considerations suggest that for each metric space, although all that is required is that X be a set of points, it can have a richer dimensional structure. The idea being that each dimension corresponds to some way in which two objects can be similar (their blue hue, their red hue, their extension in a particular direction or another, and so on). This allows us to make sense of Niiniluoto's (1988, 272-274) distinction between two different kinds of similarities: what he calls 'partial identity' and 'likeness' respec-

25 Although notice that if the relevant similarities turn out to be the same in each case, then the same space will be used and therefore the similarity measure will have to be symmetric.

tively.[26] Assume M instantiates the relevant properties $P_1, ..., P_n$ and T instantiates the relevant properties $Q_1, ..., Q_n$. If these properties are identical, i.e. if $P_i = Q_i$ for all i, then M and T are similar in the sense of being partially identical. Partial identity contrasts with likeness. M and T are similar in the sense of likeness if the properties are not identical but similar themselves: P_i is similar to Q_i for all i. So in likeness the similarity is located at the level of the properties themselves. For example, a red post box and a red London bus are similar with respect to their colour, even if they do not instantiate the exact same shade of red. According to the geometric account of similarity, partial identity is captured when M and T are located at the same point on each of the i dimensions associated with the relevant properties. Likeness is captured as long as the restriction of δ to each of those dimensions is 'small' enough, i.e. $\delta_i(M, T) \leq \epsilon$ for each of the i dimensions and for some ϵ.

To the best of my knowledge no one has developed the geometric account of similarity in the context of scientific representation. And I won't do so here, since I'm sceptical it will work. However, it is worth briefly sketching what it might look like. Recall from the previous discussion that, in order to account for the possibility of misrepresentation, the similarity account had to fall back on the idea that it's the *proposal* that M and T are similar that establishes that M represents T. This is done by a theoretical hypothesis that specifies the relevant respects in which this is proposed to be the case. This can be taken as specifying a metric space $\langle X, \delta \rangle$, where each of these relevant respects corresponds to a dimension in X. Giere's notion of the 'appropriate degree' can then be read as providing a threshold $t \in \mathbb{R}_0^+$ whereby it is proposed that $\delta(M, T) \leq t$. Notice that this is a proposal, and does not guarantee that M and T are in fact thus similar. Then, if actual similarity between M and T is taken as providing the standards of accuracy, then so long as it is the case that $\delta(M, T) \leq t$ we can say that M accurately, or faithfully, represents T.[27]

This is not without its merits, but it suffers from a serious flaw. In order to produce a metric space, the properties that are shared by models and their targets must be quantitative, in the sense that 'distances' between the points in the space are measurable by a real number. Assuming each dimension of the space corresponds to a single relevant property, we need to be able to be able to assign a real number to the comparison of how M instantiates that property to how T instantiates it. For some model-world comparisons this seems quite natural. We might compare the period of a model pendulum with the period of the pendulum in Big Ben, thereby locating the

26 These also feature in Hesse's discussion of analogies, see, for instance Hesse (1963, 66-67)

27 Notice that this can be sensitive to the purposes of the model users. Whether or not M accurately, or faithfully, represents T is only defined with respect to a theoretical hypothesis which provides a threshold t. In different context, this threshold may take different values.

model with respect to the target on a dimension corresponding to their periods. Even for properties that we might initially treat as qualitative – e.g. colour – we might be able to construct a metric space – e.g. an RGB colour space where each dimension corresponds to measure of chromaticity of red, green, or blue. But it is a very strong assumption that this can be done *for all* model-target comparisons.

I suspect that even if it could, then the way of accommodating them will make significant assumptions regarding both the 'scale' (or scales) of the space, and the way in which we can compare these comparisons across its dimensions. What I have in mind here is the fact that, by the definition of a metric space, comparisons of models with their targets on multiple dimensions need to be commensurable in some sense. Since δ is defined everywhere on $X \times X$, it requires, for example, that some distance between M and T with respect to a property P will need to be compared to the distance between M and T with respect to another property P'. We need to be able to make sense of the claim that distances in one dimension are shorter, longer, or the same as, distances in another dimension. And, given the fact that scales in each of the dimensions seem to be, in some sense, arbitrary (for example, should we measure the length of a pendulum in centimetres or inches?), how to make these comparisons remains unclear.[28] So without more being said about what justifies these inter-dimensional comparisons of similarity, and how to capture purely qualitative comparisons between properties of models and their targets, the geometric way of defining similarly does not seem like the appropriate formal framework in which to think about similarity as providing the standards of accuracy.

4.3.2 *Weighted feature matching*

The problem concerning how to compare the qualitative properties of objects with respect to their similarity is supposed to be overcome in Tversky's contrast account (1977). This account defines a gradated notion of similarity based on a weighted comparison of properties. Weisberg (2012; 2013, Chap. 8) has recently introduced this

28 In economics this corresponds to the so-called 'interpersonal comparisons of utility' (Robbins, 1938). The mathematical considerations are analogous to one another. If an agent's utility function is only specifiable up to affine transformation how do we compare specific measures of utility across agents? Similarly, if the scales on which models and their targets instantiate properties are only specifiable up to some transformation (which seems to be the case), how do we compare specific distances on each dimension with one another? It's worth noting that in the social choice literature there are some suggestions. If, for example, each agent's utility function can be measured on a ratio scale then they can be compared even if the scales are different across agents (Tsui and Weymark, 1997). But this isn't a question that has been addressed in the discussion of similarity as applied to models and their targets.

account into the philosophy of science where it serves as the starting point for his so-called 'weighted feature matching account of model world relations'.

The account introduces a set Δ of relevant properties. Let $\Delta_M \subseteq \Delta$ be the set of relevant properties instantiated by the model M; and likewise $\Delta_T \subseteq \Delta$ the set of relevant properties instantiated by the target system T. Furthermore let f be a ranking function assigning a positive real number to every subset of Δ.[29] The simplest version of a ranking function is one that assigns to each set the number of properties in the set, but rankings can be more complex, for instance by giving important properties more weight. The level of similarity between M and T is then given by the following equation (Weisberg 2012, 788, *cf.* Weisberg 2013, 144, the notation is slightly amended):

$$S(M,T) = \theta f(\Delta_M \cap \Delta_T) - \alpha f(\Delta_M - \Delta_T) - \beta f(\Delta_T - \Delta_M)$$

where α, β, and θ are weights, which can in principle take any positive value. This equation provides a 'similarity score that can be used in comparative judgements of similarity' (Weisberg, 2012, 788). The score is determined by weighing the properties the model and target have in common against those they do not.[30] In the above formulation the similarity score S can in principle vary between any two values (depending on the choice of the ranking function and the value of the weights), which makes it difficult to compare multiple similarity scores. One can then renormalise S so that it takes values in the unit interval as follows:[31,32]

$$S(M,T) = \frac{\theta f(\Delta_M \cap \Delta_T)}{\theta f(\Delta_M \cap \Delta_T) + \alpha f(\Delta_M - \Delta_T) + \beta f(\Delta_T - \Delta_M)}$$

The obvious question at this point is how the various blanks in the account can be filled. First in line is the specification of a property set Δ. Weisberg is explicit that there are no general rules to rely on and that 'the elements of Δ come from a combination of context, conceptualization of the target, and theoretical goals of the scientist' (2013, 149). Likewise, the ranking function, as well as the values of weighting parameters,

29 For non-trivial Δs, requiring that f be defined on $\mathcal{P}(\Delta)$ (i.e. the power set of Δ) is a significant assumption. Weisberg (2013, 152) concedes this and suggests that ensuring that f is additive, in the sense that $f(X \cup Y) = f(X) + f(Y)$ for $X, Y \in \Delta$, is a plausible simplifying assumption.

30 Thus this account could be seen as a quantitative version of Hesse's (1963) theory of analogy in which properties that M and T share are the positive analogy and ones they don't share are the negative analogy

31 This equation does not appear in Weisberg's presentations of his account because he further divides the elements of Δ into attributes and mechanisms. The former are the 'the properties and patterns of a system [while the latter are the] underlying mechanism[s] that generates these properties' (2013, 145). This distinction is helpful in the application to concrete cases, but for my purposes can be set aside.

32 I follow Weisberg by using S to denote both the normalised and non-normalised similarity measure, but it bears noting that these are different functions.

depend on the goals of the investigation, the context, and the theoretical framework in which the scientists operate. The problem of how to weigh similarities between a model and its target in multiple dimensions raises its head again here. The ranking function f assigns a real number to sets of properties and when calculating S these numbers are added and subtracted to each other. The weights α, β, and θ allow that these arithmetical operations can take weighted values of f as arguments but the fact remains that the definition of f requires comparisons across dimensions.

Weisberg is keen to stress that given the parameters in the above equation, his account is compatible with pretty much *any* such comparison:

> 'With no parameters set and no weighting function defined, the equation describes an infinite set of potential relations—at least one of which almost certainly holds between a model and a target' (2015, 300).

But as a result, this means that the context of the model-target comparison has to fix all of these before S can be used. Again, it is unclear what basis there could be for such comparisons.

Irrespective of these choices, the similarity score S has a number of interesting features. First, it is asymmetrical for $\alpha \neq \beta$, which makes room for the possibility of M being similar to T to a different degree than T is similar to M. So S provides the asymmetrical notion of similarity mentioned previously. Second, S has a property called maximality: everything is maximally similar to itself. Formally: $S(x, x) = 1 \geq S(x, y)$ for all objects x, y (Weisberg, 2013, 154). Thirdly, at least in some cases, and in contrast to the geometric account, it allows us to compare models and their targets with respect to their qualitative properties.

However, since the account is an elaborate version of the co-instantiation account of similarity it cannot overcome that account's basic limitations. Recall Niiniluoto's distinction between partial identity, which requires co-instantiating various properties, and likeness which requires co-instantiating appropriately similar properties (which, according to the geometric account corresponded to being located close to another another on various dimensions in the appropriate metric space). As Parker (2015, 273) notes, Weisberg's account (like all co-instantiation accounts) deals well with partial identity, but has no systematic place for likeness. To deal with likeness Weisberg would in effect have to reduce likeness to partial identity by introducing 'imprecise' properties into Δ. Parker (2015, 273) suggests that this can be done by introducing interval valued properties, for instance of the form 'the value of feature X lies in the interval $(x - \epsilon, x + \epsilon)$' where ϵ is a parameter specifying the precision of overlap. To illustrate she uses Weisberg's example of the US Corps' San Francisco Bay model and claims that in order to account for the similarity between the model and the actual Bay with respect to their Froude number Weisberg would have to claim something like:

> 'The Bay model and the real Bay share the property of having a Froude number that is within 0.1 of the real Bay's number. It is more natural to say that the Bay model and the real Bay have similar Froude numbers-similar in the sense that their values differ by at most 0.1' (Parker, 2015, 273).

In his response Weisberg (2015) accepts this and argues that he is trying to provide a reductive account of similarity that bottoms out in properties shared and those not shared. But such interval-valued properties have to be elements in Δ in order for the formal account to capture them. This means that another important decision regarding whether or not M and T are similar occurs outside of the formal account itself. The inclusion criteria on what goes into Δ now not only have to delineate relevant properties, but, at least for the quantitative ones, also have to provide an interval defining when they qualify as similar. Furthermore, it remains unclear how to account for M and T to be alike with respect to their qualitative properties. The similarity between genuinely qualitative properties cannot be accounted for in terms of numerical intervals. This is a particularly pressing problem for Weisberg, because he takes the ability to compare models and their targets with respect to their qualitative properties as a central desideratum for any account of similarity between the two (2013, 136).

In sum, although I think that the weighted feature matching account provides a more tenable formal framework in which to analyse similarities between models and their targets, it remains to be seen whether it can make good on its promise of allowing for comparisons with respect to qualitative properties. Moreover, as noted above, many of the crucial elements are determined outside of the framework (the relevant properties, weights and so on), so more needs to be said about how these elements are fixed prior to using the framework to measure model-target similarities.

4.4 SIMILARITY, STYLE, AND ACCURACY

Let's suppose that the questions raised for both the geometric and weighted feature matching accounts of similarity can be answered. And recall from the previous discussions that the most plausible version of a similarity based account of scientific representation takes the intentions of model users and their theoretical hypotheses as answering the ER-problem, with the question of whether or not models are so similar to their targets providing the standards of accurate representation. The question to ask now is whether similarity can, or does, function in this way. But before investigating this, it is worth briefly outlining how similarity might play a role in answering the problem of style, since this will allow us to get to grips with the aspirations of such an account.

Recall that the problem of style originates from the observation that different scientific models represent their targets in different ways. In the case of art, where different pictures (for example) represent their subjects in different ways, styles could be categorised by appealing to the artistic traditions in which the works are situated. In the scientific context, the question is whether models can be categorised in an analogous manner. To see how similarity might be useful in such a categorisation scheme, it is useful to think about similarity and the problem of style on both a coarse and a fine grained level.

On a more fine grained level, the flexibility of the aforementioned frameworks might allow for different styles of scientific representation to be categorised according to how the parameters are determined. The geometric account allows for different metric spaces to be used in different contexts. And if these geometric accounts can be categorised in some way (depending presumably on the dimensions that appear in the metric spaces) then this would provide a viable way of categorising at least some representations. On Weisberg's contrast account we could use his distinction between mechanisms and attributes to distinguish between different styles of representation, e.g. those that aim to be similar with respect to their attributes without being similar with respect to their mechanisms could be grouped together under one style of scientific representation.[33] Or we could further distinguish between types of properties shared, and then further fine grain the similarity measure S in the appropriate way. I think both of these suggestions are viable avenues of further research and could provide interesting ways of categorising scientific representations into different styles. However, an important question that needs to be addressed first is whether the similarity based approaches can cover *all* styles of scientific representation.

Answering the problem of style on a coarse grained level could require that similarity is taken to be one among many different scientific styles. In some instances, scientists may set out to construct a model that is similar to its target (in a manner cashed out using either of the above formal frameworks), and their intentions and theoretical hypotheses would function to establish the appropriate representation relation. Whether or not such models turned out to be thus similar could then determine whether the model was fruitful or accurate. I think that this is a plausible suggestion. I grant that similarity plays a role in the case of some (or even most) cases of scientific representation. But it is clear that adherents of the similarity approach think that

33 Weisberg (2013, 118, 148) suggests something like this when he discusses 'how possibly' modelling, which could be seen as a specific style of scientific representation.

similarity isn't one style among many, but rather underpins all scientific representation, and depending on their perspective on the demarcation problem, perhaps even epistemic representation in general.

I think these universal aspirations are problematic once we start to consider what the relevant properties of the model and targets are when it comes to comparing them. To warm up, consider the case of the London Tube map as an epistemic representation of the London Underground system and suppose your purposes for using the map are to determine whether you can get a direct Victoria line tube from Brixton to King's Cross. The relevant properties of the map are whether or not there is a light blue line between the circle marked 'Brixton' and the circle marked 'King's Cross'. The relevant property of the London Underground system is the fact that, as it turns out, the Victoria Line *does* run from Brixton to King's Cross. Now if we were to apply the weighted feature matching account to this example of using an epistemic representation we are faced with a troubling situation: what properties go in Δ? Intuitively, we want to say that the relevant property comparison here is between a colour property on the map, and a tube line property in the London Underground. But it's not clear that these can appropriately be taken to be the same property.[34] Now, one could try to claim that the relevant property is a topological one, the dots marked 'Brixton' and 'King's Cross' are connected to one another in the same way that Brixton and King's Cross are connected, and this property is shared between the map and the Underground. But this strategy won't fly. The property of the map that tells the map user that they can get a direct line is a colour one, otherwise you might have to change lines. There is no escaping the fact that it is colour properties that are associated with tube lines. And thus, it doesn't seem as though it is similarity, in the sense that the map and the Underground system instantiate the same property, that establishes that the map to be an accurate representation in this respect. Similarly, it's not obvious whether any model user would *intend* that the map and the Underground system were similar with respect to either property when establishing the representation relation in the first place. Rather, they associate colours with tube lines conventionally, and this isn't the sort of association that can be cashed out in terms of co-instantiating properties.

Conventional associations between properties of models and their targets are not limited to cartographic representation. For scientific examples, we can consider cases such as using litmus paper to represent the pH of a solution. Dip the strand of litmus in an acidic solution; if it turns red it accurately represents the solution as acidic. Are we happy to grant that the red litmus paper and acidic solution are 'similar' in virtue of the former's colour and the latter's pH? Or consider the Phillips-Newlyn machine.

34 And it is not clear that the geometric account will fare any better in this comparison, as these two properties do not seem to correspond to any dimension in a metric space.

Figure 6: The London tube map (TFL)[36]

Prima facie, it is a hydraulic machine consisting of reservoirs, pipes, and valves through which a pump pumps water depending on the width of the valves and shape of the tanks.[35] The machine is a model that represents an economy, but it remains unclear whether a theoretical hypothesis will assert that the two are similar (which properties of the hydraulic machine could be co-instantiated by an economy?).

Additional problems for using the notion of similarity as a universal condition on accurate representation (and relatedly, taking the prior representation-establishing theoretical hypotheses to universally specify similarities between models and their targets, rather than any other relation) occur when, as in some cases of epistemic representation, *distortions* play a vital role in allowing us to use representational vehicles to learn about their targets. Van Fraassen claims that it 'seems then that distortion, infidelity, lack of resemblance in some respect, may in general be crucial to the success of a

35 I present more details about the machine in Chapter 9, for now I hope this brief description suffices to illustrate my point. See Morgan (2012, Chapter 5) for further discussion.

representation' (2008, 13). He illustrates this by means of caricature (which I return to in Chapter 6 in more detail). We can think of a caricature of Margaret Thatcher that represents her as draconian. Since we know that Margaret Thatcher does not, in fact, have wings and a tail, or breathe fire, it seems that the picture is not similar to Thatcher in these respects. Yet it is these respects that play a vital role in the caricature being an accurate or faithful representation of her.[37] It's plausible that this holds in the scientific context as well. Van Fraassen uses the example of representing, by means of a drawing, two differently orientated parallelograms as being congruent to each other (2008, 13). In order to do this successfully (accurately, faithfully) they have to be drawn at different sizes, and thus be dissimilar with respect to their size. Or alternatively, we can consider idealisation assumptions in scientific models that play a vital role in their success, i.e. the accuracy of their representation, despite the fact that in these respects the model and the target are explicitly dissimilar (assuming they cannot be captured by Weisberg's interval-valued property approach). In taking proposed similarity as a universal style of representation, and thereby actual similarity as a universal standard of accurate representation, we seem committed to the idea that models have to be copies of their targets (even if only in the relevant respects to the appropriate degree) in order to be accurate. And I'm sceptical that this is the right way of thinking about scientific, or epistemic, representation in general.

For those not convinced, it is worth noting that I'm not claiming that there is no way in which these examples could be accommodated within the similarity framework. But my suspicion is that once they are so, the notion of similarity that drops out becomes relatively unhelpful. If the notion of similarity used is so flexible that it captures the aforementioned cases, then it becomes unclear whether or not the result still corresponds to our pre-theoretical understanding of similarity. And as I discuss in my own positive proposal in Chapter 9, if all the work that the notion does can be accommodated elsewhere, I see little motivation for invoking it in the first place. So to summarise where we are so far. I have argued that similarity itself cannot answer the ER-problem, at pain of conflating misrepresentation with non-representation. Moreover, even if *proposed* similarity, by means of a theoretical hypothesis, may answer the ER-problem in at least some cases, and consequently the success of this proposal providing standards of accuracy, it remains unclear whether this provides a universal answer to the problem of style. Some representations are not proposed to be similar to their targets in the relevant respects, and their success is often specifically down to the fact that they are dissimilar from them. So I think that accounts that rely solely

36 Available here `http://content.tfl.gov.uk/standard-tube-map.pdf`.

37 Those with opposing political viewpoints are welcome to substitute another caricature in place of this, one of Jeremy Corbyn perhaps.

on the notion of similarity fail to fully capture what is interesting about epistemic and scientific representation. Before moving on, however, there is yet another problem that faces the similarity theorist that needs to be discussed.

4.5 similarity and ontology

If models are supposed to be similar to their targets in the ways specified by theoretical hypotheses or commentaries, then they must be the kind of things that can be so similar.

Some models are familiar physical objects. The Army Corps of Engineers' model of the San Francisco Bay is a water basin and equipped with pumps to simulate the action of tidal flows (Weisberg, 2013); ball and stick models of molecules are made of plasticine, metal or wood (Toon, 2011; de Chadarevian, 2004); the Phillips-Newlyn model of an economy is a system of pipes and reservoirs (Morgan, 2012); model organisms in biology are animals like worms and mice (Ankeny and Leonelli, 2011); a model ship is a block of wood in a tank (Sterrett, 2006); and water-based dumb holes are used as models for black holes (Dardashti et al., 2015). For models of this kind similarity is straightforward (at least in principle) because they are of the same ontological kind as their respective targets: they are material objects. But many interesting scientific models are not like this. Two perfect spheres with a homogeneous mass distribution which interact only with each other (the Newtonian model of the sun-earth system), or a single-species population isolated from its environment and reproducing at fixed rate at equidistant time steps (the logistic growth model of a population) are what Hacking elegantly describes as 'something you hold in your head rather than your hands' (1983, 216). As a neutral term, I call these 'non-physical models'. The question then is what kind of objects non-physical models are. Giere submits that they are abstract objects: 'models in advanced sciences such as physics and biology should be abstract objects constructed in conformity with appropriate general principles and specific conditions' (2004, 745 *cf.* 1988, 81; 2010, 270).

The appeal to abstract entities brings a number of difficulties with it. The first is that the class of abstract objects is rather large. Numbers and other objects of pure mathematics, classes, propositions, concepts, the letter 'A', and Dante's Inferno are abstract objects (Rosen, 2014), and Hale (1988, 86-87) lists no less than 12 different possible characterisations of abstract objects. At the very least this list shows that there is great variety in abstract objects and classifying models as abstract objects adds little specificity to an account of what models are. A defender of the similarity account could follow Giere and counter that they limit their attention to those abstract objects that possess 'all and only the characteristics specified in the principles' (Giere, 2004, 745), where principles are general rules like Newton's laws of motion. Giere further specifies that he takes 'abstract entities to be human constructions' and that 'abstract models are definitely not to be identified with linguistic entities such as words or equations' (2004, 747). While this narrows down the choices somewhat, it still leaves many options and ultimately the ontological status of models in a similarity account remains unclear. Giere fails to expand on this ontological issue for a reason: he dismisses the problem as one that philosophers of science can set aside without loss. He voices scepticism about the view that philosophers of science 'need a deeper understanding of imaginative processes and of the objects produced by these process' (2009, 250), or that 'we need say much more [...] to get on with the job of investigating the functions of models in science' (2009, 250).

I'm not sure that this is quite the case, not least because there is an obvious yet fundamental issue with abstract objects. No matter how the above issues are resolved (and irrespective of whether they are resolved at all), at the minimum it is clear that non-physical models are 'abstract' in the sense that they have no spatiotemporal location. And as Thomson-Jones (2010) points out, this alone causes serious problems for the similarity account. The similarity account demands that models can instantiate properties and relations, since this is a necessary condition on them being similar to their targets. In particular, it requires that models can instantiate the properties and relations mentioned in theoretical hypotheses or commentaries. But such properties and relations are typically physical. And if models have no spatiotemporal location, then they do not instantiate any such properties or relations. Thomson-Jones' example of the idealised pendulum model makes this clear. If the idealised pendulum is abstract then it is difficult to see how to make sense of the idea that it has a length, or a mass, or an oscillation period of any particular time.

An alternative suggestion due to Teller is that we should instead say that whilst 'concrete objects HAVE properties [...] properties are PARTS of models' (2001, 399, original capitalisation). It is not entirely clear what Teller means by this, but my guess is that he would regard models as bundles of properties. Target systems, as concrete objects, are the sorts of things that can instantiate properties delineated by theoretical hypotheses. Models, since they are abstract, cannot. But rather than being objects instantiating properties, a model can be seen as a bundle of properties. A collection of properties is an abstract entity that is the sort of thing that can contain the properties specified by theoretical hypotheses as parts. The similarity relation between models and their targets shifts from the co-instantiation of properties, to the idea that targets instantiate (relevant) properties that are parts of the model. With respect to what it means for a model to be a bundle of properties Teller claims that the '[d]etails will vary with one's account of instantiation, of properties and other abstract objects, and of the way properties enter into models' (2001, 399). But as Thomson-Jones (2010, 294-295) notes, it is not obvious that this suggestion is an improvement on Giere's abstract objects. A bundle view incurs certain metaphysical commitments, chiefly the existence of properties and their abstractness, and a bundle view of objects, concrete or abstract, faces a number of serious problems (Armstrong, 1989).

An alternative approach, that I discuss in more detail in Chapters 7 and 10, is to take models to be, in some sense, fictions. In relation to the similarity view this has been discussed by Giere, who points out that a natural response to Thomson-Jones' problem is to regard models as akin to imaginary or fictional systems of the sort presented in novels and films. It seems true to say that Sherlock is a smoker, despite the fact that Sherlock is an imaginary detective, and smoking is a physical property. At times, Giere seems sympathetic to this view. He says:

> 'it is widely assumed that a work of fiction is a creation of human imagination ... the same is true of scientific models. So, ontologically, scientific models and works of fiction are on a par. They are both imaginary constructs' (2009, 249).

And he observes that:

> 'novels are commonly regarded as works of imagination. That, ontologically, is how we should think of abstract scientific models. They are creations of scientists' imaginations. They have no ontological status beyond that' (2010, 278).

However, these seem to be occasional slips and he has recently positioned himself as an outspoken opponent of any approach to models that likens them to literary fiction (2009).

Furthermore, as Godfrey-Smith (2009) points out, a tension remains when the fictionalist account of models is combined with a similarity view of scientific representation. The latter essentially involved comparing the properties of models (construed as fictional objects) with properties of their targets. But for many models, their relevant properties are ones that we know aren't instantiated anywhere in the actual world. There is no such thing as an infinite population. So the fictionalist needs an ontology of scientific models which accounts for them instantiating uninstantiated (at least in the actual world) properties.

In sum, even if my previous objections to the use of similarity in accounts of scientific representation can be met, since it requires that all models to which it is applicable co-instantiate relevant, which I take (at least at times) to include physical, properties, there is an ontological question regarding how non-physical models can do this. And the similarity view has yet to be equipped with a fully satisfactory answer to this question.[38]

38 Notice that this is a general instance of a more general problem for accounts of scientific representation. As I discussed in Chapter 1, although the semantic and epistemological questions surrounding scientific models don't fully specify a required ontology for non-physical models, they do bear on it. These ontological questions will continue to be discussed throughout this thesis in reference to each answer to the ER-problem, including my own.

5

STRUCTURALISM I

The structuralist conception of scientific representation originated in the so-called 'semantic view of theories' that was developed in the second half of the 20th Century.[39] The semantic view was originally proposed as an account of the nature of scientific theories rather than scientific representation. The driving idea behind the position is that scientific theories are best thought of as collections of models. This invites the questions: what are these models, and how do they represent their target systems? The first question is my problem of ontology, and the second is my ER-problem. As a first pass, defenders of the semantic view of theories take models to be mathematical structures, which represent their target systems in virtue of there being some kind of mathematical mapping (isomorphism, partial isomorphism, homomorphism, ...) between the two. I begin this chapter by briefly outlining what mathematical structures are before moving on to how invoking structure preserving mappings might serve to answer the ER-problem. Given the connections between the structuralist and similarity-based accounts of scientific representation, it should be of no surprise that a naïve appeal to the existence of such model-target mapping, which would amount to them being similar with respect to their structure, suffers problems analogous to those discussed in Chapter 4. As such, many of the moves made there, to, for example, account for the possibility of misrepresentation, arise again here.

39 Suppes (1969a, 2002), van Fraassen (1980), Suppe (1989), and Da Costa and French (2003) provide classical statements of the view. Byerly (1969), Chakravartty (2001), Klein (2013), and Portides (2005, 2010) provide critical discussions. Lutz (2015) compares the semantic view with a syntactic approach. It is worth noting that Giere, whose account of scientific representation I discussed in the previous chapter, is also associated with the semantic view, despite not subscribing to either of the positions discussed in this chapter.

5.1 STRUCTURES, ISOMORPHISMS, AND SCIENTIFIC MODELS

5.1.1 *Structures*

Almost anything, from a football match to an economy, can be referred to as a 'structure'. So the first task for a structuralist account of representation is to articulate what notion of structure it employs. A number of different notions of structure have been discussed in the literature (for a review see Thomson-Jones (2011b)), but by far the most common and widely used is the notion of structure one finds in set theory and mathematical logic. A structure $\mathcal{S}$ (sometimes described as a 'mathematical structure' or 'set-theoretic structure') of the sort found there is a pair consisting of the following elements: a non-empty set $\mathcal{D}$ of objects called the domain (or universe) of the structure and a non-empty indexed set of relations defined extensionally on $\mathcal{D}$. We can thus define a *structure* as follows:

$$\mathcal{S} = \langle \mathcal{D}, \{R_i\}_{i \in I} \rangle$$

where each R_i is a set of n-tuples, i.e elements of $\mathcal{D}^n$ for some n (where the n can vary across different relations depending on their aritary). This definition is close to, but not quite, the definition of structure that is used in logic and model theory. As they are used there, structures contain (or are sometimes identified with) an additional element called an interpretation function $\mathfrak{I}$, whose domain is a set of symbols such that for each n-ary predicate symbol $R^n, \mathfrak{I} : R^n \rightarrow \mathcal{D}^n$ (*cf.* Enderton, 2001; Boolos and Jeffrey, 1989).[40] Each of the values of this function, i.e. as it's applied each predicate symbol, could then be identified with an R_i. From a logical point of view we would define what could be called an 'interpreting structure' as a triple: $\langle \mathcal{D}, \{R_i\}_{i \in I}, \mathfrak{I} \rangle$, where $\mathfrak{I}$ is a function from some a set of symbols to $\{R_i\}_{i \in I}$. But adherents to the semantic approach usually prefer to use the word 'structure' to refer to the set theoretic object that provides the range of $\mathfrak{I}$, i.e. $\mathcal{S} = \langle \mathcal{D}, \{R_i\}_{i \in I} \rangle$.[41] See, for example, Suppes (1969a, 290-291), Lloyd (1994, 15), French (2003, 1480), French (2014, 103), and the references therein, and van Fraassen, who explicitly states that:

> 'logicians do also use "model" to refer to the interpretation itself; that may be convenient, since the structure can then be identified as the range of the

40 Interpretation functions are usually defined over constant and function symbols in the language as well, where $\mathfrak{I}$ maps constant symbols to elements of $\mathcal{D}$, and function symbols to operations on $\mathcal{D}$. I ignore this for my current purposes, since constant symbols do not play any significant role in the structuralist account, and operations can be reduced to relations (by identifying them with their graph).

41 A comparison between these approaches is discussed in Thomson-Jones (2006).

> interpretation. This is 'book-keeping'; the terminology can be adjusted for convenience in various ways' (2014, fn.1).

I shall follow suit here.

5.1.2 *Isomorphisms*

Isomorphisms are structure preserving functions between structures. Intuitively, two structures $\mathcal{S}$ and $\mathcal{S}'$ are isomorphic if they share the same relational structure, despite the fact that their domains might consist of different objects. This notion can be defined as follows:

> Two structures $\mathcal{S} = \langle \mathcal{D}, \{R_i\}_{i\in I} \rangle$ and $\mathcal{S}' = \langle \mathcal{D}', \{R_i'\}_{i\in I} \rangle$ are *isomorphic* if and only if:
>
> (i) there exists a function $f : \mathcal{D} \to \mathcal{D}'$,
>
> (ii) f is a bijection, i.e. one-to-one map from $\mathcal{D}$ onto $\mathcal{D}'$, and
>
> (iii) for each $\langle x_1, ...x_n \rangle \in \mathcal{D}^n, R_i \in \{R_i\}_{i\in I}$, and corresponding $R_i' \in \{R_i'\}_{i\in I}$: $\langle x_1, ...x_n \rangle \in R_i \Leftrightarrow \langle f(x_1), ..., f(x_n) \rangle \in R_i'$.

It bears pointing out, again, that this definition is slightly different from the standard definition used in logic and model theory. There, they take isomorphism to hold between two $\mathcal{L}$-structures, where an $\mathcal{L}$-structure interprets a set of symbols in the language $\mathcal{L}$ (like the 'interpreting structures above). In that context, in order for structures to be isomorphic, they have to interpret the same set of symbols, and the correspondence between the relations R_i and R_i' is given by the fact that they interpret the same predicate symbol (see, e.g. Enderton, 2001, 94). When structures are divorced from the language they interpret, as in the case under consideration here, the way in which the relations in $\mathcal{S}$ are associated with the relations in $\mathcal{S}'$ is being done by their index. This means that by permuting the relations of one structure in its indexed set, structures which were previously isomorphic may turn out to be non-isomorphic. This has the unsatisfactory consequence that isomorphic structures in this sense needn't be elementary equivalent i.e. satisfy the same set of sentences in some language $\mathcal{L}$, since one could permute the interpretation function as well. For the difficulties associated with this in the context of defining theories as sets, or classes, of structures in this way see Glymour (2013), Halvorson (2012, 2013), and Lutz (2015).

5.1.3 *Models*

Structuralists solve the problem of ontology by identifying scientific models with structures in the above sense. Suppes clearly states that 'the meaning of the concept of model is the same in mathematics and the empirical sciences' (1969a, 12). Lloyd argues that 'models should be understood as *structures*; in almost all of the cases I shall be discussing, they are mathematical structures, i.e., a set of mathematical objects standing in certain mathematically representable relations' (1994, 15). Likewise, van Fraassen claims that a '[s]cience represents the empirical phenomena as embeddable in certain abstract structures (theoretical models) [...] Those abstract structures are describable only up to structural isomorphism' (2008, 238). And French and Ladyman claim that 'the specific material of the models is irrelevant; rather it is the structural representation [...] which is important' (1999, 109).[42]

These structuralist accounts have typically been proposed in the semantic view of theories framework. There are differences between them, and formulations vary from author to author. However, as pointed out by Da Costa and French (2000, 119), all these accounts are committed to the idea that scientific models are mathematical structures. As such, I take it that all of these authors provide the following answer to the problem of ontology: models are structures. It remains to give an account of what these structures are ontologically speaking. Are they Platonic entities, equivalence classes, intuitionist constructions, or something else? This is a question for philosophers of logic and mathematics. Various positions are available (see, for example, Hellman (1989, 1996), Dummett (1991), Resnik (1997), and Shapiro (2000)). But philosophers of science need not resolve this issue and can pass off the burden of explanation to philosophers of mathematics. So I don't pursue this matter further here.

An extension of the standard conception of structure is the so-called partial structures approach (see, for instance, Bueno et al. (2002); Da Costa and French (2003); French (2003, 2014)). Relations were defined above by extensionally specifying the n-tuples for which they hold. This naturally allows a sorting of all n-tuples into two classes: ones that belong to the relation and ones that don't. The leading idea of partial structures is to introduce a third option: for some n-tuples it is not determined whether or not they belong to the relation. Such a relation is a partial relation and is defined as follows: let an n-ary *partial relation* be an ordered triple $\mathtt{R}^n = \langle R_1, R_2, R_3 \rangle$, such that $R_1 \cup R_2 \cup R_3 = \mathcal{D}^n$, all of R_1, R_2, and R_3 are mutually disjoint, and where R_1 is set of n-tuples of objects for which the relation holds, R_2 is the set of n-tuples for which it doesn't, and R_3 is the set of n-tuples for which it is not defined whether or

42 See also Suppes (1969b, 24); van Fraassen (1980, 64; 1991, 483; 1995, 6; 2008, 238); and Da Costa and French (1990, 249).

not the relations hold. At the limit, where $R_3 = \varnothing$, partial relations (almost) coincide with fully specified relations over a domain.[43] A *partial structure* is then defined in a way analogous to 'normal' structures discussed above, but with partial relations rather than the more familiar fully specified relations used there:

$$\mathtt{S} = \langle \mathcal{D}, \{\mathtt{R_i}\}_{i \in I} \rangle, \text{ where each } \mathtt{R_i} \text{ is a partial relation.}$$

Proponents of that approach are more guarded regarding the ontology of models. Bueno and French emphasise that 'advocates of the semantic account need not be committed to the ontological claim that models *are* structures' (2011, 890, original emphasis). This claim is motivated by the idea that the task for philosophers of science is to represent scientific theories and models, rather than to reason about them directly. French (2010) makes it explicit that according to his account of the semantic view of theories, a scientific theory is represented as a class of models, but should not be identified with that class. Moreover, a class of structures is just one way of representing a theory; we can also use an intrinsic characterisation and represent the same theory as a set of sentences satisfied by those structures in order to account for how they can be objects of our epistemic attitudes (French and Saatsi, 2006). He therefore adopts a quietist position with respect to what a theory or a model is, declining to answer the question.

There are thus two important notions of representation at play: scientific representation of targets by models, which is the job of scientists; and representation of theories and models by structures, and the representation of scientific representation in terms of structure preserving mappings, which is the job of philosophers of science. The project is a 'rational reconstruction' of scientific practice, and the quietist needn't say anything about what scientific model, or scientific representation, actually is. The question for this approach is whether the philosophical project of representing models and scientific representation as partial structures and morphisms that hold between them is an accurate or useful one. Some rational reconstructions are better than others; if the reconstruction is to work then it better be the case that it gets something right about what it is a reconstruction of. I take it that the concerns raised below remain in this context as well.

There is an additional question regarding the correct formal framework for thinking about models in the structuralist position. Landry (2007) argues that the universal reliance on set theory to define structures and morphisms is misguided, claiming that the

43 I say 'almost' because, technically speaking, such a partial relation is a pair of sets of n-tuples, whereas usual relations are identified with single sets of n-tuples. In this sense, fully specified partial relations contain slightly more information than usual relations, because they encode the entire domain they are defined over.

context determines which formal framework is appropriate to work in. For example, in the context of quantum mechanics, there are reasons to utilise group theory rather than set theory. In other contexts category theory may be the appropriate framework. Similarly, Halvorson (2012, 2016) claims that an account of scientific theories that treats theories as classes of models fails to account for important intra and inter theoretical relationships. He proposes that theories should thus be identified with categories of models instead, where the (category theoretic) morphisms between the models identify intra-theory relationships, and functors between categories identify inter-theory relationships.

Although these concerns highlight important questions regarding the nature of scientific theories, it is less obvious whether they have any significant bearing on how models represent their target systems. Even if models are construed non-set theoretically, an account of how they enter into morphisms with their target systems still needs to be given, and it's plausible that such an account will face difficulties analogous to the ones I discuss below. Landry's paper is not an attempt to re-frame the representational relationship between models and their targets (see Brading and Landry (2006) for her scepticism regarding how structuralism deals with this question). And Halvorson's category theoretic approach still takes individual models to be set theoretic structures; he just takes them to be embedded in categories rather than classes of such structures. So the question of how an individual model represents its target system is unchanged. Thus, for reasons of simplicity I focus on the structuralist view that identifies models with set-theoretic structures throughout the rest of this chapter.

5.2 NAÏVE STRUCTURALISM

So, with the appropriate formal notions thus introduced, and the structuralist answer to the problem of ontology offered, it's now time to turn to how these come to bear in answering the ER-problem. In its simplest guise, the structuralist answer to the ER-problem is the following:

STRUCTURALISM 1 A scientific model M represents a target system T if and only if M and T are isomorphic to one another.

This view is articulated explicitly by Ubbink, who claims that 'a model represents an object or matter of fact in virtue of this structure; so an object is a model [...] of matters of fact if, and only if, their structures are isomorphic' (1960, 302). Views similar to Ubbink's seem to operate in many versions of the semantic view. However, in fairness to proponents of structuralist accounts, it ought to be pointed out that for a long time representation was not the focus of attention and the attribution of (something like) STRUCTURALISM 1 to the semantic view is an extrapolation (an extrapolation that, for instance, Frigg (2006) and Suárez (2003), make). Representation became a much-debated topic in the first decade of the 21st century, and many proponents of the semantic view then either moved away from STRUCTURALISM 1, or pointed out that they never held such a view. I turn to more advanced positions shortly, but to understand what motivates such positions it is helpful to understand why STRUCTURALISM 1 fails.

This account has two *prima facie* advantages. Firstly, it offers a straightforward answer to the ER-problem which allows for surrogative reasoning: the mappings between the model and the target allow scientists to convert truths found in the model into claims about the target system. The second advantage concerns the applicability of mathematics. Mathematical structuralists take mathematics to be the study of structures (Shapiro, 2000). It is a natural move for the scientific structuralist to follow suit, which provides them with a clear explanation of how mathematics is used in scientific modelling: mathematics is applicable to the natural world because the subject matter of mathematics (mathematical structures) are isomorphic to systems found in the world.

But, like SIMILARITY 1 discussed in the previous chapter, STRUCTURALISM 1 suffers from serious flaws. The first concern, again, concerns the logical properties of scientific representation and the relation used to ground it, in this case isomorphism. Isomorphisms are reflexive, symmetric, and transitive. Since scientific representation has none of these logical properties, STRUCTURALISM 1 fails to capture the properties of scientific representation.

A second concern is how we can make sense of the idea that target systems can enter into isomorphisms. The notion is only defined on *pairs* of *structures*. And, at least *prima facie* target systems aren't structures; they are physical objects like planets, molecules, populations of organisms, and economies. The relation between structures, abstract mathematical objects, and physical targets is a serious question. Are we supposed to take physical objects *to be* abstract structures? To *instantiate* abstract structures? Or something else entirely? I devote the next chapter to investigating how the structuralist can answer this question.

Relatedly, although the account has the resources to successfully deal with mathematical models, it remains unclear how to apply it to models that aren't like this: concrete models like the Phillips-Newlyn machine, or scale models of ships in water tanks. Suppes (1969a) claims that these models can be ascribed a structure in some sense or another. But, like the question of target-end structure, how physical systems can be said to 'be', or 'have', structures remains at this point rather unclear. For the purposes of this chapter I grant the structuralist the assumption that physical systems, targets and concrete models, are (or at least have) structures.

Thirdly, like similarity, isomorphism alone is insufficient for scientific representation. Not all structures that are isomorphic to one another represent each other. In the case of similarity this case was brought home by Putnam's thought experiment with the ant crawling on the beach; in the case of isomorphism a look at the history of science will do the job. Furthermore, as Frigg (2006, 10) notes, many mathematical structures were discovered and discussed long before they were used in science. Non-Euclidean geometries were studied by mathematicians before Einstein used them in the context of spacetime theories, and Hilbert spaces were studied by mathematicians prior to their use in quantum theory. If representation were nothing more than isomorphism, then it would be the case that Riemann discovered general relativity and that Hilbert invented quantum mechanics. This jars with our intuitions, which provides one reason to doubt that isomorphism on its own does not establish representation.

Another concern is how STUCTURALISM 1 deals with a version of the 'mistargeting' objection discussed previously. Granted, isomorphism is more restrictive than similarity: not everything is isomorphic to everything else. But isomorphism is still too easy to come by to correctly identify the class of systems a structure represents. The root of the difficulty is that the same structures can be found in different target systems. As Frigg (2003, 33-35) points out, the $F = G\frac{m_1 \times m_2}{r^2}$ law of Newtonian gravity is the mathematical skeleton of Coulomb's law of electrostatic attraction, $F = k\frac{q_1 \times q_2}{r^2}$. The mathematical structure of the pendulum is also the structure of an electric circuit with a condenser and solenoid (Kroes, 1989). Linear equations, partial differential equations, and probability spaces and yet more kinds of mathematical objects, are all ubiquitous in physics, biology, economics, and other scientific fields. And, certain geometrical structures are found in many different systems; consider how many spherical things there are. This shows that the same structure can be exhibited (in some sense) by more than one kind of target system. Structures are 'multiply realisable' (Frigg, 2003, 33). If representation is explicated solely in terms of isomorphism, then we have to conclude that, say, a model of a celestial orbit is also a model of a model of the electrostatic interaction between two charged particles. But this seems wrong. Hence, it is difficult to see how the notion of being isomorphic to a model M can be used to delineate the systems that M represents.[44]

44 Van Fraassen (1980, 66), mentions a similar problem when he discusses 'unintended realisations' and attempts to avoid it by claiming that it will 'disappear when we look at larger observable parts of the world'. Even if there are multiply realisable structures to begin with, as science progresses we will realise that such targets have different structural features. Once we focus on a sufficiently large part of the world in enough detail, no two targets will have the same structure. But relying on future science to explain

Fourthly, as we have seen in the last section, a misrepresentation is one that represents its target as having features it doesn't have. In the case of an isomorphism account of representation this presumably means that the model represents the target as having structural properties that it doesn't have. However, isomorphism demands identity of structure: the structural properties of the model and the target must correspond to one another exactly. Any model that misrepresents the structure of its target (of which there are many; I don't think anyone believes that any model perfectly represents the structure of its target system) will not be isomorphic to the target. By the lights of STRUCTURALISM 1 it is therefore not a representation at all. Like simple similarity accounts, STRUCTURALISM 1 conflates misrepresentation with non-representation.

And fifthly, like the similarity account, STRUCTURALISM 1 has a problem with non-existent targets because no model can be isomorphic to something that doesn't exist. If there is no ether, a model can't be isomorphic to it. Hence models without targets are not dealt with by STRUCTURALISM 1.

So, to summarise: STRUCTURALISM 1 (i) prescribes the wrong logical properties to scientific representation; (ii) needs to provide an account of where the target end-structures come from; (iii) doesn't provide sufficient conditions on scientific representation, (a) since it ignores both the transition from when structures were studied in mathematics to when they came to be used in the sciences and (b) the fact that multiple target systems exhibit the same structure; (iv) conflates (structural) misrepresentation with non-representation; and (v) has nothing to say about non-existing targets.

5.2.1 *Choosing a morphism*

From the above objections, it could be argued that at least (i) and (iv) can be met by weakening STRUCTURALISM 1, whilst keeping the spirit of the account. The central idea is to replace isomorphism in the definition of scientific representation with a weaker structure preserving mapping. There are two (non-exclusive) ways of going about this. The first is to move to the partial structures framework outlined above, and define an isomorphism in that context. The second is to look at weaker mappings than isomorphism. The formal details of these approaches are spelt out first, and then I investigate the resulting answers to the ER-problem.

If the structures in question are taken to be partial structures, then the notion of a partial isomorphism can be defined as follows. Let $\mathtt{S} = \langle \mathcal{D}, \{\mathtt{R_i}\}_{i \in I} \rangle$ and $\mathtt{S}' =$

---

how models work today seems unconvincing. It is a matter of fact that we currently have models that represent gravitational attraction and electrostatic interaction, and waiting for future science to provide us with more detailed accounts of such target systems seems irrelevant with respect to what our current models actually represent (Frigg, 2003, 34-35).

$\langle \mathcal{D}', \{\mathtt{R}'_\mathtt{i}\}_{i\in I}\rangle$ be two partial structures. Then f is a *partial isomorphism* from $\mathcal{D}$ to $\mathcal{D}'$ if and only if

(i) f is a bijection, i.e. one-to-one map from $\mathcal{D}$ onto $\mathcal{D}'$, and

(i) For all $\langle x_1, ..., x_n\rangle \in \mathcal{D}$ and for each $\mathtt{R_i} = \langle R_1, R_2, R_3\rangle \in \{\mathtt{R_i}\}_{i\in I}$ and corresponding $\mathtt{R}'_\mathtt{i} = \langle R'_1, R'_2, R'_3\rangle \in \{\mathtt{R}'_\mathtt{i}\}_{i\in I} : \langle x_1, ..., x_n\rangle \in R_1 \Leftrightarrow \langle f(x_1), ..., f(x_n)\rangle \in R'_1$ and $\langle x_1, ..., x_n\rangle \in R_2 \Leftrightarrow \langle f(x_1), ..., f(x_n)\rangle \in R'_2$.

Other relevant mappings include homomorphisms (Bartels, 2006; Lloyd, 1994; Mundy, 1986), isomorphic embeddings (van Fraassen, 1980, 1997, 2008; Redhead, 2001), and Δ/Ψ-morphisms (Swoyer, 1991). It's worth noting that there is some confusion in the literature about how to define homomorphism (see Pero and Suárez (2015) for a useful discussion), so I offer multiple definitions to clarify how they differ.[45] Since Swoyer's Δ/Ψ-morphisms require further elaboration on his non-standard use of 'structure', and it doesn't solve the problems associated with the below mappings, I do not discuss it here.

f is a *homomorphism* from $\mathcal{S} = \langle \mathcal{D}, \{R_i\}_{i\in I}\rangle$ to $\mathcal{S}' = \langle \mathcal{D}', \{R'_i\}_{i\in I}\rangle$ if and only if for all $\langle x_1, ..., x_n\rangle \in \mathcal{D}^n$ and each $R_i \in \{R_i\}_{i\in I}$ and corresponding $R'_i \in \{R'_i\}_{i\in I} : \langle x_1, ...x_n\rangle \in R_i \Rightarrow \langle f(x_1), ..., f(x_n)\rangle \in R'_i$.

f is a *surjective homomorphism* if it is a homomorphism that is surjective (i.e. onto D').

f is a *faithful homomorphism* from $\mathcal{S} = \langle \mathcal{D}, \{R_i\}_{i\in I}\rangle$ to $\mathcal{S}' = \langle \mathcal{D}', \{R'_i\}_{i\in I}\rangle$ if and only if:

(i) For all $\langle x_1, ..., x_n\rangle \in \mathcal{D}^n$ and each $R_i \in \{R_i\}_{i\in I}$ and corresponding $R'_i \in \{R'_i\}_{i\in I}$: $\langle x_1, ...x_n\rangle \in R_i \Rightarrow \langle f(x_1), ..., f(x_n)\rangle \in R'_i$, and

(ii) For all $\langle f(x_1), ..., f(x_n)\rangle \in \mathcal{D}'^n$ and each $R'_i \in \{R'_i\}_{i\in I}$ and corresponding $R_i \in \{R_i\}_{i\in I} : \langle f(x_1), ..., f(x_n)\rangle \in R'_i \Rightarrow \langle x_1, ..., x_n\rangle \in R_i$.

A structure $\mathcal{S} = \langle \mathcal{D}, \{R_i\}_{i\in I}\rangle$ can be *isomophically embedded* in a structure $\mathcal{S}' = \langle \mathcal{D}', \{R'_i\}_{i\in I}\rangle$ if and only if:

(ii) For all $\langle x_1, ..., x_n\rangle \in \mathcal{D}^n$ and each $R_i \in \{R_i\}_{i\in I}$ and corresponding $R_i \in \{R'_i\}_{i\in I}$: $\langle x_1, ...x_n\rangle \in R_i \Leftrightarrow \langle f(x_1), ..., f(x_n)\rangle \in R'_i$, and

(iii) f is an injection, i.e. one-to-one map from $\mathcal{D}$ to (but not necessarily onto) $\mathcal{D}'$.

45 In all of the below, that $f : \mathcal{D} \to \mathcal{D}'$ is left as implicit.

Hopefully the pertinent differences between these mappings is clear. An isomorphism demands that the two structures have the same structure in the sense that each relation is fully preserved in both directions. A homomorphism from $\mathcal{S} \rightarrow \mathcal{S}'$ simply demands that the relations in $\mathcal{S}$ are preserved in the image of the homomorphism. A surjective homomorphism requires that f map to every element in the domain of $\mathcal{S}'$ (but as I use it allows for there to be relations in $\mathcal{S}'$ that are not preserved). A faithful homomorphism from $\mathcal{S} \rightarrow \mathcal{S}'$ requires that the relations are preserved in both structures, but differs from an isomorphism in not requiring that the function be a bijection, and thus $\mathcal{S}'$'s preserved relations are restricted to the image of the faithful homomorphism. An isomorphic embedding of $\mathcal{S}$ into $\mathcal{S}'$ is an injection that preserves both sets of relations, where $\mathcal{S}'$'s are restricted to the image of the embedding. Each of these have straightforward analogs in the partial structures framework. When I need to, I use the term 'morphism' to remain neutral between these mappings.

If we replace the term 'isomorphism' in Structuralism 1 with any of the above mappings, we can consider how the resulting account fares with respect to the issues concerning the logical properties of scientific representation, and whether it can handle the problem of misrepresentation.

With respect to the logical properties, weaker mappings do fare better than isomorphism. None of the above mappings are symmetric, which captures the idea that in general, models represent their target systems but not the other way around. However, all of them are reflexive, and some of them remain transitive. So the weaker mappings alone don't suffice to establish the logical properties we expect of scientific representation.

The problem of misrepresentation is more subtle. To start with the partial structures approach, Bueno and French claim that:

> 'With the introduction of partial isomorphism and homomorphism, no requirement is made that the structures that are used to represent other structures do so with perfect accuracy' (2011, 888).

For sure, the claim that there is a partial isomorphism f from a scientific model M to a target system T allows for some degree of misrepresentation. For instance, it allows that there are objects, or n-tuples of objects, that are not known to be in the extension of a relation in M (i.e. are in R_3 for some R_i) that, as a fact of the matter either are, or are not, in the extension of the corresponding relation in T. And if partial homomorphism is used instead, then it allows that there are relations in the target that don't correspond to any relation in the model. The first of these cases corresponds to cases where we are ignorant about whether or not target objects are related by various relations, and the second to cases where models are abstractions of their targets: models that represent only some of the relations in the target system. In this sense it is true that the partial structures framework has a degree of flexibility that the standard view does not.

But, despite Bueno and French's claims to the contrary, it remains unclear whether it stretches far enough. The pertinent cases of misrepresentation I have in mind are where scientific models are distortive representations; in the terminology of the partial structures approach, where the model represents certain objects as being in the extension of relations that *they are explicitly not in the extension of*. These distortions can come about through deliberate misrepresentation – cases of representation which from our current perspectives are known to be radically inaccurate in certain respects – or from the fact that our models, despite our best efforts, are still mistaken about the structure of their target systems. With respect to the former kind of misrepresentation, consider for example the case introduced in Chapter 1 of modelling a sand box as a two dimensional lattice, sand grains as integers associated with cells, and the toppling rule defined in terms of grains of sand moving to their neighbour cells. In this case, presumably, the partial structure will include a function on time that governs how sand grains in the model move. This function will, again presumably, correspond to a 'function' in the sandbox that governs the actual dynamic behaviour of sand grains. But actual sand grains don't instantly 'jump' in axial directions, and then stay still (assuming their action doesn't induce another toppling event), they exhibit inertia: it takes some time for them to accelerate, and some time for them to come to rest. So if the partial homomorphism maps a 'sand grain' in the model to a sand gain in the sandbox, then the function in the model will ascribe the sand grain a location which explicitly doesn't correspond to the actual location of the sand grain. But this sort of distortion isn't captured by the introduction of partial relations in any obvious way. With respect to the latter sort of misrepresentation; it's difficult to imagine what sort of partial homomorphism could be set up between, for example, Thomson's plum pudding model of hydrogen atoms (which presumably would include a relation between locations and charges) and actual hydrogen atoms (which are misrepresented by the model in the sense that the positive charge is concentrated in the centre of the atom). The plum pudding model explicitly relates locations away from the centre, and locations at the centre, of the model with uniform positive electric charge. As a fact of the matter, this is not so in the case of the actual hydrogen atom. Again, we have an instance where something in the model is in the extension of a relation in the model, and what it corresponds to is explicitly not in the corresponding relation in the world. This precludes a partial homomorphism. So it is at best unclear how partial homomorphisms, let alone partial isomorphisms, can account for these kinds of distortive representations.

The other morphisms outlined above fare no better with respect to these sorts of representations either. The most plausible approach, that claims a model M scientifically represents a target system T if and only if there is a homomorphism from M to T, allows for there to be relations in the target that are not represented by the model, but it still demands that every relation in the model accurately represent a relation in the target. A surjective homomorphism does worse, since it requires that every object in the target is represented: it precludes abstraction on the level of objects. A faithful homomorphism does even worse, since it requires that every relation in the target be accurately represented, at least with respect to their restrictions to the image of the homomorphism. So, again, a weaker mapping does not seem to have the resources to deal with misrepresentation, at least of certain kinds.[46]

Isomorphic embeddings are motivated by a slightly different worry. van Fraassen (1980) uses them to cash out his structuralist constructive empiricism, i.e. the idea that we should be committed to what scientific models tell us about the observable world, and remain agnostic with respect to what they say about the unobservable world. An isomorphic embedding then, is an embedding of what he calls 'the appearances' into a theoretical model. The appearances are taken to be the structure of the observable world (more on this in the next chapter), and the embedding ensures that a theoretical model represents the appearances. Again, this approach cannot deal with misrepresentation of the structure of observable phenomena. Consider, for example, how the population of predator and prey fish in the Adriatic sea change through time. This is, presumably, something we can observe. But this would mean that no mathematical structure that failed to accurately represent the way that these populations change through time (i.e. any mathematical structure that does not not embed the appearances), could represent the fish. This doesn't seem right: not all population growth models are accurate representations. I suspect for this reason, van Fraassen (2008), as discussed in the next chapter, moves on from using a structure preserving mapping to answering the ER-problem, to using it as providing the standards of accuracy instead.

Another option, as suggested by Muller is to give up on the idea that there is a specific morphism that relates all representing models with their targets, and instead choose a 'tailor-made morphism on a case by case basis' (2011, 112). Muller is explicit that this suggestion presupposes that there is 'at least one resemblance' (2011, 112) between model and target because 'otherwise one would never be called a representation of the other' (2011, 112). While this may work in some cases, it is not a general solution. Models that are gross distortions of their targets (such as the sandpile model or the logistic model of a population) may not enter into any structure preserving mapping

46 Pero and Suárez (2015) provide a useful discussion where they formally capture the sorts of misrepresentation where each of these mappings fail.

with their target systems. More generally, as Muller admits, his solution 'precludes total misrepresentation' (2011, 112). So in effect it just limits the view that representation coincides with correct representation of at least part of the target by at least part of the model.[47] However, this is too restrictive a view of representation. I take it that total misrepresentations may be useless, but they are representations nevertheless.

So although replacing isomorphism with a weaker mapping in STRUCTURALISM 1 provides an answer to the ER-problem which fares better than the purist account, it still cannot fully account for the logical properties of scientific representation, nor all instances of misrepresentation. Moreover, these were just two of five objections I provided at the end of the previous section, and the remaining three objections have not be addressed. For these reasons I don't think that a structure preserving mapping, of any kind, alone will suffice to ground scientific representation. Like naïve similarity, the account needs additional resources if it is to meet these objections, and, as there, it would seem that giving up on the idea that representation can be naturalised will provide, at least part of the story.

5.3 NUANCED STRUCTURALISM

Most of these problems can be resolved by making the same move that lead to SIMILARITY 3: introduce agents and proposed theoretical hypotheses into the account of representation. This delivers the following:

STRUCTURALISM 2 : A scientific model M represents a target system T if and only if a model user provides a theoretical hypothesis H specifying that M and T are appropriately morphic to one another.[48]

Something similar to this was suggested by Adams (1959, 259) who appeals to the idea that physical systems are the *intended* models of a theory in order to differentiate them from purely mathematical models of a theory. This suggestion is also in line with van Fraassen's recent pronouncements on representation. He offers the following as his 'Hauptstatz' of a theory of representation: '*There is no representation except in the sense that some things are used, made, or taken, to represent things as thus and so*' (2008, 94, original emphasis). Likewise, Bueno submits that 'representation is an *intentional act* relating two objects' (2010, 94, original emphasis), and Bueno and French point

47 I remain deliberately vague by what I mean by 'part' here, because this could be cashed out in multiple ways in the structuralist context, for instance by taking restrictions of the model to subsets of its domain, or restrictions of the model with respect to some of its relations, and so on. Muller's suggestion of choosing 'tailor-made' morphisms would be compatible with each of the aforementioned suggestions.

48 STRUCTURALISM 2 is really a family of positions depending on the specific morphism invoked.

out that using one thing to represent another thing is not only a function of (partial) isomorphism but also depends on pragmatic factors 'having to do with the use to which we put the relevant models' (2011, 885).[49] This, of course, reliquishes the idea of an account which reduces representation to intrinsic features of models and their targets. At least one extra element, the model user, also features in whatever relation is supposed to constitute the representational relationship between M and T. In a world with no agents, there would be no scientific representation.

This seems to be the right move. Like Similarity 3, Structuralism 2 accounts for the logical properties of representation, and as long as the truth of the theoretical hypothesis isn't baked into the conditions on representation, then it has no problem with misrepresentation. Requiring that a model user propose that a model is appropriately morphic to a target system also accounts for the fact that many mathematical structures were studied in their own right by mathematicians before being used to represent any target system and allows us to differentiate between models that are used to represent different target systems with 'the same' structure. In fact, with the exception of the problem of non-existing targets, it solves all of the issues I highlighted at the end of section 5.2. But, again as in the case of Similarity 3, this is Pyrrhic victory as the role of the morphism has shifted. The crucial ingredient is the agent's intention and a morphism has in fact become either a representational style or normative criterion for accurate representation (or both). Let's see how they fare in response to those problems.

5.3.1 *Structures, style, accuracy, and demarcation*

The problem of style is to identify representational styles and characterise them. Isomorphism offers one style of representation: one can represent a system by coming up with a model and proposing that the two are appropriately morphic to one another. This style also offers a clear-cut standard of accuracy: the representation is accurate if the hypothesised isomorphism holds; it is inaccurate if it doesn't.

This is a neat answer. The question is what status it has vis-à-vis the problem of style. Is the isomorphism style merely one style among many other styles which are yet to be identified, or is it in some sense privileged? Just as I granted in the previous section that intended similarities might provide one style, amongst others, I can grant that intended isomorphisms might provide one style, amongst others. However, that

49 However, for reasons that I find unclear, they remain sceptical that the intentions of model users be directly written into the conditions on scientific representation (*cf.* French, 2003).

fact that structuralists emphasise isomorphisms, and other morphisms, to such an extent, it seems that they think that they cover *all* styles of scientific representation.

This claim seems to conflict with scientific practice. Many representations are distortions, and known to be. Theoretical hypotheses that specify that models have to be isomorphic to their targets in order to be accurate miss out ways in which models can still allow successful surrogative reasoning about their targets, despite not having exactly the same structure as them. Some models in statistical mechanics have an infinite number of particles and the Lotka-Volterra model represents ecological systems without unlimited food for the prey, and where predators and prey do not reproduce continuously. These models can be used fruitfully, at least in certain contexts, and for certain purposes, despite the fact that they are not isomorphic to their target systems.

This raises the question whether other mappings such as homomorphisms, or embeddings would fare any better. They would, I think, provide multiple different styles of scientific representation, but they would not fill all of the gaps. Again, although they deal well with models that are, in some sense, abstractions of their targets, either by allowing scientific representations to reflect our ignorance as to whether certain objects are in the extension of certain relations (partial isomorphisms), or allow our models to be accurate despite the fact that they fail to represent all relations on the target domain (homomorphisms), they still cannot account for the fact that models which are distortions can nevertheless be accurate representations of their target systems. From a morphism perspective all one can say about such idealisations is that they are failed morphism representations (or morphism misrepresentations). This is rather uninformative. One might try to characterise these idealisations by looking at how they fail to be appropriately morphic to their targets, but I doubt that this is going to get us very far. Understanding how distortive idealisations work requires a positive characterisation of them, and I cannot see how such a characterisation could be given within the morphism framework. So one has to recognise styles of representation other than the proposed morphisms.

Structuralism's stand on the demarcation problem is by and large an open question. Unlike similarity, which has been widely discussed across different domains, structuralism is tied closely to the formal framework of set theory, and it has been discussed only sparingly outside the context of the mathematised sciences. An exception is French (2003), who discusses structuralist accounts in the context of pictorial representation. He discusses in detail Budd's (1993) account of pictorial representation and points out that it is based on the notion of a structural isomorphism between the structure of the surface of the painting and the structure of the relevant visual field. Therefore representation is the perceived isomorphism of structure (French, 2003, 1475-

1476).[50] In a similar vein, Bueno claims that the partial structures approach offers a framework in which different representations – among them 'outputs of various instruments, micrographs, templates, diagrams, and a variety of other items' (2010, 94) – can be accommodated. This would suggest an isomorphism account of representation at least has a claim to being a universal account covering representations across different domains.

This approach faces a number of questions. First, neither a visual field nor a painting is a structure, and the notion of there being an isomorphism in the set theoretic sense between the two at the very least needs unpacking. The theory is committed to the claim that paintings and visual fields have structures, but, as I discuss in the next subsection, this claim faces serious issues. Second, Budd's theory is only one among many theories of pictorial representation, and most alternatives do not invoke isomorphism. So there is a question whether a universal claim can be built on Budd's theory. In fact, there is even a question about isomorphism's universality within scientific representation. Non-mathematised sciences work with models that aren't structures. Godfrey-Smith (2006), for instance, argues that models in many parts of biology are imagined concrete objects. There is a question whether isomorphism can explain how models of that kind represent.

This points to a larger issue. The structuralist view is a rational reconstruction of scientific modelling, and as such it has some distance from the actual practice. Some philosophers have worried that this distance is too large and that the view is too far removed from the actual practice of science to be able to capture what matters to the practice of modelling (this is the thrust of many contributions to Morgan and Morrison (1999); see also Cartwright (1999)). Although some models used by scientists may be best thought of as set theoretic structures, and the way that they represent their targets might be grounded in a proposed morphism, there are many where this seems to contradict how scientists actually talk about, and reason with, their models. Obvious examples include physical models like the Phillips-Newlyn machine, and the the San Francisco Bay model (Weisberg, 2013), but also systems such as the idealised pendulum or imaginary populations of interbreeding animals. Such models have the strange property of being concrete-if-real and scientists talk about them as if they were real systems, despite the fact that they are obviously not. (Thomson-Jones, 2010, 295-298) dubs this 'face value practice', and provides a useful discussion of how unnatural the structuralist position looks when applied to such models. The question remains how to apply the structuralist framework beyond purely mathematical models.

50 This point is reaffirmed by Bueno and French (2011, 864-865); see Downes (2009, 423-425) for a critical discussion

So, to sum up. STRUCTURALISM 2 fares much better than STRUCTURALISM 1 as an answer to the ER-problem. But it is no longer a morphism between models and their targets that answers this problem. Instead, it is a *proposed* morphism, by means of a theoretical hypothesis. Morphisms themselves now provide the standards of accuracy, or a style, instead. Even if the stylistic concerns raised in this section, along with the idea that distortive models can still be accurate representations of their target systems, at least in some sense, can be addressed, there remains an additional question for any structuralist account. Target systems are physical objects: atoms, planets, populations of rabbits, economic agents, etc. By definition, all of the mappings suggested – isomorphism, partial isomorphism, homomorphism, or isomorphic embedding – only hold between two structures. Claiming that a set theoretic structure is isomorphic to a part of the physical world is *prima facie* a category mistake. In order to make sense of the idea that a model is isomorphic (or appropriately morphic) to its target, we have to assume that the latter somehow exhibits a certain structure. But what does it mean for a target system – a part of the physical world – to exhibit a structure, and where in the target system is the structure located? That is the focus of the next chapter.

6

STRUCTURALISM II

As far as I can see, there are two places where the structuralist might look for structures at the target-end of their morphisms. The first is in data. Data can be seen as mathematical objects, and the sorts of things that can enter into morphisms. The structuralist is at liberty then, to answer the ER-problem by invoking a theoretical hypothesis specifying a morphism between a scientific model and data. If STRUCTURALISM 2 is correct, this would then establish that a scientific model represents data. If the morphism holds, then we can grant that the representation is accurate. Alternatively, the structuralist could try to locate the target-end structure 'in' the physical target system itself, either by appealing to the idea that parts of the world can instantiate structures, or even are, at base, structures themselves. This would allow physical systems to enter into morphisms. I discuss both of these strategies in order, but it's worth noting that structuralists themselves disagree with respect to which option to take. Van Fraassen himself describes the suggestion that physical systems can enter into isomorphisms as a ' "dormative virtue" response [which is not] only [...] merely verbal, but [...] also hijacks a term from mathematics for unwarranted use elsewhere' (2010, 549).

6.1 DATA AND PHENOMENA

What are data? Data are what are gathered in experiments. When observing the motion of a planet across the sky, for instance, we choose a coordinate system and at consecutive instants of time observe the position of the planet in this coordinate system. These observations are written down, and data thus gathered are the 'raw data'. The raw data then is then cleansed and regimented: anomalous data points are rejected; measurement errors taken into account; scientists take averages; and usually idealise the data, for instance by replacing discrete data points by a continuous function. Often, but not always, the result is a smooth curve through the data points that satisfies certain theoretical desiderata.[51] I call the result of this process a 'data model'.[52] These resulting data models can be treated as set theoretic structures. In many cases the data points are numeric and the data model is a smooth curve through these points. Such a curve is a relation over $\mathbb{R}^n$ (for some n), or subsets thereof, and hence it is a structure in the requisite sense.[53]

Suppes (1969b) was the first to suggest that data models are the targets of scientific models: models don't represent parts of the world; they represent data structures. This approach has then been adopted by van Fraassen, when he declares that '[t]he whole point of having theoretical models is that they should fit the phenomena, that is, fit the models of data' (1981, 667). He has defended this position on multiple occasions (1980, 64; 1985, 271; 1989, 229; 1997, 524; 2002, 164; 2008, Chapter 7). So models don't represent planets, atoms, or populations; they represent data that are gathered when performing measurements on planets, atoms, or populations.

This revisionary point of view has met with stiff resistance. Muller articulates the unease about this position as follows:

> 'The best one could say is that a data structure $\mathfrak{D}$ seems to act as simulacrum of the concrete actual being **B** [...] But this is not good enough. We don't want simulacra. We want the real thing. Come on' (2011, 98).

Muller's point is that science aims to represent real systems in the world and not data structures. Van Fraassen calls this the 'loss of reality objection' (2008, 258). He points out that 'phenomena are actual objects, events, and processes, while [data models] are the products of our independent intellectual activity' (2008, 259), and accepts that the structuralist must ensure that models represent target systems, rather than finishing the story at the level of data. He addresses this issue in detail and offers a solution which I turn to below. But before going on, it's worth clarifying the objection in a little more detail, and to this end I briefly revisit the discussion about phenomena and data which took place in the 1980s and 1990s.

51 Harris (2003) and van Fraassen (2008, 166-168) elaborate on this process.

52 Van Fraassen's discussion throws up a terminological issue that needs regimenting to avoid confusion. Throughout this chapter I use 'data model' rather than van Fraassen's 'surface model' to refer to the end result of the cleaning and idealising process. I also use 'scientific model' in place of van Fraassen's 'theoretical model'.

53 The example of numerical data is illustrative. As van Fraassen notes, the process of creating data models is not restricted to 'number assigning', and the resulting structures do not have to have $\mathbb{R}$ as their domain. For example, a measurement procedure may only provide an ordinal ranking, and therefore deliver a different kind of structure (van Fraassen, 2008, 158-160). This has no bearing on the discussion in this section.

Bogen and Woodward (1988) and Woodward (1989) introduced the distinction between phenomena and data. The difference is best understood by thinking about an example: the discovery of weak neutral currents (Bogen and Woodward, 1988, 315-318).[54] The model represents particles, neutrinos, nucleons, and the Z^0 particle, and their interactions. The data in this example were 290,000 photographs produced at the bubble chamber in CERN, and only 100 provided evidence for the existence of weak neutral currents. The question then is whether the model represents particles and their interactions, or whether it represents these photographs (or a data model extracted from them). I think the following observation should persuade us (if it's even required) that the former is the case.

As Bogen and Woodward point out, whilst scientists in CERN were taking pictures of interactions in bubble chambers, scientists at the NAL in Chicago were also performing an experiment to detect weak neutral currents. But their research involved an entirely different experimental set-up. Rather than bubble chamber photographs, the NAL experiment recorded patterns of discharge across electronic particle detectors. This meant that the sorts of data gathered at CERN were completely different to the sorts of data gathered at NAL. So different, in fact, that it's difficult to see how the model that represents particles interacting with in each in a particular manner could be appropriately morphic to both of them. And yet both experiments provided evidence for the existence of weak neutral currents, and so both were taken to provide evidence for the same scientific model. As such, I take it that the model, and indeed models more generally, represent phenomena not data.[55]

If data are not the ultimate targets of scientific models, then what role do they play in scientific modelling? Following Frigg (2003, 74), I take it that they perform an evidential function. Data provide *evidence* for the presence, and features, of phenomena that are in turn represented by scientific models (*cf.* Bogen and Woodward, 1988, 305). A certain pattern in a bubble chamber photograph, or a certain pattern of recorded electrical discharge, is evidence for the existence of weak neutral currents. How data provide this evidential support is a difficult question that goes beyond the scope of this thesis. But I want to stress that I'm not claiming that data are irrelevant to the practice of scientific modelling. I'm just pointing out that it's phenomena, not data, that are the ultimate targets of scientific models.[56]

54 In the following discussion of this example I draw heavily upon Frigg (2003, 71-72).

55 What 'phenomena' are, ontologically speaking is an additional question. I take it that they are a diverse collection of objects, states of affairs, processes, mechanisms, amongst other kinds of things (*cf.* Bogen and Woodward, 1988, 321). But this doesn't matter for my current purposes. All that matters is that they are distinct from data, and that they are not mathematical structures.

56 Notice that this does not even preclude data being represented. All I am claiming is that somewhere in this whole story there has to be an account of how scientific models represent physical phenomena.

Those who think that data models are the ultimate targets of scientific models have three options available to them in light of this. Firstly, they can appeal to radical empiricism. By taking phenomena to exist independently of the data we gather from them we have left the realm of observable entities. I doubt that this response will help much. Firstly, note that it even rules out representing 'observable phenomena'. To use van Fraassen's (2008, 254-260) example, which is discussed in more detail below, we can consider the case of a population biologist representing the growth of a deer population through time. If all that is represented is a data model, then strictly speaking, what is represented are data, rather than the deer population itself. Traditionally, empiricists would readily accept that deer, and the rates at which they reproduce, are observable phenomena. Denying that they are represented, by replacing them with data models, seems to be an implausible move. Secondly, as Frigg (2003, 75) points out, regardless of whether phenomena should be understood realistically (Bogen and Woodward, 1988) or anti-realistically (McAllister, 1997), it is still phenomena, not data, that models represent. By suspending belief in the existence of weak neutral currents, it doesn't follow that the model discussed above all of a sudden represents bubble chamber photographs or recorded patterns of electrical discharge.

The second reply is to invoke a chain of representational relationships. Brading and Landry (2006, 575) point out that the connection between a model and the world can be broken down in two parts: the connection between a model and a data model, and the connection between a data model and the world. So the structuralist could claim that scientific models represent data models in virtue of an isomorphism between the two and additionally claim that data models in turn represent phenomena. But the key questions that need to be addressed here are (a) what establishes the representational relationship between data models and phenomena, and (b) why, if a scientific model represented some data model, which in turn represented some phenomenon, that would establish a representational relationship between the model and the phenomenon itself. With respect to the first question, Brading and Landry argue that it cannot be captured within the structuralist framework. The question has just been pushed back: rather than asking how a scientific model qua mathematical structure represents a phenomenon, we now ask how a data model qua mathematical structure represents a phenomenon.[57] With respect to the second question, although scientific representation is not antitransitive, it is intransitive.[58] Suppose I take a photograph of one of Picasso's portraits of Dora Maar. The photograph represents the picture, the picture represents Dora Maar, but the photograph doesn't represent Dora Maar

57 Van Fraassen's (2008, Chapters 6,7) account of (a) is discussed below.

58 A relation R is antitransitive if $\forall x \forall y \forall z((Rxy \wedge Ryz) \rightarrow \neg Rxz)$. R is intransitive if $\neg \forall x \forall y \forall z((Rxy \wedge Ryz) \rightarrow Rxy)$ *cf.* fn 11 above.

(*cf.* Frigg, 2002, 11-12). To illustrate, Figure 7 represents the portrait (along with the author of this thesis), but not Dora Maar herself.

Figure 7: A picture that represents a portrait of Dora Maar, but doesn't represent Dora Maar

So more needs to be said regarding how a scientific model representing a data model, which in turn represents the phenomenon from which data are gathered, establishes a representational relationship between the first and last element in the representational chain.

The third reply is due to van Fraassen (2008). His 'Wittgensteinian' solution is to diffuse the loss of reality objection by appealing to features in the context of using models to represent their targets.[59] He claims that despite the distinction between data and phenomena, for a given scientist, in a given context, there is no difference between accurately representing the two. That accurately representing data is the same as accurately representing the system that provided it is claimed to be a 'pragmatic tautology ... [something that is] ... logically contingent but undeniable nonetheless' (2008, 259).

Van Fraassen's argument for this is one of the most significant contributions of his (relatively) recent *Scientific Representation: Paradoxes of Perspective* (2008), but it hasn't received the attention it deserves. I can only speculate about why, but I suspect that it is: in part due to the considerable novelty of many of the central notions used; in part due to the fact that the argument is spread out throughout the book, interwoven

59 The material in the remainder of this chapter is drawn from Nguyen (2016).

with substantial broader discussions of representation, measurement, and empiricism; and in part due to a style of presentation that is often difficult to penetrate. In fact, the project of extracting a coherent position from the rich and intricate lines of thought is beset with exegetic challenges. Since this is certainly his most developed account of scientific representation, and, in my opinion, the most developed of all the structuralist accounts on offer, it's worth investigating his argument for this claim in detail. This is my task for the remainder of this section.

But before turning to his argument, it is worth clarifying that he takes a morphism (or more specifically, an isomorphic embedding) to establish that a pre-existing representational relationship is *accurate* (as per something like Structuralism 2), rather than establishing the representational relationship itself (as per Structuralism 1).[60] This is despite the fact that he starts the discussion with the claim that the central question to be addressed is:

> '*How can an abstract entity, such as a mathematical structure, represent something that is not abstract*, something in nature?' (2008, 240, original emphasis).

But he then shifts to the question of how a structure can do so accurately:

> 'The question how an abstract structure can represent something ... is just this: *how, or in what sense, can such an abstract entity as a model "save" or fail to "save" this concrete phenomenon*?' (2008, 245, original emphasis).

And then, when presenting his solution, he couches it in terms of 'fit', 'match', 'empirical adequacy' and so on, and explicitly states:

> 'If a model were offered to represent the phenomenon, that structural relation [a model-phenomenon morphism] would determine whether the model was adequate with respect to its purpose' (2008, 249-250).

But, as mentioned above, van Fraassen is explicitly aware that phenomena are not the sorts of things that can enter into morphisms. And he accepts that the structuralist cannot finish the sort at the level of data. He phrases the objection as follows:

> 'Oh, so you say that the only 'matching' is between data models and theoretical [scientific] models. Hence the theory does not confront the observable phenomena, those things, events, and processes out there, but only certain representations [i.e. data models] of them.' (2008, 258).

60 See Thomson-Jones (2011a) for a discussion of how this distinction is important to van Fraassen's project.

He then admits that '[a]n empiricist account of what the sciences are all about must absolutely answer this objection' (2008, 258). Without an answer, the structural empiricist is left in the uncomfortable position, whereby it is data, the 'products of our independent intellectual activity', not phenomena, that are the ultimate targets of scientific models. Now onto his argument for the pragmatic equivalence between taking a model to accurately represent a data model, and taking it to accurately represent the phenomenon from which the data were gathered. The argument utilises notions that are relatively novel in the literature on scientific representation. So the first task is to outline these. This is a necessary first step in any critical evaluation of van Fraassen's developed philosophical position.

6.1.1 *Toolbox*

i. HAUPTSTATZ: *'There is no representation except in the sense that some things are used, made, or taken, to represent things as thus or so'* (2008, 23, *origional emphasis*). There are two important things to note about this. Firstly, it is clearly non-reductive as it invokes the intentions and acts of agents. Secondly, it involves representation-as, rather than representation-of.[61] Van Fraassen (2008, 16) explicitly refers to Goodman (1976) as the source of the distinction, and following them I assume that x is a representation-of y if and only if x denotes y. Representation-as is stronger: x represents y as thus or so if and only if x denotes y and attributes certain features to y. If y has those features then x accurately represents y with respect to them. To use one of van Fraassen's examples, the proper name 'Margaret Thatcher' is a representation of Margaret Thatcher, since it denotes her. But a caricature of Margaret Thatcher also represents her *as* thus or so, e.g. if she is depicted with horns and a tail then it represents her as being draconian (2008, 13-15).

ii. USE OF REPRESENTATIONS: Hauptsatz makes clear that representations only represent when they are used to do so. But in addition, certain representations have particular uses: 'they are typically produced for a certain use, with a certain purpose or goal' (2008, 76). Using maps to navigate provides an illustrative example: '[a] map is designed to help one get around in the landscape it depicts' (2008, 76). Throughout this discussion I assume that the analogous use of models is to generate predictions about their target systems. This is supported by van Fraassen's analogy between using a map to navigate and using the Aviation Model (AVN) for weather forecasting, i.e. to generate predictions about the weather (2008, 77).

61 Van Fraassen doesn't develop what establishes representation-as. In Part iii I do this, but outside of the structuralist framework.

III. LOGICAL SPACE: Representations, or at least scientific models, are associated with 'logical spaces'. This is a very general notion. Examples include PVT space in elementary gas theory, phase spaces in classical mechanics, and Hilbert spaces in quantum theory (2008, 164). Locations in PVT space are combinations of pressure, volume and temperature. Routes through a phase space are possible trajectories of an object, and locations in a Hilbert space are possible quantum states of a system.

IV. SELF-LOCATION: A necessary condition on using a map to navigate, or a model to predict, is that the user self-locates in the logical space provided. They 'must be in some pertinent sense able to relate him or herself, his or her current situation, to the representation' (2008, 80). In order to navigate with a map, the users must be able to locate themselves in the terrain depicted and associate that location with an area on the map. They distinguish a particular map region as representing where they are, they orient the map to correspond to the direction they're facing, and so on. In doing so they locate themselves with respect to the map. When it comes to scientific models, van Fraassen claims:

> 'Suppose now that science gives us a model which putatively represents the world in full detail. Suppose even we believe that this is so. Suppose we regard ourselves as *knowing* that it is so. Then still, before we can go on to use that model, to make predictions and build bridges, we must locate ourselves with respect to that model. So apparently we need to have something *in addition* to what science has given us here. The extra is the self-ascription of location' (2008, 83, original emphasis).

It's worth clarifying what 'self-location' could mean in the spaces under consideration. Although suggested by van Fraassen's cartographic analogy, I presume that it doesn't require that the model user locate herself in logical space. When it comes to measuring the pressure of his tire (2008, 181), what would it mean for van Fraassen to locate himself in PVT space? Van Fraassen is 100psi? A more charitable reading of 'self-location' is that the model users themselves actively locate the target system in logical space. And this proceeds in two steps. The model user first adopts a certain perspective towards the target by taking it to be the sort of thing that can be located in the logical space provided by the model. For example: van Fraassen takes the tire to be the sort of object that can be located in PVT space.

But although this may be a necessary condition on using a model to generate a prediction, it is not the condition van Fraassen has in mind when he invokes the cartographic analogy. It isn't enough that a navigator is located somewhere in the terrain depicted; we need to delineate a specific point, or at least an interval or region, of the space. This is the second step in self-location.

When it comes to generating predictions using scientific models, this is done by inputting the target's initial and boundary conditions:

> 'The AVN itself requires *input* to be run at all, of course: namely initial conditions and lateral boundary conditions obtained from operational weather centers in the relevant area [...] The model presents a space of possible states and their evolution over time – the input locates the weather forecaster in that space, at the outset of the forecasting process' (2008, 78, original emphasis).

Self-location demands that it is not enough that the system is in fact thereby located, but the model user must perform an act of location. To speak loosely, the user distinguishes a region in logical space with the claim 'that target system is there'.

V. MEASUREMENT AS LOCATION IN LOGICAL SPACE: 'the act of measurement is an act – performed in accordance with certain operational rules – of locating an item in logical space' (2008, 165). And these measurements deliver data models. As van Fraassen notes, the location needn't be a point, but can be a region (2008, 165). This can, but doesn't have to, be the result of measurement imprecision. Even a perfectly precise pressure reading p determines only a region of PVT space since there are multiple volume-temperature pairs compatible with p.

VI. MEASUREMENT AS REPRESENTATION: locating a system in logical space involves representing it as thus or so. This form of representation is not established by a morphism (recall van Fraassen's worry about invoking a 'dormative virtue'). Instead, data models represent because:

> 'A measurement is a physical interaction, set up by agents, in a way that allows them to gather information. The outcome of a measurement provides a representation of the entity (object, event, process) measured' (2008, 179-180).

A data model represents the system measured as having the features corresponding to the region of logical space where it is thereby located. If the system has those features, the data model is accurate.

VII. PRAGMATIC TAUTOLOGY: 'a pragmatic tautology is a statement which is logically contingent, but undeniable nevertheless. Similarly, a pragmatic contradiction is a statement that is logically contingent, but cannot be asserted' (2008, 259). Moore's paradox – utterances of the form 'P and it is not the case that I believe that P' - is a classic example of the latter. They are logically contingent - their form is an agent i asserting '$P \land \neg B_i(P)$', where $B_i(P)$ means i believes that P - and neither conjunct semantically entails the negation of the other (if they did, i would be clairvoyant). Such sentences are pragmatic contradictions because, in the context of i asserting P, i commits herself to believing P. It is this commitment that, when combined with the second conjunct, makes the sentence unassertable. Since van Fraassen's account of scientific representation does not involve linguistic representation, his argument requires generalising from the assertablity of sentences to certain acts of representation.

With the above notions in mind, we can now turn to van Fraassen's argument for the pragmatic equivalence between taking scientific models to accurately represent data and phenomena. My primary interest here is not the relationship between data and phenomena. For my current purposes I simply grant that data represent the systems from which they were gathered (as per *VI. Measurement as representation*). I further grant that morphisms play a role in establishing whether a scientific model represents, accurately or otherwise, data. I'm concerned with representational relationships between scientific models and phenomena. Although accurately representing a data model D might provide us with evidence that M is an accurate representation of T, this does not establish any representational relationship between M and T. Without this *Loss of Reality* remains.

6.1.2 *The Wittgensteinian move*

Van Fraassen's resolution to *Loss of Reality* is to claim that in the context of use there's no difference between accurately representing data and phenomena. The reasoning, which is found in pages 254-60, is illustrated with an example. I present it here before reconstructing the argument that underpins it. The example in question concerns only observable features of a target system (the observable-unobservable distinction is largely irrelevant in the current context). Focusing on observables makes it clear how important the Wittgensteinian move is to van Fraassen's project. If he fails to establish the pragmatic equivalence with respect to observable phenomena, then they fail to feature in his structuralist account of scientific representation. The result is a far more radical anti-realist position than has been offered previously, and is a far more radical position than I suspect van Fraassen would accept.

He begins by considering a scientist representing the growth of the deer population in the Princeton region. The scientific model used includes assumptions about environmental features: luscious gardens, the council's culling instinct, its tendency to experiment with birth-control measures for the local animal population, and so on. The data model D is supplied by a graph constructed from cleaned up data points gathered by field researchers measuring samples of 'values of various parameters over time' (2008, 255). Van Fraassen does not specify which parameters are measured, but given that the theory concerns the deer population growth, I assume that the scientist literally counts deer in representative regions throughout the duration of the experiment. So the graph plots the number of deer against time. The target system is the deer population itself.

The scientist has a model M about deer population growth and argues that M is morphic to D. Van Fraassen imagines a philosophical interlocutor, arguing that although M accurately represents D, the question is whether M accurately represents the population itself (2008, 249). The scientist showing the interlocutor that it matches D does not establish this.

Van Fraassen replies that the scientist has 'no leeway' to deny that the model accurately represents the actual population without withdrawing the graph altogether (2008, 256). According to him, the scientist should say:

> 'Since this is *my* representation of the deer population growth, there is *for me* no difference between the question whether [M] fits the graph [i.e. the data model] and the question whether [M] fits the deer population growth. If I were to opt for a denial or even a doubt, though without withdrawing my graph, I would in effect be offering a reply of form:
>
> - The deer population growth in Princeton is thus or so, but the sentence "The deer population growth in Princeton is thus or so" is not true, for all I know or believe' (2008, 249, original formatting).

And since a scientist who replied this way would be faced with a Moorean paradox, the scientist simply cannot doubt that the model accurately represents the target system whilst accepting that it accurately represents the graph. This is supposed to establish the pragmatic equivalence between the two.

That's the example, now let's work out why the scientist might be forced into such a position. In the rest of this section I reconstruct the argument for this conclusion in detail. I break it down into three sub-arguments, and show how the notions laid out in the previous subsection are utilised. It's important to notice that the first two arguments — which establish that the scientist must locate the target in the logical space of the model in order to use it at all, and that this is done with the graph –- are explicitly concerned with representation *simpliciter*. This accounts for the scientist's claim that 'this [data model] is *my* representation of the deer population growth' (2008, 249, emphasis in the original). The third argument then shifts to the question of accurate representation in an attempt to establish that, for that scientist, there is 'no difference between the question whether [M] *fits* the graph and the question whether [M] *fits* the deer population growth' (2008, 249, emphasis added). The first premise in the third argument makes it explicit how van Fraassen requires that the necessary act of representation established in the first two arguments must generate doxastic commitments, i.e. commit the model user to certain beliefs, if the third argument is to generate the pragmatic equivalence.

(A) The argument for self-location:

A1. A scientist *S* is using *M* to represent a target system *T* for certain purposes *P*. (Premise)

A2. If *S* is using *M* to represent a target *T* for purposes *P*, then *S* must self-locate in the logical space, *L*, provided by the model. (Premise)

A3. *S* must self-locate in *L*. (From A2 and A3)

M is a model of deer population growth, *T* is the target deer population, and the scientist is using *M* to represent *T* for the purpose of generating a prediction (*II. Use of Representations*). *M* provides a logical space *L*, the space of possible deer populations and their growth through time (*III. Logical Space*). A necessary condition on using *M* to generate a prediction about *T* is self-location in *L* (*IV. Self-location*).

(B) The argument from self-location to representation-as:

B1. *S* self-locates in *L* using a data model *D*. (Premise specifying A3)

B2. If *S* uses *D* to self-locate in *L*, then *S* uses *D* to represent *T* as thus or so (Π). (Premise)

B3. *S* uses *D* to represent *T* as Π. (From B2 and B3)

Argument (A) required that the scientist self-locate in *L*. In van Fraassen's example, this is done using a data model *D*, a graph of the deer population. When *S* uses *D* to represent the target system, *S* locates *T* in the logical space provided by the model (*V. Measurement as location*). Locating *T* in a region of *L* requires representing *T* as having the features corresponding to that region (*VI. Measurement as representation*). Let Π be the conjunction of predicates that corresponds to that region. This may be a region, not a point, so these predicates are of the form 'the magnitude of P_i is in region Δ'. In this instance P_i is the size of the deer population at particular times and the size of Δ corresponds to the potential measurement error induced by the counting process and the generalisation from representative samples to the population as a whole. So when using D to locate T in logical space the scientist represents T as Π.

(C) The argument from representation-as to the pragmatic tautology:

c1. The (pragmatic) content of S using D to represent T as Π includes S believing that T is Π. (Premise)

c2. If S is able to take M to accurately represent D, but not T, then S is able to express disbelief in any proposition concerning T that S commits herself to in using D to represent T. (Premise)

c3. If S is able to take M to accurately represent D, but not T, then S is able to express disbelief that T is Π. (from B3, C1, and C2)

c4. It is not the case that S is able to express disbelief that T is Π (whilst using D to represent T), on pain of pragmatic contradiction. (Premise)

---

c5. It is not the case that S is able to take M to accurately represent D, but not T. (From C3 and C4)

I return to C1 and C2 below. C3 follows from B3, C1 and C2. S represents T as Π (B3), and in doing so commits herself to the belief that T is Π (C1). This instantiates the universal quantifier in C2 delivering C3. C4 is the instance of Moore's paradox that van Fraassen is concerned with. He claims that if the scientist were to accept that M accurately represents D but not T, whilst using D to represent T as Π, S would be offering a reply of in the form of Moore's paradox ('the deer population is thus or so but ...'). Taking D to represent T is analogous to asserting the first conjunct. Denying that M accurately represents T is analogous to asserting the second conjunct (*VII. Pragmatic tautology*). This generates the pragmatic equivalence between accurately representing T and D (C5).

6.1.3 *The argument scrutinised*

With the argument reconstructed, I now turn to my critical discussion. My objections are the following. Firstly, the pragmatics of representation don't induce doxastic commitments: acts of representation don't commit the agent doing the representing to any relevant beliefs. So C1 is false. Secondly, one option available to van Fraassen is to amend C1 to the claim that S takes D to accurately represent T as Π. But this isn't supported by (A) and (B): it would require that in order to use a scientific model to generate a prediction, the model user must believe the inputted initial/boundary conditions. This is false. My final objection concerns C2, I argue that without an account of scientific representation (irrespective of accuracy), it's difficult to get a grip on what it would mean for S to deny that M accurately represents T.

The pragmatics of representation

The argument has the following macro structure. Models are used to generate predictions about their targets and a necessary condition on doing this is that the user locate the target in the model's logical (A). This is typically done with a data model, and when S uses a data model to locate a target system T in such a way, S represents T as Π (B). So far so good.

C1 is vital for rest of the argument, since it is the move from S representing T as Π to pragmatically committing herself to the belief that T is Π that is required to generate the pragmatic tautology. Using the data model to represent the target system is supposed to commit S to the belief that the deer population is thus or so in a way analogous to asserting the first conjunct of the Moorean paradox. The denial that the model accurately represents the deer population then provides the analogy with asserting the second.

But all arguments (A) and (B) established is that S represents T as Π. And acts of representation do not incur the same pragmatic commitments as acts of assertion. Consider the example of representing Margaret Thatcher as draconian. A caricaturist can represent Thatcher as such without committing herself to the belief that Thatcher is draconian. There is a vital pragmatic difference between acts of representation and assertions. If the caricaturist were to assert that Margaret Thatcher was draconian, then she would commit herself to believing such. But the caricaturist doesn't do this; she merely represents Thatcher in such a way. The artist could have been commissioned to draw the caricature despite having only a vague idea of who Thatcher was, and no knowledge about her time as Prime Minister. The artist can reasonably draw the caricature, thereby representing her as draconian, whilst at the same time remaining agnostic about her character. The same point applies to scientific representation: S's act of representing the target system in a certain way doesn't pragmatically commit her to the belief that the target is that way.

It pays to be careful here. My claim does not concern whether or not S actually believes that T is Π, it is a conceptual point regarding the pragmatics of assertion and representation. Presumably in most cases, model users do believe that the initial/boundary conditions used are (at least approximately) accurate. But this does not establish that an agent's act of representing something in a particular way commits that agent to any particular beliefs in the way that acts of assertion do in the traditional version of Moore's paradox. So C1 is false, *S*'s act of representing a target system as thus or so doesn't commit *S* to the belief that the target is thus or so. Therefore (C) is unsound.

A possible response is to invoke a weaker doxastic attitude than belief as being incurred in representing a target system. And although this attitude might not deliver the Moorean paradox van Fraassen discusses, it may deliver a closely related pragmatic contradiction that still allows a version of (C) to go through. In other contexts van Fraassen invokes the attitude of acceptance (Muller and van Fraassen, 2008). Accepting a theory, or model, is to take it to be empirically adequate: to believe its observable content and to remain agnostic about its unobservable content (acceptance is typically applied to scientific models, but here I'm considering applying it to data). So, what happens if, in using D to represent T as Π, S commits herself to accepting that T is Π? Well that depends on T and Π. We can distinguish between the observable and unobservable content of $\Pi(T)$, denoted $\Pi(T)^O$ and $\Pi(T)^U$ respectively. If S accepts $\Pi(T)$, then S commits herself to believing $\Pi(T)^O$ and being agnostic about $\Pi(T)^U$, i.e. not believing $\Pi(T)^U$ or $\neg\Pi(T)^U$ (see Muller and van Fraassen, 2008, 204).

For neither of these types of content will acceptance do the work required. Regarding observable content we are back where we started. Accepting that Thatcher is draconian entails believing that she is. And an agent can represent her in such a way without taking on this commitment. Regarding unobservable content, S accepting $\Pi(T)^U$ entails $\neg B_s(\Pi(T)^U)$ and $\neg B_s(\neg\Pi(T)^U)$. But this will not generate a pragmatic contradiction when combined with the second conjunct of van Fraassen's instance of Moore's paradox, i.e. $\neg B_S(\Pi(T))$ (even restricted to its unobservable content).

Invoking acceptance when an agent uses a data model to represent a target doesn't work. But the above discussion suggests another available strategy available to van Fraassen. It proceeds in two steps. Firstly, introduce a weaker act than assertion – call it entertaining – and assume that an act of entertaining that P incurs a commitment to not believing $\neg P$. Again this alone doesn't generate a pragmatic contradiction when combined with $\neg B_i(P)$. But it does when combined with $B_i(\neg P)$. The second step is to move from a Moorean paradox of the form $P \wedge \neg B_i(P)$ to one of the form $P \wedge B_i(\neg P)$ – i.e. from sentences like 'it's raining and I don't believe it's raining' to 'it's raining and I believe that it's not raining'. (C) then becomes (C*):

C1*. The (pragmatic) content of S using D to represent T as Π includes S not believing that it is not the case that T is Π. (Premise)

C2*. If S is able to take M to accurately represent D, but not T, then S is able to express belief in the negation of any proposition concerning T that S commits herself to in using D to represent T. (Premise)

C3*. If S is able to take M to accurately represent D, but not T, then S is able to express belief that it is not the case that T is Π. (From B3, C1*, and C2*)

C4*. It is not the case that S is able to express belief that it is not the case that T is Π (whilst using D to represent T), on pain of pragmatic contradiction. (Premise)

C5*. It is not the case that S is able to take M to accurately represent D, but not T. (From C3* and C4*)

Assuming that an act of representation is an act of entertaining, in using D to represent T as Π, S pragmatically commits herself to not believing that it is not the case that T is Π , i.e. $\neg B_S(\neg\Pi(T))$ (C1*). Further assume that S denying that M accurately represents T whilst accepting it accurately represents D, induces a commitment to believing that it is not the case that T is Π (C2* and C3*). This is a stronger commitment than assumed in C, $B_S(\neg\Pi(T))$ rather than $\neg B_S(\Pi(T))$. Under these assumptions, if S were to take M to accurately represent D but not T, whilst at the same time using D to represent T, she would be offering a reply with the following commitments: 'It's not the case that I believe that T isn't Π and I believe that T isn't Π'. This would be a pragmatic contradiction.

(C1*) requires that in using D to represent T as Π, S entertain that T is Π, and therefore commit herself to not believing that T isn't Π. However, the following example shows that even this commitment is not incurred by acts of representation. Consider a different caricaturist representing Margaret Thatcher as draconian. This time assume that the Labour Party has commissioned the caricature and the artist is a staunch Conservative. He goes ahead and draws the picture because he is desperate for the money. In drawing the caricature the artist represents Thatcher as draconian, but he certainly doesn't believe it. In fact, he explicitly believes that she isn't draconian to the extent that he sings her praises whilst drawing the caricature. This makes him feel better about drawing something that goes so strongly against his political beliefs. Now, if, in representing Thatcher as draconian, the artist commits himself to not believing that she isn't, then his act of drawing her as such whilst singing the negation would be a pragmatic contradiction. But although a strange situation, this isn't the case. Acts of representing that P don't incur the pragmatic commitment to $\neg B_i(\neg P)$. So C1* is false, and (C*) unsound.

From self-location to belief

Despite van Fraassen's phrasing, the above concerns suggest that argument (C) shouldn't start from the premise that *S* uses *D* to represent *T* as Π, but rather *S* takes *D* to accurately represent *T* as Π. Rather than:

> 'Since [*D*] is *my* representation of the deer population growth, there is *for me* no difference between the question whether [*M*] fits [*D*] and the question whether [*M*] fits the deer population growth' (2008, 256, original emphasis)

the scientist should say:

> Since *I* take *D* to be an accurate representation of the deer population growth, there is *for me* ...

It's plausible that in taking *D* to accurately represent *T* as Π, *S* commits herself to believing that *T* is Π. But since B3 only got us as far as representation, the preceding argument needs amending. Argument (A) stays as it is. In order to use a model to generate a prediction, the model user must self-locate in its logical space. (B) gets revised to ($B^{\dagger}$):

$B1^{\dagger}$. *S* self-locates in *L* using a data model *D*. (Premise specifying A3)

$B2^{\dagger}$. If *S* uses *D* to self-locate in *L*, then *S* takes *D* to accurately represent *T* as Π. (Premise)

$B3^{\dagger}$. *S* takes *D* to accurately represent *T* as Π. (From $B2^{\dagger}$, $B3^{\dagger}$)

And if this can be established then a revised version of (C) runs as follows[62]:

$C1^{\dagger}$. The (pragmatic) content of *S* taking *D* to accurately represent *T* as Π includes *S* believing that *T* is Π. (Premise)

$C2^{\dagger}$. If *S* is able to take *M* to accurately represent *D*, but not *T*, then *S* is able to express disbelief in any proposition concerning *T* that *S* commits herself to in taking D to accurately represent T. (Premise)

62 ($C^{\dagger}$) is a revised version of (C), not (C*), but my criticisms can be run against a revised version of the latter as well.

C3†. If S is able to take M to accurately represent D, but not T, then S is able to express disbelief that T is Π. (From B3†, C1†, and C2†)

C4† It is not the case that S is able to express disbelief that T is Π (whilst using D to represent T), on pain of pragmatic contradiction. (Premise)

C5†. It is not the case that S is able to take M to accurately represent D, but not T. (From C3†, C4†)

But although (C†) seems plausible in isolation, the argument as a whole is not, for B2† is false.

To see why, recall what self-location required. The model user had to adopt a certain perspective towards the target by taking it to be the sort of thing that could be located in the model's logical space. She then had to delineate an area within that space for the target. This was a necessary condition on generating a prediction using the model. But neither of these steps commits the agent to any beliefs. In particular, in using a data model to self-locate in a model's logical space, the model user does not thereby commit herself to the data's accuracy.

Consider again the example of the deer population. In order to use her model to generate a prediction about its size, the scientist had to input an initial number of, and fitness values for, the deer. The model allows the scientist to make any number of predictions about the future size of the population. If the scientist inputs a low fitness value – imagine a pro-cull council – then the model will predict a small future population. If the scientist initially assumes that the deer population is too large for the region to support, then the model will predict population decline. And so on. The scientist can use the model to generate numerous predictions about the deer population regardless of whether or not she believes these values to be accurate. All of these inputs serve to delineate the logical space of the model, and some input is necessary to generate a prediction about the target. But she is not required to believe them.

Other examples abound. Some are of scientists failing to believe that the logical space of a model is correct. Ptolemaic models can be used to generate predictions about planetary orbits without the user believing that those planets in fact are located anywhere in the model's logical space. State of the art global climate models (GCMs) contain variables that are known to describe model-processes with no direct real-world correlates. These variables – in that context typically referred to as 'parameters' – are loosely related to sub-grid processes such as small-scale convection and cloud coverage. However, their values depend to a large extent on details of the particular computational scheme used rather than on the state of the world. So we have here a case where scientists don't believe that the logical space is correct (at least not completely correct), and yet they pick values for certain variables in order to make calculations (Bradley et al., 2016). And this is no isolated instance; one can find similar cases, for example, in economics (Friedman, 1953) and population dynamics.[63]

The problems don't end here. Even supposing that the scientist believes that the logical space is correct, they still needn't believe that the target is located in the region delineated by the model input. For example, representative concentration pathways (RCPs) are used to locate the global climate in the logical space of GCMs. They supply concentration trajectories of the main forcing agents of climate change. One particular pathway, RCP2.6, requires that we essentially eliminate greenhouse gas emissions immediately, something that no one believes is, or will be, the case. And yet RCP2.6 is widely used to generate numerous predictions about the global climate (see the 2013 IPPC report (Stocker et al.), Ch.12 in particular).

The point is that the scientists can use models to generate predictions about target systems without adopting any epistemological position towards the model, or where the target is located in its logical space. As stressed previously, this isn't to say that scientists *don't* believe that their data models and initial/boundary conditions are (at least approximately) accurate. My claim is that this belief is not a *necessary* condition on using a model to generate a prediction. As such it is not part of the pragmatic content of locating a target in logical space. And this is what van Fraassen requires. So, although ($C^\dagger$) may seem plausible in isolation, it rests on ($B^\dagger$) for its support, which in turn rests on the false premise $B2^\dagger$. So the argument as a whole is unsound.

Representation and accurate representation

I hope by now to have shown that van Fraassen's argument fails. But before moving on, it is worth highlighting a further issue with the argument. It concerns accurate representation, and the question of what establishes the prior representational relationship is not addressed. I suspect that van Fraassen would fall back on his *Hauptsatz* and claim that representation cannot be analysed beyond this. But this does not help when we look at C2 (or its variants). In particular, what would it mean for S to deny that M accurately represents T? Van Fraassen's phrasing suggests that in doing so S would take M to represent T, but to do so inaccurately. In this sense the deer scientist would be effectively asserting the second conjunct of van Fraassen's version of Moore's

63 See for example Weisberg and Reisman (2008) who offer individual-based versions of the Lotka-Volterra model that start from the assumption that individuals move about on a 30×30 toroidal lattice.

paradox (the sentence 'the deer population is thus or so' is not true for all I know or believe).

But this needn't be the question the philosophical interlocutor asks. Rather than asking whether the scientific model matches the phenomenon, they can ask whether the model represents it in the first place (perhaps by doubting whether STRUCTUALISM 2 is the correct answer to the ER-problem). And if the scientist were to doubt that M represents T in this sense, then (C) (and its variants) will again fail irrespective of my previous criticisms. I have established that acts representation failed to incur doxastic commitment. But what about representational denial, as it occurs in C2/C2*/C2†? Consider a caricature that depicts David Cameron as draconian. Denying that it represents Margaret Thatcher doesn't incur a commitment to believing (or disbelieving) anything about Thatcher, other than she isn't the one caricatured. That an agent incur any doxastic commitments in the denial of representational relationships is even less plausible than him incurring them when affirming them.

For these reasons, van Fraassen's attempt to establish a pragmatic equivalence between taking a scientific model to accurately represent data (via a morphism), and accurately represent the phenomenon from which the data was gathered doesn't seem to work. This leaves the question of target-end structures unanswered. I'm sceptical that data provide them, so it's now time to turn to the alternative option available to the structuralist: locate the target-end structures in the phenomena themselves.

6.2 THE STRUCTURE OF THE WORLD

Locating target-end structures 'in' the target themselves can be done in at least two different ways: one can either argue that target systems, in some sense, *instantiate* structures; or alternatively one can attempt to *identify* target systems with structures. I discuss each of these approaches in turn.

6.2.1 *Structure instantiation*

The idea that target systems instantiate structures is intuitively straightforward. Recall the definition of a structure given in the previous chapter:

$$\mathcal{S} = \langle \mathcal{D}, \{R_i\}_{i \in I} \rangle$$

where $\mathcal{D}$ is a domain of objects, and each R_i^n an extensionally defined n-ary relation over $\mathcal{D}^n$. The idea then, is that a physical system can instantiate such a structure in virtue of consisting of a collection of parts that enter into various physical relations with one another. The objects can then be used to define the domain of individuals, and by considering the physical properties and relations purely extensionally, we arrive at a class of extensional relations defined over that domain (see for instance Suppes' (2002, 22) discussion of the solar system). This supplies the required notion of structure. We might then say that physical systems instantiate a certain structure, thus providing a candidate for the proposed morphism to link to a scientific model.

As an example consider the methane molecule. The molecule consists of a carbon atom and four hydrogen atoms grouped around it, forming a tetrahedron. Between each hydrogen atom and the carbon atom there is a covalent bond. One can then take the atoms to be the objects and the bonds to be the relations. Denoting the carbon atom by a, and the four hydrogen atoms by $b, c, d,$ and e, we obtain a structure $\mathcal{S}$ with the domain $\{a, b, c, d, e\}$ and the relation $R_1^2 = \{\langle a, b\rangle, \langle b, a\rangle, \langle a, c\rangle, \langle c, a\rangle, \langle a, d\rangle, \langle d, a\rangle, \langle a, e\rangle, \langle e, a\rangle\}$ which is the extensional counterpart to the relation of being connected by a covalent bond.

The main problem facing this approach is the underdetermination of target-end structure. Underdetermination threatens in two distinct ways. Firstly, in order to determine the structure determined by a target system, a domain of objects is required. What counts as an object in a given target system is a substantial question (Frigg, 2006). One could just as well choose bonds as objects and consider the relation to be the sharing of a node with another bond. Denoting the bonds by $\alpha, \beta, \gamma,$ and δ, we obtain a structure $\mathcal{S}'$ with the domain $\{\alpha, \beta, \gamma, \delta\}$ and the relation $R_1'^2 = \{\langle \alpha, \beta\rangle, \langle \beta, \alpha\rangle, \langle \alpha, \gamma\rangle, \langle \gamma, \alpha\rangle, \langle \alpha, \delta\rangle, \langle \delta, \alpha\rangle, \langle \beta, \gamma\rangle, \langle \gamma, \beta\rangle, \langle \beta, \delta\rangle, \langle \delta, \beta\rangle, \langle \gamma, \delta\rangle, \langle \delta, \gamma\rangle, \}$. Obviously $\mathcal{S}$ and $\mathcal{S}'$ are not isomorphic (their domains don't even have the same cardinality). So which structure is picked out depends on how we think about the system; depending on which parts we take to be objects and which parts we take to be relations, very different structures can emerge.

I take it that there is nothing special about the methane molecule in this regard, and that any target system can be carved up in different ways, each of which delivers an alternative structure instantiated by the target. So I think that the lesson generalises: there is no such thing as *the* structure of a target system. Systems only have a structure when certain things are taken as objects, and others are taken as relations. And there are many non-equivalent ways of thinking about target systems in this way. So in order to make sense of the structuralists' idea that scientific representation (accurate or otherwise) is grounded in the fact that scientific models are appropriately morphic to their targets, we have to take into account how the target system is 'carved up'. Under one carving of a target T, a model M might be morphic to T, under another it might not. But then this carving has a role to play in establishing the representation relation. So an account of scientific representation should explicate this role.

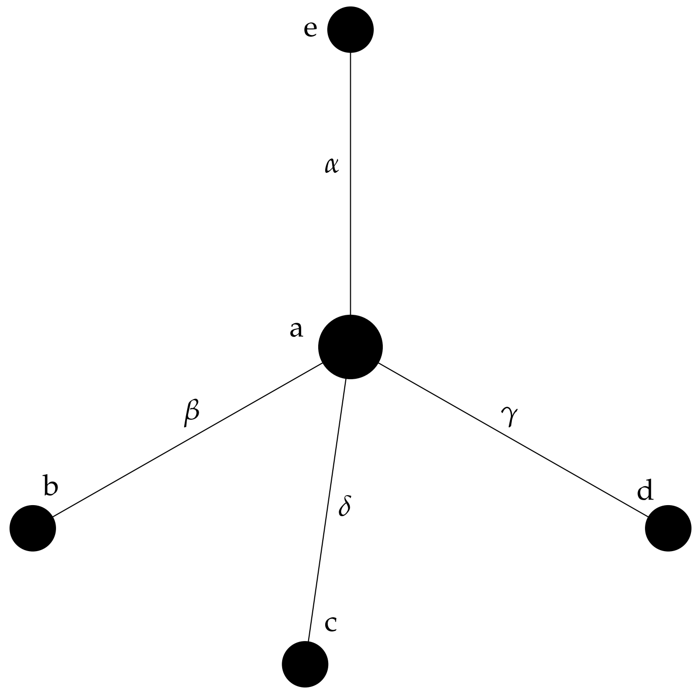

Figure 8: The structure of a methane molecule (*cf.* Frigg, 2006, 57)

One way of thinking about how we 'carve up' target systems is the way in which we describe them. In the case of the methane molecule we can describe it as a system of atoms related by bonds, which gives rise to $\mathcal{S}$, or a system of bonds related by sharing a node with one another, which gives rise to $\mathcal{S}'$ instead. And when the system is described as in the former case, the latter structure is, in some sense, illegitimate (and *mutatis mutandis* when it is described as in the latter case). I think it's plausible that introducing the notion of a 'target-system-under-a-description' will provide a way in which to discriminate between the multiple structures that arise in the manner discussed above.

Whether this is an acceptable way of dealing with the problem of multiple structures depends on how austere one's conception of models is. The semantic view of theories, which is most strongly associated with the structuralist accounts of scientific representation, was in part an attempt to divorce scientific representation from any linguistic trappings. Many proponents of the view emphasised that the model-world relationship ought to be understood as a purely structural relation, without any reference to language. Van Fraassen, for instance, rather boldly, claims that 'no concept which is essentially language dependent has any philosophical importance at all' (1980, 56) and takes it that '[t]he semantic view of theories makes language largely irrelevant' (1989, 222). And other proponents of the view, while less vocal about the irrelevance of language, have not assigned language a systematic place in their analysis of theories or scientific representation.

For someone who wants to exorcise language from scientific representation the above argument is bad news. However, a more liberal position could integrate descriptions in the package of modelling, but this would involve abandoning the idea that representation can be cashed out solely in structural terms. Bueno and French have recently endorsed such a position. They accept the point that different descriptions lead to different structures and explain that such descriptions would involve 'at the very least some minimal mathematics and certain physical assumptions' (2011, 887). Likewise, 'Munich' structuralists explicitly acknowledge the need for a concrete description of the target-system (Balzer et al., 1987, 37-39). This is a plausible move, but those endorsing this solution have to concede that there is more to representation than structures and morphisms.

The second way in which structural indeterminacy can surface is via Newman's theorem. The theorem essentially says that any system instantiates any structure, the only constraint being cardinality.[64] Hence, any structure of cardinality Γ is isomorphic to a target of cardinality Γ because the target instantiates any structure of cardinality Γ.

The argument can be run as follows. Assume $\mathcal{M} = \langle \mathcal{D}, \{R_i\}_{i \in I} \rangle$ is a scientific model, $|\mathcal{D}| = \Gamma$ (where the vertical bars denote cardinality in the usual way), and $\mathcal{M}$ represents a target system whose domain $\mathcal{D}'$ is such that $|\mathcal{D}'| = \Gamma$. Then we can prove that there exists a structure $\mathcal{T}$, whose domain is $\mathcal{D}'$ and $\mathcal{M}$ is isomorphic to $\mathcal{T}$. We do so by first noting that since $|\mathcal{D}| = |\mathcal{D}'|$, there exists a bijection $\pi : \mathcal{D} \rightarrow \mathcal{D}'$. Then for each $R_i^n \in \{R_i\}_{i \in I}$ we can define a relation $R_i'^n$ as follows: $R_i'^n = \{\langle y_1, ..., y_n \rangle \in \mathcal{D}'^n : \langle \pi^{-1}(y_1), ..., \pi^{-1}(y_n) \rangle \in R_i^n\}$. Doing this for all $R_i^n \in \{R_i\}_{i \in I}$, and letting the indices of the constructed relations on $\mathcal{D}'$ match the indices of the relations on $\mathcal{D}$ from which they were constructed, we arrive at an indexed set of relations $\{R_i'\}_{i \in I}$ on the n-ary products of $\mathcal{D}'$. Then, by letting $\mathcal{T} = \langle \mathcal{D}', \{R_i'\}_{i \in I} \rangle$, we have provided a structure of the target system that is isomorphic to $\mathcal{M}$ by construction. Since we started with an arbitrary $\mathcal{M}$, and the only constraint on the construction is that the target system have a domain whose cardinality matched that of the model's domain, it follows that the target system will instantiate a structure isomorphic to any scientific model whatsoever, as long as their cardinalities match.[65] And this means that the proposal that the target and model are isomorphic in STRUCTURALISM 2 will be correct

64 A very similar conclusion is reached in Putnam's so-called model-theoretic argument; see Demopoulos (2003) for a discussion.

65 See Ketland (2004); Frigg and Votsis (2011); Ainsworth (2009) for further discussions of this objection.

just so long as the two have the same cardinality (other morphisms are satisfied even more easily). Which will mean that accurate representation boils down to getting the cardinality of the target right.

Newman's theorem is both stronger and weaker than the argument from multiple conceptualisations of the target. It's stronger in that it provides more alternative structures. It's weaker in that many of the structures it provides are 'unphysical' because the relations defined may be purely set theoretical combinations of elements. By contrast, the multiple conceptualisations of the target pick out structures that a system can reasonably be seen as possessing in a substantial sense. The above response of picking out target-end structures via descriptions might work again here. By describing the relations on the domain of the target we rule out ones that don't correspond to something we have described, thereby, at least in some cases, ruling out the relations constructed by Newman's theorem. In the literature, solutions to Newman's objection either turn on linguistic considerations about how to to construct so-called 'Ramsey-sentences' and/or require that among all structures formally instantiated by a target system one is singled out as being the 'true' or 'natural' structure of the system (Ainsworth (2009) provides as a useful summary of the different solutions). The first family of solutions is not available to the structuralist, since they have no systematic place for linguistic formulations of theories or models. Moreover, how to single out a single structure as being the 'natural' or 'true' one instantiated by a target system in the structuralist framework remains unclear.

6.2.2 *The world is a structure*

An alternative way of picking out the structures of the target system is more radical. Rather than appealing to the idea that they *instantiate* them, one might instead *identify* target systems with structures. If this is the case then there is no problem with the idea that they can be morphic to a scientific model. One might expect ontic structural realists to take this position. If the world fundamentally is a structure, then there is nothing mysterious about the notion of an morphism between a model and the world. Surprisingly, some ontic structuralists have been hesitant to adopt such a view (see French and Ladyman (1999, 113) and French (2014, 195)). Others, however, seem to endorse it. Tegmark (2008), for instance, offers an explicit defence of the idea that the world simply is a mathematical structure. He defines a seemingly moderate form of realism – what he calls the 'external reality hypothesis (ERH)' – as the claim that 'there exists an external physical reality completely independent of us humans' (Tegmark, 2008, 102) and argues that this entails that the world is a mathematical structure (his 'mathematical universe hypothesis') (Tegmark, 2008, 102). His argument for this is based on the idea that a so-called 'theory-of-everything' must be expressible in a form that is devoid of human-centric 'baggage' (by the ERH), and the only theories that are devoid of such baggage are mathematical, which, strictly speaking, describe mathematical structures. Thus, since a complete theory of everything describes an external reality independent of humans, and since it describes a mathematical structure, the external reality itself is a mathematical structure.

This approach stands or falls on the strengths of its premise that a complete theory of everything will be formulated purely mathematically, without any 'human baggage', which in turn relies on a strict reductionist account of scientific knowledge (Tegmark, 2008, 103-104). Discussing this in any detail goes beyond my current purposes. But it is worth noting that Tegmark's discussion is focused on the claim that fundamentally the world is a mathematical structure. Even if this were the case, it seems irrelevant for many current scientific models, whose targets aren't at this level. When modelling an aeroplane wing we don't refer to the fundamental super-string structure of the bits of matter that make up the wing, and we don't construct wing models that are isomorphic to such fundamental structures. So Tegmark's account offers no answer to the question about where structures are to be found at the level of non-fundamental target systems.

So, as discussed in Chapter 1, the ER-problem is not distinct from questions of ontology, and any structuralist answer to the former requires that there is a target-end structure with which scientific models could be appropriately morphic to. Where to find this structure remains, by and large, an open question. Appealing to data fails to account for the fact that, ultimately at least, scientific models represent phenomena. The alternative approach, of appealing to the idea that phenomena themselves are 'structured' faces underdetermination issues that have yet to be resolved within the structuralist framework.

7

FICTIONS AND THE DIRECT VIEW

The theories of representation I have discussed so far take it for granted that there are model systems and construe representation as a relation between two entities; these and their target systems. Following Toon (2012, 43), this can be called the *indirect view* of scientific representation.[66] In the case of physical models this relation simply relates two physical objects. In the case of non-physical models, the idea is that descriptions of models pick out non-physical model systems, which in turn represent target systems. Such an approach can be clearly seen in the views discussed in Chapters 4, 5, and 6: these non-physical model systems are either abstract objects, fictions, or mathematical structures, and these objects are the vehicles of scientific representation in virtue of some (possibly intended) similarity relation between them and their targets.

This can be contrasted with the *direct view* which aims to do without model systems at all and takes scientific descriptions to directly describe (in a certain fashion) target systems themselves.[67,68] In this chapter I outline the direct view associated with Toon (2010b,a, 2012) and Levy (2015) and argue that it is unmotivated; that it has trouble accounting for models lacking target systems (for one reason or another); and that the comparison of 'models' and their targets remains undeveloped.

7.1 WHO NEEDS MODEL-SYSTEMS ANYWAY?

To understand the direct view, it is useful to briefly outline what it is a reaction to. According to the indirect view, scientific descriptions – be they descriptions of specific

66 Weisberg (2007) takes this as a defining feature of scientific modelling and argues that cases of scientific representation which are not indirect should not be considered cases of scientific modelling.

67 Levy (2012), who discusses the direct vs. indirect views in the context of fictional accounts of the ontology of scientific models refers to the former as the 'worldly fiction' view, and the latter as the 'whole cloth' view. In his more recent discussion of the issue he uses Toon's terminology (Levy, 2015).

68 In the case of physical models the idea would be that they too describe, in some sense, target systems, without treating them as model systems in their own right. I return to this when explicating the direct account below.

models ('imagine an infinite population of rabbits breeding at such and such a rate'), or classes of models ($T = 2\pi\sqrt{m/k}$, where T = an oscillation period, m is a mass term, and k a spring constant) – pick out model systems (an infinite population of rabbits, ideal pendulums) which in turn represent target systems (an actual finite population of rabbits, an actual bob bouncing on a spring). As discussed in Chapter 1, this approach faces the problem of ontology; what are these model systems when they are not physical?

Levy (2015) discusses two ways of answering this question; these model systems can be taken to be *abstract* objects, without physical properties, or they can be taken to be *concrete* but *hypothetical* objects, to which we can ascribe physical properties, even if we do not bear spatiotemporal relations to them in any straightforward sense. One popular way of cashing out the latter approach is to appeal to the idea that model systems are fictions in the same way that the characters and situations in Lord of the Rings, or the Hitchhiker's Guide to the Galaxy, are fictions (Godfrey-Smith, 2006; Frigg, 2010a,b).

Both Levy and Toon think that the idea that models are fictions is a step in the right direction, but they both argue that appealing to the idea that they are stand-alone systems is unnecessary. Rather than treating them as systems in their own right – entirely distinct from their targets – we should instead take model descriptions to be fictional in the same way that we (or at least Levy and Toon) take historical fiction to be fictional. To borrow Levy's example, *I Claudius*, both the book and the TV series, is a fiction. But we don't need to invoke a separate 'world' of which every sentence in the book is true. Rather, we can treat the book as being directly about *actual* historical Rome, albeit in such a way that the sentences needn't necessarily be literally true of it. By bypassing model systems entirely they take it that they have avoided the problem of ontology altogether.

So let's turn to the details of their accounts. Since Levy (2015, 790) sees his approach as 'largely so complementary to Toon's' I present Toon's account before turning to Levy's.

Toon (2010b,a, 2012) takes as his point of departure Walton's (1990) pretense theory. At the heart of this theory is the notion of a game of make-believe. The simplest examples of these games are children's plays (Walton, 1990, 11). In one such play we imagine that stumps are bears and if we spot a stump we imagine that we spot a bear. In Walton's terminology the stumps are *prop*s, and the rule that we imagine a bear when we see a stump is a *principle of generation*. Together a prop and principle of generation prescribe what is to be imagined. If a proposition is so prescribed to be

Figure 9: A statue of Napoleon on horseback (public domain)[70]

imagined, then the proposition is fictional in the relevant game.[69] The set of propositions actually imagined by someone need not coincide with the set of all fictional propositions in the game. It could be the case that there is a stump somewhere that no one has seen and hence no one imagines that it's a bear. Yet the proposition that the unseen stump is a bear is still fictional in the game.

Walton considers a vast variety of different props. In the current context two kinds are particularly important. The first are objects like statues. Consider a statue showing Napoleon on horseback (Toon, 2012, 37). The statue is the prop, and the games of make-believe for it are governed by certain principles of generation that apply to statues of this kind. So when seeing the statue we are mandated to imagine, for instance, that Napoleon has a certain physiognomy and certain facial expressions. We are not mandated to imagine that Napoleon was made of bronze, or that he hasn't moved for more than 100 years.

The second important kind of props are works of literary fiction. In this case the text is the prop, which together with principles of generation appropriate for novels of a certain kind, generates fictional truths by prescribing readers to imagine certain things. For instance, when reading The War of the Worlds we are prescribed to imagine that the dome of St Paul's Cathedral has been attacked by aliens and now has a gaping hole on its western side (Toon, 2012, 39).

69 The term 'fictional' has nothing to do with falsity; rather it indicates that the proposition is 'true in the game'.

70 Available here https://commons.wikimedia.org/wiki/File:Napoleon_statue_cherbourg.jpg#/media/File:Napoleon_statue_cherbourg.jpg.

In Walton's theory something is a representation if it has the social function of serving as a prop in a game of make believe, and something is an object of a representation if the representation prescribes us to imagine something about the object. In the above examples the statue and the written text are the props, and Napoleon and St Paul's Cathedral, respectively, are the objects of the representations.

The crucial move now is to say that models are props in games of make-believe. Specifically, material models – such as an architectural model of the Forth Road Bridge – are like the statue of Napoleon: the model is the prop and the Bridge is the object of the representation (Toon, 2012, 39). The same observation applies to non-physical models, such as the mechanical model of a bob bouncing on a spring. The model portrays the bob as a point mass and the spring as perfectly elastic. Rather than thinking of the model as a system in its own right, we are urged to treat the model description as representing the real ball and spring system in the same way in which a literary text represents its objects (Toon, 2012, 39-40): the model description prescribes imaginings about the real system – we are supposed to imagine the real spring as perfectly elastic and the bob as a point mass. This supplies the following direct account of scientific representation:

direct view: '*M* is a model-representation if and only if *M* functions as a prop in a game of make-believe' (Toon, 2012, 62)

where *M* ranges over physical models and model-descriptions in the case of non-physical models.

It should now be apparent how Toon's account is a direct view of modelling. Theoretical model descriptions represent actual concrete objects, e.g an actual bob bouncing on a spring. There is no intermediary entity of which model descriptions are literally true and which are doing the representing. Models prescribe imaginings about a real world target, and that is what representation consists of. The same holds in the case of physical models; although they are objects in their own right, the way they represent their targets is by prescribing us to imagine certain things about their targets.

It's worth briefly outlining how Direct View answers the questions laid out in Chapter 1. Firstly, as least as the definition is stated, the view appears committed to the idea that there is no demarcation between scientific and non-scientific epistemic representation. Just as the Phillips-Newlyn machine prescribed imaginings about an economy, the statue prescribes imaginings about Napoleon. Both function as props in games of make-believe. Indeed Direct View goes as far as to include all works of literary fiction (or at least all that Walton's approach reasonably applies to) within its purview. However, Toon does note a possible distinction between scientific modelling and works of fiction. He suggests that the former involves the representation of various different objects at different times (i.e. in virtue of prescribing us to imagine things about different target systems in different contexts), whereas 'it seems that the object (or objects) of a work of fiction are usually fixed: paintings or novels rarely represent different objects at different times' (2012, 63).

As previously mentioned, the account doesn't face the problem of ontology, since the only 'objects' involved beyond target systems are physical models and model-descriptions. The account has little to say about the problem of style, or the applicability of mathematics. But this is not to say that they couldn't be incorporated within the framework. One way of doing this could be to group together different props, or principles of generation, in such as way as to form the locus of a classification system by styles. For example, the principles of generation associated with particle physics seem very different to the principles of generation associated with population biology. Alternatively, material models seem to prescribe imaginings about their targets in a different manner to the way linguistic model-descriptions do. Similarly, the account could be extended in such a way that when the model-descriptions are mathematical in character, they prescribe us to imagine that their targets are governed by mathematical equations, an insight that could be extended to provide a more fully fledged account of the applicability of mathematics and one that to the best of my knowledge, hasn't been considered in the literature concerning that question. The problem of misrepresentation is dealt with in a straightforward manner, since there is no expectation that what a model prescribes us to to imagine about a target systems be literally true of that system.

Unfortunately, the account fares less well with respect to answering the ER-problem in such a way that accounts for surrogative reasoning and provides standards of accuracy, and in addition, has trouble dealing with models without target systems.

7.2 WE NEED MODEL-SYSTEMS

7.2.1 *Model-target comparisons*

Answering the ER-problem requires providing an account of in virtue of what models represent their targets that explains how they are used to reason about their targets. The surrogative reasoning condition, in Toon's framework, amounts to explaining the relationship between claims we are prescribed to imagine about target systems, and claims that we should import to target systems. The inferential use of models is the inferential relationship between these two types of claims. According to DIRECT VIEW, models represent their targets simply by prescribing us to imagine certain things about them. But imagining that the target has a certain feature does not tell us how the imagined feature relates to what we should take the target to actually have, and so there is no mechanism to transfer model results to the target. Imagining the pendulum bob to be a point mass tells us nothing about which, if any, claims about point masses are also true (or better, which we should *take* to be true) of the real bob.

As far as I can see, there are two ways that Toon could explain the relationship between fictional propositions and imported propositions. The first is to simply take them to be the same. Every proposition prescribed to be imagined, by either a concrete model or a model description, should also be imported directly onto the relevant target system. The second is to introduce a mechanism of transfer that relates fictional and imported propositions. I discuss both of these options in order.

The first option doesn't sit well with scientific practice for two reasons. Firstly, it's not the case that everything that a model prescribes us to imagine gets imported onto its target system, at least by competent model-users. In Chapter 6 I discussed state of the art global climate models that contain variables that are known to describe model-processes with no direct real-world correlates. These provide a case where the model descriptions prescribe us to imagine something about a target system, which explicitly is not imported. In fact, it's plausible that the model users don't even imagine that the target has any process corresponding to what's going on in the model whatsoever, let alone imagine them but withhold imputation.

The second, and more worrying, issue is that, even restricted to fictional propositions that are taken to correspond, in some sense, to something in the target system, the idea that an imagined proposition is just taken to hold directly in the target seems implausible. When we imagine that an inclined plane is frictionless we don't import this onto the inclined plane. Rather, we import something like 'the fiction on the surface of the plane is low enough to be negligible'. And the same applies to countless other instances of scientific modelling; what a model (or model-description) prescribes us to imagine about a target rarely, if ever, corresponds exactly to what a competent model user imports to the target itself. And if this is what Toon has in mind, then under the assumption that the standards of accuracy associated with the account are that what is imagined is true, his account has the unfortunate consequence that the majority, if not all, of scientific models are inaccurate representations, without telling us anything about why they are useful.

An alternative approach is to distinguish between fictional propositions and those the model user should import to the target system. For instance, in some cases it could be the case that when a model prescribes the user to imagine something about the target, then the model user should take this to be 'approximately' true, or 'true enough'. In the case of the frictionless plane model, the model user might import that the claim 'the surface of the plane is frictionless' is approximately true when literally applied to the plane itself. But clearly the details of the transfer mechanism between imagined propositions and what is imported need to be spelt out, and Toon provides little information about how to go about doing this.

In response to these difficulties, Levy (2015) invokes Yablo's (2014) notion of 'partial truth'. The core idea of this view is that a statement is partially true

> 'if it is true when evaluated only relative to a subset of the circumstances that make up its subject matter – the subset corresponding to the relevant content-part' (Levy, 2015, 792).

Levy submits that this will also work for a number of cases of modelling and illustrates it using how the ideal gas model explains Boyle's law, the claim that the pressure of a fixed amount of gas at a constant temperature is inversely proportional to its volume. He writes:

> 'The model is an instruction to regard (imagine) a real world gas as if it had various features (including non-colliding molecules). When it is used to explain real-world Boylean behavior, we are in effect told that because of the specified features, the gas behaves in a Boylian way. This, we know full well, cannot be true as stated, because the gas simply doesn't have all the specified features (in particular, its molecules collide all the time). Here we bring in partial truth: the model (or, more precisely, the derivation of its Boylean behavior) is partially true and partially untrue: true with respect to the role of energy distribution, but false with respect to the role of collisions' (Levy, 2015, 794).

The ideal gas model includes a content part pertaining to molecular energy distribution of gas particles and a part pertaining to the lack of collisions between these particles. It is true with respect to the first part, and false with respect to the second. This means it's partially true, and this suffices for it to explain Boyle's law, an observed proportionality between the pressure and volume of an actual gas to which the model has been applied.

The notion of partial truth is used to augment Direct View in order to explain how we use models to reason about target systems, and provide standards of accuracy. The idea is that when a model prescribes us to imagine something about a target system, what we are to imagine is supposed to be partially true of the system, and if this is the case, then the model is an accurate, or fruitful representation of that system. This certainly seems like a step in the right direction, but a lot hangs on whether the notion of partial truth can do the job for all models. Levy is explicit that

> 'there are other sorts of cases, which may be harder to capture in terms of partial truth. Sometimes, a model treats some factor in a system in an idealized fashion, and this factor modulates the relationship between two other factors (where the latter, let's suppose, are depicted accurately). In this kind of case [...] it is often not obvious what the model says about the relationship between the non-idealized factors, and whether that can be treated as a partial truth (Levy, 2015, 795).

Consider the model of the sand pile introduced in Chapter 1. We can suppose that, according to Direct View, the description prescribes us to imagine, of an actual sand pile, that it sits on a discrete two dimensional lattice; sand grains are cubes that land on cells; and move as governed by the toppling rule. According to Levy, when using this model to represent an actual sand pile the model user should take these claims to be partially true. But this doesn't seem right. The content circumstances that make up the subject matter of the model are actual sand grains and their motions. But no one thinks that sand grains are perfect cubes, or that they move as governed by the defined toppling rule. At best, we might say that what the model invites us to imagine is 'partially true' in the sense that *some* of the claims it invites us to imagine (that sand piles end up in SOC states) are *approximately* true. But it is unclear how to accommodate the idea that some of the claims a model invites us to imagine are approximately true using Yablo's notion of partial truth, at least in cases where the relevant content of the model includes idealised claims that are known to be, strictly speaking, false. The fact sand grains are not perfect cubes, and do not move according to the toppling rule, is a crucial part of the model and cannot be clearly 'partitioned out' in the way that Levy suggests the role of collisions can in the ideal gas model. So, at best, it remains an open question whether the notion of partial truth will be able to provide the appropriate standards of accuracy required by using Direct View as an answer to the ER-problem.

7.2.2 *Target-less models*

The next issue is that not all models have a target system, which is a serious problem for a view that analyses representation in terms of imagining something about a target. Toon is well aware of this issue and calls them 'models without objects' (2012, 76). Some of these are models of discredited entities like the ether and phlogiston, which were initially thought to have a target but then turned out not to have one. But not all models without objects are errors: architectural plans of buildings that are never built or models of experiments that are never carried out fall into the same category.

Toon addresses this problem by drawing another analogy with fiction. He points out that not all novels are like The War of the Worlds, which has an actual object, St Paul's Cathedral, of which we are invited to imagine things about. Passages from Dracula, for instance, 'do not represent any actual, concrete object but are instead about fictional characters' (2012, 54). Models without a target are like passages from Dracula. So the solution to the problem is to separate the two cases neatly. When a model has a target then it represents that target by prescribing imaginings about the target; if a model has no target it prescribes imaginings about a fictional character.

Toon immediately admits that models without targets 'give rise to all the usual problems with fictional characters' (2012, 54). However, he seems to think that this is a problem we can live with because the more important case is the one where models do have a target, and his account offers a neat solution there. So what his account requires is a sharp distinction between how models with targets work compared to how models without targets work. In the former case, they prescribe us to imagine that their targets have certain features; in the latter case, they prescribe us to imagine that some fictional target system has certain features.

This bifurcation of imaginative activities raises questions. The first is whether it squares with scientific practice. In some cases we are mistaken about whether or not a model has a target system: we think that the target exists but then find out that it doesn't (as in the case of ether and phlogiston). But does that make a difference to the imaginative engagement with a phlogiston model of combustion? Even today we can understand and use such models in much the same way as its original protagonists did, and knowing that there is no target seems to make little, if any, difference to our imaginative engagement with the model. Of course the presence or absence of a target matters to many other issues, most notably surrogative reasoning (there is nothing to reason about if there is no target!), but it seems to have little importance for how we imaginatively engage with the scenario presented to us in a model. In other cases it is simply left open whether there is a target when the model is developed. In elementary particle physics, for instance, a scenario is often proposed simply as a suggestion worth considering and only later, when all the details are worked out, the question is asked whether this scenario bears an interesting relation to what happens in nature, and if so what the relation is. So, again, the question of whether there is or isn't a target seems to have little, if any, influence on the imaginative engagement of physicists with scenarios in the research process.

In contrast to Toon, Levy offers a radical solution to the problem of models without targets: there aren't any! He first broadens the notion of a target system, allowing for models that are only loosely connected to targets (2015, 796-979). To this end he appeals to Godfrey-Smith's (2009) notion of 'hub-and-spoke' cases: families of models where some are simple and well understood (which makes them the hub models) and the others are connected to them via conceptual links (spokes) but include all sorts of additional modelling assumptions necessary to accurately represent specific target systems. Godfrey-Smith mentions population growth models as examples. He assumes that something like the Lotka-Volterra model introduced in Chapter 1 is the 'hub' model. And once it's applied to actual populations, which involves introducing additional assumptions about the reproduction and death rates of prey and predators, taking into account multiple different populations competing for resources and so on, then we arrive at a 'spoke' model. According to Levy, we should take the Lotka-Volterra hub model as a model with a 'generalised target', population growth in the abstract.[71] More detailed models of specific target systems, based on the Lokta-Volterra equations, are then the 'spokes' attached the hub model.

If something that looks like a model doesn't meet the requirement of having even a generalised target, then it's not a model at all. Levy mentions structures like Conway's game of life and observes that they are 'bits of mathematics' rather than models (2015, 979). This eliminates the need for fictional characters in the case of target-less models.

Notice that although Levy may be correct that that we do not have to deal with things like Conway's Game of Life when developing an account of modelling, it seems far fetched to make the same claim about Toon's other examples of target-less models. Things like models of ether or phlogiston, or architectural models of things that are never built, strike me as clear cut instances of scientific representation, despite the fact that they don't have target systems. So I don't think that an account that remains silent about them is fully satisfactory.

Toon and Levy differ with respect to how they treat target-less models. The former takes them to work as per Direct View; but they prescribe us to imagine things about fictional systems, rather than any concrete target system. The latter attempts to minimise the instances of target-less models by appealing to the idea of a generalised target. And for cases where this won't work, he is happy to bite the bullet and accept

71 This claim is in tension with my discussion in Chapter 1 where I made it explicit that Volterra was interested in a specific target population, fish in the Adriatic sea, rather than population growth in general. The tension arises because historically the model had a specific target system, but has been superseded by more complex models that account for more and more details about that system. The question why the original, relatively simple, model is still discussed and investigated in population biology, despite the fact that it is no longer used to represent specific systems, is answered, according to Levy, by the fact that it has 'population growth' as a generalised target.

that they are not models at all, and thus does 'not think we need to tailor our account of modeling to accommodate [them]' (2015, 979).

The question, then, is whether either of these approaches is acceptable. As I discuss in the next subsection, one of the primary motivations for adopting Direct View in the first place is its apparent ontological parsimony. Rather than having to invoke abstract objects (fictional, mathematical, or something else entirely), scientific modelling is explained in terms of homely imaginings about actual objects. No ontological inflation is required. But in having to appeal to fictional (in the case of Toon), or generalised (in the case of Levy) target systems, the accounts are no more ontologically parsimonious (at least with respect to ontological types) than the indirect accounts. Both have to invoke at least some non-physical objects of which models are about.[72] So one of the key motivations for Direct View over its competitors is lost.

7.2.3 *The direct view is unmotivated*

My main issue with the direct view is that I don't think it is very well motivated. It is clear that both Toon, and Levy's primary interests are ontological considerations, rather than attempting to address the ER-problem first, and then considering the implications this has for the problem of ontology as a question of secondary importance. Both of them motivate their accounts by rejecting the idea of thinking about models as abstract entities, preferring instead to think about them as concrete hypothetical systems akin to works of fiction. But then, they argue that once we do this, then we should no longer adopt an indirect view of scientific representation, since their preferred (Waltonian) account of fiction does not require invoking fictional-systems that exist independently of the sorts of objects we are prescribed to imagine things about. Toon writes:

> 'So if we are to understand model-systems in the same way that Walton understands fictional characters then it seems we would conclude that there are no model-systems' (2012, 58)

and goes on to argue against Frigg's (2010a) attempt to reconcile a Waltonian view of the ontology of models with an indirect view of how they represent.

Similarly, Levy argues that:

72 In the case of Levy's generalised target systems the targets are things like population growth or complex behaviour, not applied to any particular system, and these are not concrete things like rabbits, stumps, or clock pendulums.

> '[W]e have good reasons to treat models as concrete and to assign a central role to the imagination in model development and analysis. However, identifying models with imaginary entities, even if in the end they are nothing but mere make-believe, is problematic because it is hard to make sense of model-target comparisons on such a view' (2015, 791)

and goes on to argue that by shifting to a direct view of representation instead, the model-target comparisons can be handled by Yablo's partial truth of what models prescribe us to imagine.

My preferred take on the relationship between the problem of scientific representation and the question of the ontological status of models is the other way around. I think we should first attempt to answer the ER-problem, and then look at the ontological implications such answers have. I grant that this may be a matter of taste, but I think that semantic considerations are more tractable that ontological ones. From the discussions in the previous subsections, I hope I have made it clear that the direct account, as it turns out, isn't any more ontologically parsimonious that its competitors, and therefore doesn't answer the ontological questions in a way that gives it an edge over the indirect view. But most importantly, DIRECT VIEW fails to provide an account of the representational relationship between models and their targets. Toon has nothing to say about this question, and although Levy's attempt to invoke the notion of partial truth is a step in the right direction, it fails to deal with distortive models.

I think if we were to focus on semantic considerations first and foremost, leaving ontological questions as secondary, then the path towards the DIRECT VIEW becomes much less attractive. As noted at the beginning of this chapter Weisberg (2007) takes indirectness to be a defining characteristic of scientific modelling, and by abandoning it, Levy and Toon assimilate modelling with other kinds of scientific practice. This misses out an important feature of how scientists use models to represent their target systems. Volterra didn't start out by closely investigating fish in the Adriatic Sea, and imagining certain things of them. He explicitly constructed a scientific model, and then used that to represent the fish. Bak and his collaborators didn't start off by directly imagining things about sand piles, they constructed a two dimensional lattice and investigated its dynamical properties under their toppling rule.[73] Models can be constructed, and investigated, without reference to any target system whatsoever, and DIRECT VIEW, by focusing on ontology first and foremost, makes no room for these types of investigations.

---

73 In fact, it's not clear whether they were interested in sand piles at all. In Bak et al. (1987) they first offer a model consisting of a system of pendulums, and then reinterpret the formalism in terms of sand piles.

This is not to say that I would hold onto the methodological principal of semantics first, ontology second, come what may. If it were the case that models *have* to be thought of as fictional, and that the *only* way to make sense of model-target comparisons in this context is to appeal to the direct view of representation, then so be it.[74] But I remain unconvinced that either of these claims is true. As I discuss in Part iii my own account of scientific representation is compatible with various accounts of the ontological status of models, and even if models are best construed as fictions, this doesn't not preclude us comparing their properties with properties of the target systems. Moreover, my account provides a unified framework in which to think about target-less models that does not require invoking either fictional or generalised targets, and makes sense of the fact that models can be investigated in their own right.

74 However, even Levy (2012, 746) accepts that, at least when it comes to models with actual targets, both the indirect and direct views are compatible with scientific practice.

8

INFERENTIALISM

In this chapter I discuss accounts of scientific representation that analyse representation in terms of the inferential role of scientific models. On the previous accounts discussed, a model's inferential capacity dropped out of whatever it was that was supposed to answer the ER-problem: proposed morphisms or similarity relations between models and their targets, for example. The accounts discussed in this section build the notion of surrogative reasoning directly into the conditions on epistemic representation.

In contrast to the ideas discussed in the previous chapters, no canonical 'inferentialist' positions have emerged in the debate, presumably at least in part because of their considerable novelty. For this reason, this chapter is explicitly structured around the individual authors Suárez (2004, 2015), Suárez and Solé (2006), Contessa (2007), Ducheyne (2012), Hughes (1997), and Frigg (2010a). Their accounts are discussed in the order listed here.

8.1 DEFLATIONISM

Suárez argues that we should adopt a 'deflationary or minimalist attitude and strategy' (2004, 770). I discuss deflationism in some detail below but in order to formulate and discuss Suárez's theory of representation at least a preliminary idea of what is meant by a deflationary attitude is needed. In fact, two different notions of deflationist are in operation in his account. The first is

> 'abandoning the aim of a substantive theory to seek universal necessary and sufficient conditions that are met in each and every concrete real instance of scientific representation ... necessary conditions will certainly be good enough' (2004, 771).

I call the view that an answer to the ER-problem should provide only necessary conditions *n*-deflationism ('*n*' for necessary). The second notion is that we should seek 'no deeper features to representation other than its surface features' (2004, 771) or 'platitudes' (Suárez and Solé, 2006, 40) and that we should deny that an analysis of a concept 'is the kind of analysis that will shed explanatory light on our use of the concept' (Suárez, 2015, 39). I call this position *s*-deflationism ('*s*' for surface level feature). As far as I can tell, Suárez intends his account of representation to be deflationary in both senses. He provides the following answer to the ER-problem.

INFERENTIAISM 1 '[*M*] represents [*T*] only if (i) the representational force of [*M*] points towards [*T*], and (ii) [*M*] allows competent and informed agents to draw specific inferences regarding [*T*] ' (2004, 773).

In keeping with his *n*-deflationism INFERENTIALISM 1 features a material conditional rather than a biconditional and hence provides necessary (but not sufficient) conditions for *M* to represent *T*. Let's see how each of these conditions satisfy *s*-deflationism.

The first condition is designed to make sure that *M* and *T* indeed enter into a representational relationship, and Suárez stresses that representational force is 'necessary for any kind of representation' (2004, 776). But explaining representation in terms of representational force seems to shed little light on the matter as long as no analysis of representational force is offered. Suárez addresses this point by submitting that the first condition can be 'satisfied by mere stipulation of a target for any source' (2004, 771).[75] One might think that this is denotation, a condition that features in the accounts discussed below and in the next part of this thesis, but Suárez stressed that this is not what he intends for two reasons. Firstly, he takes denotation to be a substantive relation between a model and its target, and the introduction of such a relation would violate the requirement of *s*-deflationism (Suárez, 2015, 41). Secondly, *M* can denote *T* only if *T* exists. And thus, including denotation as a necessary condition on scientific representation 'would rule out fictional representation, that is, representation of nonexisting entities' (Suárez, 2004, 772), and 'any adequate account of scientific representation must accommodate representations with fictional or imaginary targets' (Suárez, 2015, 44). The second issue is one that besets other accounts of representation too, in particular similarity and isomorphism accounts. In Chapter 9 I provide a solution that keeps denotation as a condition. So non-existent targets need not necessarily be a reason to ban denotation from a theory of representation. The first reason, however, goes right to the heart of Suárez's account: it makes good on the s-deflationary condition that nothing other than surface features can be included in an account of representation. At a surface level one cannot explicate 'representational force' at all, and any attempt to specify what representational force consists in is a violation of *s*-deflationism.

75 Suárez uses 'source' as I use 'vehicle' to refer to the object (e.g. picture or model) doing the representing

The second necessary condition, that models allow competent and informed agents to draw specific inferences about their targets, is in fact just the surrogative reasoning condition I introduced in Chapter 1, now taken as a necessary condition on scientific representation. The sorts of inferences that models allow are not constrained. Suárez points out that the condition 'does not require that [*M*] allow deductive reasoning and inference; any type of reasoning – inductive, analogical, abductive – is in principle allowed (Suárez, 2004, 773). A problem for this approach is that we are left with no account of how these inferential rules are generated: what is it about models that allows them to licence inferences about their targets, or what leads them to licence some inferences and not others? Contessa makes this point most stridently when he argues that:

> 'On the inferential conception, the user's ability to perform inferences from a vehicle [model] to a target seems to be a brute fact, which has no deeper explanation. This makes the connection between epistemic representation and valid surrogative reasoning needlessly obscure and the performance of valid surrogative inferences an activity as mysterious and unfathomable as soothsaying or divination' (2007, 61).

This seems correct, but Suárez can dismiss this complaint by appeal to *s*-deflationism. Since inferential capacity is supposed to be a surface level feature of scientific representation, we are not supposed to ask for any elucidation about what makes an agent competent and well informed, or how inferences are drawn.

For these reasons Suárez's account is deflationary both in the sense of *n*-deflationism and of *s*-deflationism. His position provides a conception of representation that is cashed out in terms of an inexplicable notion of representational force and of an inexplicable capacity to ground inferences. This is very little indeed. It is the adoption of a deflationary attitude that allows him to block any attempt to further unpack these conditions and so the crucial question is: why should one adopt deflationism? I turn to this question below, but it is useful to first see how the account fares with respect to the additional questions and conditions of adequacy discussed in Chapter 1.

The account provides a neat explanation of the possibility of misrepresentation:

> 'Part (ii) of this conception accounts for inaccuracy since it demands that we correctly draw inferences from the source about the target, but it does not demand that the conclusions of these inferences be all true, nor that all truths about the target may be inferred' (Suárez, 2004, 776).

Models represent their targets only if they license inferences about them. They represent them accurately to the extent that the conclusions of these inferences are true.

With respect to the representational demarcation problem it is clear that the conditions of Inferentialism 1 are met by non-scientific as well as scientific epistemic representations, so, at least as it stands without sufficient conditions, there is no clear way of demarcating between the different kinds of epistemic representation. And given that Suárez illustrates his account with a large range of representations, including diagrams, equations, scientific models, and non-scientific representations such as artistic portraits (like Velázquez's portrait of Innocent X), and that he explicitly states that 'if the inferential conception is right, scientific representation is in several respects very close to iconic modes of representation like painting' (2004, 777) I presume that he agrees with this outcome.

Given the wide variety of types of representation that this account applies to, it's unsurprising that Suárez has little to say about the ontological problem. The only constraint that Inferentialism 1 places on the ontology of models is that '[i]t requires [M] to have the *internal structure* that allows informed agents to correctly draw inferences about [T]' (Suárez, 2004, 774, emphasis added). And relatedly, since the account is supposed to apply to a wide variety of entities, including equations and mathematical structures, the account implies that mathematics is successfully applied in the sciences, but in keeping with the spirit of *s*-deflationism no explanation is offered about how this is possible.

Suárez does not directly address the problem of style, but a minimalist answer emerges from what he says about representation. On the one hand he explicitly acknowledges that many different kinds of inferences are allowed by the second condition in Inferentialism 1. In the passage quoted above he mentions inductive, analogical and abductive inferences. This could be interpreted as the beginning of classification of representational styles. On the other hand, Suárez remains silent about what these kinds are and about how they can be analysed. This is unsurprising because spelling out what these inferences are, and what features of the model ground them, would amount to giving a substantial account, which is something Suárez wants to avoid.

I now return to the question about the motivation for deflationism. As I have discussed, a commitment to deflationism about the concept is central to Suárez's approach to scientific representation. But deflationism comes in different guises, which Suárez illustrates by analogy with deflationism with respect to truth. Suárez (2015) distinguishes between the 'redundancy' theory (associated with Frank Ramsey, also referred to as the 'no theory' view), 'abstract minimalism' (associated with Crispin Wright) and the 'use theory' (associated with Paul Horwich).[76] What all three are

76 Here I concentrate on what Suárez says about these accounts, rather than their original presentations.

claimed to have in common is that they accept the disquotational schema – i.e. instances of the form: 'φ' is true if and only if φ. Moreover they 'either do not provide an analysis in terms of necessary and sufficient conditions, or if they do provide such conditions, they claim them to have no explanatory purchase' (Suárez, 2015, 37).

Starting with the redundancy theory of truth (abstract minimalism and the use theory are discussed in the next section) Suárez claims that it is characterised by the idea that 'the terms 'truth' and 'falsity' do not admit a theoretical elucidation or analysis. But that, since they can be eliminated in principle – if not in practice – by disquotation, they do not in fact require such an analysis' (2015, 39). So, as Suárez characterises the position, the redundancy theory denies that any necessary and sufficient conditions for application of the truth predicate case be given. He argues that:

> 'the generalization of this 'no-theory theory' for any given putative concept X is the thought that X neither possesses nor requires necessary and sufficient conditions because it is not in fact a 'genuine', explanatory or substantive concept' (2015, 39).

This is supposed to motivate *n*-deflationism.[77]

This approach faces a number of challenges. First, the argument is based on the premise that if the no theory view is good for truth it must be good for representation. This premise is assumed tacitly. There is, however, a question whether the analogy between truth and representation is sufficiently robust to justify subjecting them to the same theoretical treatment. Surprisingly, Suárez offers little by way of explicit argument in favour of the no theory account of (or indeed any deflationary approach to) scientific representation. In fact, the natural analogue of the linguistic notion of truth is accurate epistemic representation, rather than epistemic representation itself, which may be more appropriately compared with linguistic meaning. Second, the argument insinuates that no theory view is the correct analysis of truth. This, however, is far from an established fact. Different positions are available in the debate and whether deflationism (or any specific version of it) is superior to other proposals remains a matter of controversy (see, for instance, Künne (2003)). But as long as it remains unclear whether the no theory account about truth is a superior position, it's hard to see how one can muster support for the no theory approach to scientific representation by appealing to deflationism about truth.

More significantly, a position that allows only necessary conditions on epistemic representation faces a serious problem. While such an account allows us to rule out

77 However, even here, it remains unclear why the approach would motivate giving even necessary conditions.

certain scenarios as instances of epistemic representation (for example a proper name doesn't allow for a competent and well informed language user to draw any specific inferences about its bearer and Callender and Cohen's salt-shaker discussed in Chapter 2 doesn't allow a user to draw any specific inferences about Madagascar), the lack of sufficient conditions doesn't allow us to rule in any scenario as an instance of epistemic representation. So on the basis of Inferentialism 1 we are never in a position to assert that a particular model actually is a scientific representation, which is an unsatisfactory situation.

8.2 inference, representation, and levels of abstraction

The other two deflationary positions in the debate over truth Suárez draws on are abstract minimalism and the use theory. Suárez characterises the use theory as being based on the idea that 'truth is nominally a property, although not a substantive or explanatory one, which is essentially defined by the platitudes of its use of the predicate in practice' (2015, 40). Abstract minimalism is presented as the view that while truth is 'legitimately a property, which is abstractly characterized by the platitudes, it is a property that cannot explain anything, in particular it fails to explain the norms that govern its very use in practice' (2015, 40). Both positions imply that necessary and sufficient conditions for truth can be given. But on either account, such conditions only capture non-explanatory surface features. This motivates *s*-deflationism.

Since *s*-deflationism explicitly allows for necessary and sufficient conditions, Inferentialism 1 can be extended to provide necessary and sufficient conditions on scientific representation:[78]

inferentialism 2 : A model *M* represents a target *T* if and only if (i) the representational force of *M* points towards *T*, and (ii) *M* allows competent and informed agents to draw specific inferences regarding *T*.

Depending on how conditions (i) and (ii) are interpreted, we arrive at an analog of either the use theory, or abstract minimalism.

If one takes the conditions to refer to 'features of activates within a normative practise, [that] do not stand for relations between sources and target' (Suárez, 2015, 46), then we arrive at a 'use-based' account of scientific representation. In order to understand a particular instance of a model *M* representing a target *T* we have to understand

78 Which also seems to be in line with Suárez and Solé (2006, 41) who provide a formulation of inferentialism with a biconditional

how scientists go about establishing that M's representational force points towards T, and the inferential rules, and particular inferences from M to T, they use and make.[79]

Plausibly, such a focus on practice amounts to looking at the inferential rules employed in each instance, or type of instance, of scientific representation. This, however, raises a question about the status of any such analysis vis-á-vis the general theory of representation as given in Inferentialism 2. There seem to be two options. The first is to affirm Inferentialism 2's status as an exhaustive theory of representation. This, however, would imply that any analysis of the workings of a particular model would fall outside the scope of a theory of representation because any attempt to address Contessa's objection would push the investigation outside the territory delineated by *s*-deflationism. Such an approach seems to be overly purist.

The second option is to understand Inferentialism 2 as providing abstract conditions that require concretisation in each instance of scientific representation. Studying the concrete realisations of the abstract conditions is then an integral part of the theory. This approach seems plausible, but it remains unclear whether this approach is truly 'deflationary'. Thus understood, the view becomes indistinguishable from a theory that accepts the surrogative reasoning condition, along with a condition that demands models are about their targets in a way that targets are not about their models, and then analyses these conditions in a pluralist spirit, that is, under the assumption that these conditions can have different concrete realisers in different contexts. If these realisers are substantial properties of the activities of the normative practice of science, then each specific instance of scientific representation will not be analysed in a deflationary manner, even if the 'concept' in general is analysed solely in terms of platitudes.

One might worry that these responses unfairly stack the deck against inferentialism and point out that different inferential practises can be studied within the inferentialist framework. One way of making good on this idea would be to submit that the inferences from models to their targets should be taken as conceptually basic, denying that they need to be explained; in particular, denying that they need to be grounded by any (possibly varying) relation(s) that might hold between models and their targets. Such an approach is inspired by Brandom's inferentialism in the philosophy of language where the central idea is to reverse the order of explanation from representational notions – like truth and reference and so on – to inferential notions – such as the validity of argument (Brandom, 1994, 2000). Instead, we are urged to begin from the inferential role of sentences (or propositions, or concepts, and so on) – that is the role that they

79 It's worth reiterating that Suarez provides no argument for why we should adopt these positions, restricting himself to pointing out the analogy between deflationary accounts of truth and deflationary accounts of scientific representation.

play in providing reasons for other sentences (or propositions etc.), and having such reasons provided for them – and from this reconstruct their representational aspects.

Such an approach is developed by de Donato Rodríguez and Zamora Bonilla (2009) and seems like a fruitful avenue of research. The idea being that each investigation into the inferential rules utilised in each instance, or type of instance, of scientific representation will likely be a substantial (possibly sociological or anthropological) project.[80] And there is a tension between this philosophical approach and Suárez's *s*-deflationary aspirations, at least if these are taken to require that nothing substantial can be said about scientific representation in each instance, as well as in general. One way to address this issue is to note that once the inferential capacities of models are taken as conceptually basic, and that investigating these requires investigating the practice of scientists using them in particular cases, the notion of 'representation' has little role to play. Rather, as philosophers we are closer to sociologists or anthropologists, whose investigations should focus on the scientific practice of using models, rather than addressing things like the ER-problem or the other questions discussed in Chapter 1. Such an approach could be in line with Price's (2011) project of eliminating the concept of representation from our philosophical agenda. His central idea is that to adopt a global expressionism: we shouldn't think that any 'statement' functions by 'matching' some fact in the world, their primary function is not to describe. Applying this insight for scientific models: we shouldn't think that they function by 'matching' (read: 'representing') some target system in the world. If we are to properly understand how they work we should instead investigate how, as part of the natural world themselves, scientists go about using models.

Investigating scientific models within the broader philosophical frameworks associated with the likes of Brandom and Price goes beyond the scope of this thesis. However, I think that such an investigation is worthwhile for at least two reasons. Firstly, Brandom's approach of reversing the notions of inference and representation in terms of their conceptual basicness stands or falls with how fruitful such a reversal is. I grant that good philosophy cannot be done without taking something as primitive, but the basis for this decision should be sensitive to the outcome of doing so. If, by taking the notion of inference as conceptually basic we arrive at a fruitful picture of how we understand the world, then the approach stands as a worthwhile alternative to the representational alternative (which I take to have proven its worth in the development of things areas like semantics and formal logic, and, as I discuss later in the thesis, understanding how scientific models represent in the way I advise furthers our understand of the models themselves). Investigating scientific models – i.e. one of our

80 By 'substantial', I mean will uncover inferential practises that are described in a level of detail that go beyond Suárez's platitudes.

primary ways of investigating the world – from this perspective would seem like an interesting test-bed for Brandom's project.

Secondly, Brandom and Price's criticisms of the representational approach are targeted primarily at linguistic representation. Price in particular criticises how philosophers have utilised the concept of the 'truth' of a proposition, rather than understanding how speakers use the proposition in discursive contexts. Even if their criticisms bite there, it remains to be seen whether they carry over into the context of model-based representation. If they do, then the frameworks gain further support. If not, then the representational alternative should be preferred.

Moving onto Suárez's third deflationary analogy, if the conditions in INFERENTIALISM 2 are taken to be abstract platitudes then we arrive at an abstract minimalism. Although INFERENTIALISM 2 defines the concept of scientific representation, the definition does not suffice to explain the use of any particular instance of scientific representation for:

> 'on the abstract minimalism here considered, to apply this notion to any given concrete case of representation requires that some additional relation obtains between [*M*] and [*T*], or a property of [*M*] or [*T*], or some other application condition' (Suárez 2015, 48 *cf.* Suárez and Solé 2006).

Hence, according to this approach, representational force and inferential capacity are taken to be abstract platitudes that suffice to define the concept of scientific representation. However, because of their level of generality, they fail to explain any particular instance of it. To do this requires reference to additional features that vary from case to case. These other conditions can be 'isomorphism or similarity' and they 'would need to obtain in each concrete case of representation' (Suárez 2015, 45 *cf.* 2004, 773, 776 and Suárez and Solé 2006, 45). These extra conditions are called the 'means' of representation – the relations that scientists exploit in order to draw inferences about targets from their models – and are to be distinguished from conditions (i) and (ii) – the 'constituents' of representation – that define the concept (Suárez 2003, 230; 2010, 93-94; 2015, 46; Suárez and Solé 2006, 43). We are told that the means cannot be reduced to the constituents but that 'all representational means (such as isomorphism and similarity) are concrete instantiations, or realisations, of one of the basic platitudes that constitute representation' (Suárez and Solé, 2006, 43), and that 'there can be no application of representation without the simultaneous instantiation of a particular set of properties of [*M*] and [*T*], and their relation' (Suárez and Solé, 2006, 44).

Such an approach amounts to using conditions (i) and (ii) to answer the ER-problem, but again with the caveat that they are abstract conditions that require concretisation in each instance of scientific representation.[81] In this sense it is immune to Contessa's objection about the 'mysterious' capacity that models have to licence about their targets. They do so in virtue of more concrete relations that hold between models and their targets, albeit relations that vary from case to case. The key question facing this account is to fill in the details about what sort of relations concretise the abstract conditions. But again such an approach faces a similar problem to the above. Even if *s*-deflationism applies to scientific representation in general, an investigation into each specific instance will involve uncovering substantial relations that hold between models and their targets, which again seems to conflict with Suárez's *s*-deflationary approach.

I take these observations to indicate that Suárez's project could be developed into a substantial account of scientific representation. Depending on one's broader philosophical commitments, this could be done by adopting an inferentialist position – taking the notion of inference as conceptually basic, investigating the inferences associated with normative scientific practise, and then reconstructing the notion of representation from them (or, following Price, abandoning the notion of representation altogether)[82] – or alternatively, investigating more concrete relations that hold between models and their targets in various instances. Both of these approaches would be in line with Inferentialism 2, but they would entail that more can be said about scientific representation, at least in its instances, than I suspect Suárez himself would be happy with.

8.3 reinflating inferentialism

In response to these difficulties Contessa claims that 'it is not clear why we should adopt a deflationary attitude from the start' (2007, 50) and provides a 'interpretational account' of scientific representation that is still, at least to some extent, inspired by Suárez's account, but without being deflationary. Contessa claims:

> '[t]he main difference between [the] interpretational conception [...] and Suárez's inferential conception is that the interpretational account is a substantial account – interpretation is not just a "symptom" of representation; it is what makes something an epistemic representation of a something else' (2007, 48).

81 Abstract minimalism differs from the use-based account in virtue of where these concrete conditions are found. In the former it is taken to be properties of models and their targets, in the latter is is taken to be properties of the practise of using scientific models.

82 See Price (2011, Chapter 14) for a discussion of the differences between these projects in the context of the philosophy of language.

Contessa introduces the notion of an interpretation of a model, in terms of its target system, as a necessary and sufficient condition on epistemic representation:

INTERPRETATION: 'A [model M] is an epistemic representation of a certain target [T] (for a certain user) if and only if the user adopts an interpretation of the [M] in terms of [T]' (2007, 57 *cf.* 2011, 126-127).

A loose way of characterising what it means to adopt an interpretation of a model in terms of target is that the user 'takes facts about the vehicle to stand for (putative) facts about the target' (2007, 57). Contessa offers a precise formulation of a specific type of interpretation – a so-called 'analytic interpretation' – that first requires that the model user identify, in both the vehicle and the target, relevant sets of objects, relations (including properties) and relations:

> 'An analytic interpretation of a vehicle in terms of the target identifies a (nonempty) set of relevant objects in the vehicle ($\Omega^V = \{o_1^V, ..., o_n^V\}$) and a (nonempty) set of relevant objects in the target ($\Omega^T = \{o_1^T, ..., o_n^T\}$), a (possibly empty) set of relevant properties of and relations among objects in the vehicle ($P^V = \{^nR_1^V, ..., ^nR_m^V\}$, where ^nR denotes an n-ary relation and properties are construed as 1-ary relations) and a set of relevant properties and relations among objects in the target ($P^T = \{^nR_1^T, ..., ^nR_m^T\}$), and a set of relevant functions from $(\Omega^V)^n$ – that is, the Cartesian product of Ω^V by itself n times – to Ω^V ($\Phi^V = \{^nF_1^V, ..., ^nF_m^V\}$, where ^nF denotes an n-ary function) and a set of relevant functions from $(\Omega^T)^n$ to Ω^T ($\Phi^T = \{^nF_1^T, ..., ^nF_m^T\}$)' (Contessa, 2007, 57).

Notice that this explicitly does not require that the relevant sets exhaust the objects, relations, or functions in either the vehicle or the target. The definition of an analytic interpretation is then given as follows:

> 'A user adopts an *analytic interpretation* of a vehicle in terms of a target if and only if:
>
> 1. The user takes the vehicle to denote the target,
> 2. The user takes every object in Ω^V to denote one and only one object in Ω^T and every object in Ω^T to be denoted by one and only one object in Ω^V,
> 3. The user takes every n-ary relation in P^V to denote one and only one relevant n-ary relation in P^T and every n-ary relation in P^T to be denoted by one and only one n-ary relation in P^V,

4. They take every n-ary function in Φ^V to denote one and only one n-ary function in Φ^T and every n-ary function in Φ^T to be denoted by one and only one n-ary function in Φ^V' (Contessa, 2007, 58).

Analytic interpretations licence inferences from models to their targets in the following manner:

'An analytic interpretation underlies the following set of inference rules:

RULE 1: If o_i^V denotes o_i^T according to the interpretation adopted by the user, it is valid for the user to infer that o_i^T is in the target if and only if o_i^V is in the vehicle,

RULE 2: If o_i^V denotes o_i^T, ..., o_n^V denotes o_n^T, and ${}^nR_k^V$ denotes ${}^nR_k^T$ according to the interpretation adopted by the user, it is valid for the user to infer that the relation ${}^nR_k^T$ holds among $o_1^T, ..., o_n^T$ if and only if ${}^nR_k^V$ holds among $o_1^V, ..., o_n^V$,

RULE 3: If, according to the interpretation adopted by the user, o_i^V denotes o_i^T, ..., o_n^V denotes o_n^T, and ${}^nF_k^V$ denotes ${}^nF_k^T$, it is valid for the user to infer that the value of the function ${}^nF_k^T$ for the arguments $o_1^T, ..., o_n^T$ is o_i^T if and only if the value of the function ${}^nF_k^V$ is o_i^V for the arguments $o_1^V, ..., o_n^V$' (Contessa, 2007, 61).

So according to the interpretational account, a vehicle represents its target in virtue of a model user adopting an analytic interpretation of the former in terms of the latter. This proceeds by the model user identifying sets of relevant objects, relations, and functions in both the vehicle and the target, and setting up one-to-one directed denotation relations from elements in the vehicle's relevant sets to elements in the target's relevant sets. With this established the model user can infer from facts about the vehicle – things like a particular n-tuple of objects being in the extension of a relation – to (purported) facts about the target – the n-tuple of target objects that are the denotatum of the n-tuple of vehicle objects are in the extension of the a target relation that is the denotatum of the vehicle relation.[83]

INTERPRETATION is a non-deflationary account of scientific representation; most (but not necessarily all) instances of scientific representation involve a model user adopting

83 At first sight Contessa's interpretation may appear to be equivalent to setting up an isomorphism between model and target. This impression is correct in as far as an interpretation requires that there be a one-to-one correspondence between relevant elements and relations in the model and the target. However, unlike the isomorphism view, Contessa's interpretations are not committed to models being structures. Relations (and functions) can be interpreted as full fledged relations (and functions) rather than purely extensionally specified sets of n-tuples, and Contessa restricts himself to 'relevant' sets of each.

an analytic interpretation towards a target. The capacity for surrogative reasoning is then seen as a symptom of the more fundamental notion of a model user adopting an interpretation of a model in terms of its target. For this reason the adoption of an analytical interpretation is a substantial sufficient condition on establishing the representational relationship. Contessa focuses on the sufficiency of analytic interpretations rather than their necessity and adds that he does 'not mean to imply that all interpretation of vehicles [models] in terms of the target are necessarily analytic. Epistemic representations whose standard interpretations are not analytic are at least conceivable' (2007, 58). Even with this in mind, it is clear that he intends that there be some interpretation as a necessary condition on epistemic representation.

With the account presented, it's now time to see how well the account fares with respect to the questions and conditions of adequacy laid out previously. Modulo the caveat about non-analytical interpretations, INTERPRETATION provides necessary and sufficient conditions on epistemic representation and hence answers the ER-problem. Furthermore, unlike the naïve similarity and structuralist accounts, it does so in a way that is not symmetric: interpreting a model in terms of a target does not entail interpreting a target in terms of a model. Contessa does not comment on the applicability of mathematics but since his account shares the structuralist emphasis on relations and one-to-one model-target correspondence, Contessa can appeal to the same sort of account of the applicability of mathematics as was appealed to there. With respect to the demarcation problem, Contessa is explicit that '[p]ortraits, photographs, maps, graphs, and a large number of other representational devices' also satisfy [INTERPRETATION] (2007, 54). Since nothing in the notion of an interpretation seems restricted to scientific models, it is plausible to regard INTERPRETATION as a universal theory of epistemic representation.[84] As such, INTERPRETATION seems to deny the existence of a substantial distinction between scientific and non-scientific epistemic representations (at least in terms of their representational properties). It remains unclear how INTERPRETATION addresses the problem of style. With respect to the question of ontology, INTERPRETATION itself places few constraints on what scientific models are, ontologically speaking. All it requires is that they consist of objects, properties, relations, and functions. Contessa (2010) himself takes (at least some models) to be abstract objects, and thus one would assume that the relevant features of such models that feature in their analytic interpretations are themselves abstract.

Given Contessa's adherence to the distinction between representation and accurate representation one might think that INTERPRETATION would be designed to handle the possibility of misrepresentation. There are, however, issues facing the account in this

84 A conclusion that is also supported by the fact that Contessa uses the example of the London underground map to motivate his account.

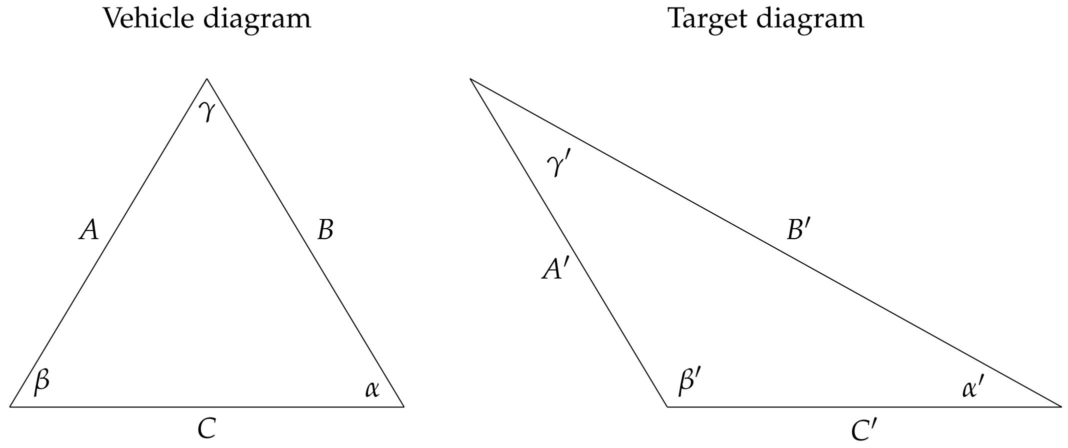

Figure 10: An equilateral triangle and an obtuse triangle (*cf.* Shech, 2014, 13)

regard that deserve to be briefly sketched. The way that Contessa takes INFERENTIALISM to deal with the possibility of misrepresentation is by distinguishing between *valid* and *sound* inferences. Valid inferences are instances of Rules 1,2 and 3 above, but they needn't have true conclusions; they needn't be sound. It's an interesting question, though, how such a inference, if licenced by an analytic interpretation, could have a false conclusion. In fact, Shech (2014) argues that this cannot be the case. He does so by first inviting us to consider the instance of epistemic interpretation displayed in Figure 10.

He claims that any analytic interpretation of the vehicle in terms of the target will guarantee that all valid inferences licenced by such an interpretation will be sound. He argues for this as follows. Suppose we identify A, B, C, α, β, and γ as the relevant objects in the vehicle, A', B', C', α', β', and γ' as the relevant objects in the target, $\alpha = \beta = \gamma$ and $A = B = C$ as the relevant relations in the vehicle, and $\beta' > \gamma' > \alpha'$ and $B' > A' > C'$ as the relevant relations in the target. And we notice that $\alpha = \beta = \gamma$ is a relation in the vehicle. What inference(s) does this licence? Well this turns on what analytic interpretation we adopt, but Shech seems to assume that the only possibilities available that we let A denote A', α denote α', and so on. Further suppose that we identify $\alpha = \beta = \gamma$ as a relation in the vehicle. The question, then, is what relation in the target should $\alpha = \beta = \gamma$ be taken to denote? Shech claims that the possible denotable relations are $B' > C' > A'$ and $\beta > \gamma > \alpha$. But in both cases, any inference drawn from such an interpretation will be sound, thus seemingly ruling out an inference that is valid but unsound, and thus ruling out the possibility of misrepresentation.

But this is an artifact of the limited analytic interpretations he considers. In contrast, consider the following analytic interpretation. Let A denote A', B denote B', and C denote γ', α denote α', β denote β' and γ denote C'. Let the property *angle* in the vehicle (whose extension $R_1^V = \{\alpha, \beta, \gamma\}$) denote the property *angle* in the target (whose extension $R_1^T = \{\alpha', \beta', \gamma'\}$), and the property *line* in the vehicle (whose extension $R_2^V = \{A, B, C\}$) denote the property *line* in the target (whose extension $R_2^T = \{A', B', C'\}$).[85] Then, from the fact that C instantiates *line* (or extensionally, $C \in R_2^V$), and the preceding interpretation, the user can validly infer that γ' instantiates *line* (or extensionally, $\gamma' \in R_2^T$). This is false, so we have a valid but unsound inference licenced by the above analytic interpretation. Thus, Shech's objection turns on a mistaken reading of Contessa's account. Some analytic interpretations will licence more truths than others. But that there exist some that don't licence falsehoods does not entail that the interpretational account of epistemic representation cannot deal with misrepresentation.[86]

This shows that the interpretational account can deal with misrepresentation where properties that are instantiated somewhere in the target system are represented as being instantiated elsewhere in the system. But this does not exhaust all instances of misrepresentation, and Shech is right to point out that Contessa's account, as presented, cannot deal with inferences where the model represents the target as having properties that it does not have, i.e. that are not instantiated anywhere in the target system. For these properties cannot feature in the relevant set of relations in the target to be denoted. For example, it cannot be the case that an ideal pendulum represents the string of the actual pendulum as having no mass. Surprisingly, this is the very example that Contessa uses to illustrate his interpretational account. He claims that:

> '[O]ne does not need to believe that the string from which a certain pendulum hangs is massless in order to adopt an interpretation of the ideal pendulum according to which the string is massless. The knowledgeable user knows perfectly well that, since no real string is massless, the inference, though valid, is not sound' (Contessa, 2007, 59-60).

Now it is clear that the property *masslessness* does not exist in the target system. But this means that the property cannot be taken to be a member of P^T, the set of relevant properties the user identifies in the target system when setting up her analytic

85 Here I distinguish between properties and their extensions, but this is not necessary for my argument to go through.

86 Consider for example the analytic interpretation exhausted by letting A denote A'. The only inference this will licence is that A' is in the target, which is true. Such an interpretation is accurate in the sense that the conclusion of every inference it licences is true, but this isn't problematic for the interpretational account, it's just not a very informative instance of epistemic representation.

interpretation, since as stated, these properties are properties present in the system. So since Contessa is explicit that the inference from the masslessness of the pendulum string in the model to the masslessness of the actual string is a valid, but unsound inference, his definition of an analytic interpretation requires amendment.

Shech (2014, 14) suggests, and I agree with him, that we can interpret Contessa's definition of an analytic interpretation more liberally. When identifying the relevant sets of objects, relations, and functions in the target system, we do not have to restrict ourselves to those that are actually exist in the system. Instead, we can appeal to the idea that objects, relations, and functions are *purportedly* in the system: hypothesised to exist in the system even if they 'are not semantically true' (Shech, 2014, 14) of the target.

With this in mind, we can now make sense of the claim that the inference from the masslessness of the model string to the masslessness of the actual string in the following way. The model user identifies a string in the model and collects it in Ω^V and a string in the target and collects it in Ω^T. Further, the user identifies *masslessness* as being in the model, and collects it in P^V, and identifies *masslessness* as being *purportedly* in the target, and collects it in P^T.[87] Then, from the fact that the model string instantiates *masslessness*, and the fact that the analytic interpretation adopted means taking the model string to denote the actual string, and *masslessness* in the model to denote *masslessness* in the target, the model user can validly infer that the string in the target instantiates *masslessness*. And this is a valid, but unsound inference.

Shech rejects this approach since 'by the same account we could adopt an interpretation of a target in terms of a vehicle in which both are identical, and still extract inferences that are valid but not sound' (Shech, 2014, 14).[88] To see why this is the case he utilises the idea of using the vehicle diagram above as an epistemic representation of itself. This time let the vehicle (the equilateral triangle) be an epistemic representation of itself, and adopt an analytic interpretation such that $A, B, C, \alpha, \beta, \gamma$ all denote themselves, but let the equality between angles (in the vehicle) denote a purported inequality between angles (in the target). From this analytic interpretation we can infer from the fact that $\beta = \gamma = \alpha$ that $\beta > \gamma > \alpha$. Shech concludes:

> 'This is a valid inference that is not sound. But since the vehicle and the target are identical this is not a case of misrepresentation. How can a vehicle misrepresent a target if the two are identical?!' (2014, fn.23).

87 I do not suggest that these objects and properties exhaust the relevant ones of either the vehicle or the target.

88 As a terminological issue, Contessa presents his account as interpreting the vehicle in terms of the target, not the target in terms of a vehicle, I assume this is a typographical mistake on Shech's part.

One person's *modus tollens* is another's *modus ponens*. Why should we think that if an object represents itself, it must be an accurate representation of itself? As a counterexample consider the sentence φ = 'this sentence is written in italics'. φ represents itself, but does so inaccurately. It asserts, of itself, that it is written in italics, and this is not the case.[89] For another example we can consider an appropriately altered example of Goodman's.

First consider the picture of Obama (Figure 11). It's standardly taken to represent Obama (and represents him as wearing a dark suit and tie against a light shirt). Now consider the coloured inverted picture of Obama (Figure 12). Assume that it is known that this picture is colour inverted. Then I take it that it still represents Obama as wearing a dark suit and tie against a light shirt: a competent interpreter of the picture would be able to infer this if they knew how the colour inversion mapped shades to shades. Both Figure 11 and Figure 12 are equally accurate representations of Obama. The sort of interpretation that we apply to the former (where colours are mapped to themselves) are more entrenched than the sort we would have to apply to the latter once we were aware of the colour inversion software, but increased entrenchment doesn't make the representation any more accurate. The two pictures, appropriately interpreted, licence the exact same inferences.

Now mix things up a bit. First suppose that Figure 11 was taken to represent itself, rather than Obama. Hopefully this is fairly straightforward. Under a standard interpretation (where colours are mapped to themselves), the picture would serve as an excellent representation of itself. However, suppose that the same picture were taken to represent itself, but an onlooker thought that it was doing so under a colour inversion interpretation. In this case, the model user would infer that the picture itself had been colour inverted and would infer entirely different facts about that which it represents (in fact in such a situation it would be correct to say that Figure 11 is a more accurate representation of Figure 12, than it is of itself). So we have another example of something that represents itself, but does so inaccurately, and I take it as an attribute of Contessa's account that it explains how this sort of thing can come about.

So it seems like Shech's objections fail. Appropriately amended, Contessa's account does allow for the possibility of misrepresentation. The problem with Contessa's account is not that it is false, but rather than it leaves things unsaid. For example, Inter-

89 Although if 'is written in italics' were taken to denote *was typeset in LaTeX* then φ would be an accurate representation of itself. This would amount to adopting an alternative analytic interpretation, and again, that one allow for only sound inferences is not a problem for the interpretational account.

89 Available here `https://commons.wikimedia.org/wiki/File:Official_portrait_of_Barack_Obama.jpg#/media/File:Official_portrait_of_Barack_Obama.jpg`.

91 The image was generated using Apple Preview.

Figure 11: Obama (public domain)[90]

Figure 12: Obama Inverted[91]

PRETATION has nothing to say about target-less models, and Contessa tells us nothing about either how the relevant systems (models and targets) are cleaved into objects, relations and functions, nor how these are associated with one another. Moreover, his demand on one-to-one denotation relations seems rather strict, even when restricted to 'relevant' sets of these things. Other interpretations could be appealed to here, but Contessa provides no insight about what these look like.

A variant on Contessa's account is offered by Ducheyne. The details of the account, which I won't state precisely here for want of space, can be found in Ducheyne (2012, 83-86). The central idea is that each relevant relation (or function) specified in the interpretation, holds precisely in the model, and corresponds to the same relation (or function) that holds only approximately (with respect to a given purpose) in the target. For example, the low mass of an actual pendulum's string approximates the masslessness of the string in the model. The one-to-one correspondence between (relevant) objects, relations, and functions in the model and target is retained, but the notion of a user taking relations in the model to denote relations in the target, is replaced with the idea that the relations in the target are approximations of the ones they correspond to. Ducheyne calls this the PRAGMATIC LIMITING CASE account of scientific representation (the pragmatic element comes from the fact that the level of approximation required is determined by the purpose of the model user).

However, if this account is to be an improvement on Contessa's, then more needs to be said about how a target relation can 'approximate' a model relation. Ducheyne relies on the fact that relations are such that 'we can determine *the extent to which* [they hold] empirically' (2012, 83, emphasis added). This suggests that he has quantifiable relations in mind, and that what it means for a relation R (or function) in the target to approximate a relation R' (or function) in the model is a matter of comparing numerical values, where a model user's purpose determines how close they must be if the former is to count as an approximation of the latter. But whether this exhausts the ways in which relations can be approximations remains unclear. Hendry (1998), Laymon (1990), Liu (1999), Norton (2012), and Ramsey (2006) among others, offer discussions of different kinds of idealisations and approximations, and Ducheyne would have to make it plausible that all these can be accommodated in his account.

More importantly, Ducheyne's account has problems dealing with misrepresentations. Although it is designed to capture models that misrepresent by being approximations of their targets, it remains unclear how it deals with models that are outright mistaken. For example, it seems a stretch to say that Thomson's model of the atom is an approximation of what the Schrödinger's model tells us about atoms, and it seems unlikely that there is a useful sense in which the relations that hold between electrons in Thomson's model 'approximate' those that hold in reality. Similarly, I take it that no one would take a Ptolemaic model of the solar system to be an 'approximation' of the actual solar system, and yet it still seems that such a model would genuinely count as a scientific representation (albeit a poor one). But this does not mean that they are not scientific representations of the atom or solar system; it's just that they are incorrect ones. It does not seem to be the case that all cases of scientific misrepresentation are instances where the model is an approximation of the target (or even conversely, it is not clear whether all instances of approximation need to be considered cases of 'misrepresentation' in the sense that they licence falsehoods about their targets). So it seems as though Contessa's account fares better as it stands, where the relations and functions of models can be associated as conventionally as the model user likes with relations and functions (purportedly) in the target system, without Ducheyne's added constraint that they be related via approximation.

8.4 DDI

The penultimate account I want to discuss in this chapter is Hughes' Denotation, Demonstration, and Interpretation account of scientific representation (1997; *cf.* Hughes

2010, Chapter 5). This account inspired both the inferential (see Suárez 2004, 770; 2015) and interpretational accounts (Contessa, 2011, 126) discussed above.

Quoting directly from Goodman (1976, 5), Hughes takes a model of a physical system to 'be a symbol for it, stand for it, refer to it' (Hughes, 1997, 330). Presumably the idea is that a model denotes its target in the same way that a proper name denotes its bearer, or, stretching the notion of denotation slightly, a predicate denotes elements in its extension.[92] This is the first 'D' in 'DDI'. What makes models epistemic representations and thereby distinguishes them from proper names, are the demonstration and interpretation conditions.

The demonstration condition, the second 'D' in 'DDI', relies on a model being a 'secondary subject that has, so to speak, a life of its own. In other words, [a] representation has an internal dynamic whose effects we can examine' (1997, 331).[93] The two examples offered by Hughes are both models of what happens when light is passed through two nearby slits. One model is mathematical where the internal dynamics are 'supplied by the deductive, resources of the mathematics they employ' (1997, 331), the other is a physical ripple chamber where they are supplied by 'the natural processes involved in the propagation of water waves' (1997, 332).

Such demonstrations, on either mathematical models or physical models, are still primarily about the models themselves. The final aspect of Hughes' account – the 'I' in 'DDI' – is interpretation of what has been demonstrated in the model in terms of the target system. This yields the predictions of the model (1997, 333). Unfortunately Hughes has little to say about what it means to interpret a result of a demonstration on a model in terms of its target system, and so one has to retreat to an intuitive (and unanalysed) notion of carrying over results from models to targets. Now Hughes is explicit that he is not attempting to answer the ER-problem, and that he does not even offer denotation, demonstration and interpretation as individually necessary and jointly sufficient conditions for scientific representation. He prefers the more 'modest suggestion that, if we examine a theoretical model with these three activities in mind, we shall achieve some insight into the kind of representation that it provides' (1997, 339).

I'm not sure how to interpret Hughes' position in light of this. On one reading, he can be seen as describing how we use models. As such, DDI functions as a diachronic account of what a model user does when using a model in an attempt to learn about a target system. We first stipulate that the model stands for the target, then prove what we want to know, and finally 'transfer' the results obtained in the model back

92 Hughes (1997, 330) notes that there is an additional complication when the model has multiple targets but this is not specific to the DDI account and is discussed in more detail in Part iii.

93 That models have an 'internal dynamic' is all that Hughes has to say about the problem of ontology.

to the target. Details aside, this picture seems by and large correct. The problem with the DDI account is that it does not explain why and how this is possible. Under what conditions is it true that the model denotes the target? What kinds of things are models that they allow for demonstrations? How does interpretation work; that is, how can results obtained in the model be transferred to the target? These are questions an account of scientific representation has to address, but which are left unanswered by the DDI account thus interpreted. Accordingly, DDI provides an answer to a question distinct from the ER-problem. Although a valuable answer to the question of how models are used, it does not help us too much here, since it presupposes the very representational relationship we are interested in between models and their targets.

An alternative reading of Hughes' account emerges when we consider the developments of the structuralist and similarity conceptions discussed previously, and the previous discussion of deflationism: perhaps the very act of using a model, with all the user intentions and practices that it brings with it, constitutes the scientific representation relationship itself. And as such, perhaps the DDI conditions could be taken as an answer to the ER-problem:

DDI-ER: A scientific model M represents a target T if and only if M denotes T, an agent (or collection of thereof) S exploits the internal dynamic of M to make demonstrations D, which in turn are interpreted by the agent (or collection thereof) to be about T.

This account comes very close to INTERPRETATION as discussed in the previous subsection. And as such it serves to answer the questions set out in Chapter 1 in much the same way. But in this instance, the notions of what it means to 'exploit an internal dynamic', and 'interpret the results' of this to be about T, need further explication. If 'interpretation' is cashed out in the same way as Contessa's analytic interpretation, then the account will be vulnerable to the same issues as those discussed previously. In another place Hughes endorses Giere's semantic view of theories, which he characterises as connecting models to the target with a theoretical hypothesis (1998, 121). This suggests that an interpretation is a theoretical hypothesis in this sense. If so, then Hughes's account collapses into a version of Giere's.

Given that Hughes describes his account as 'designedly skeletal [and in need] to be supplemented on a case-by-case basis' (1997, 335), one option available is to take the demonstration and interpretation conditions to be abstract (in the sense of abstract minimalism discussed above), which require filling in each instance, or type of instance, of epistemic representation. As Hughes notes, his examples of the internal dynamics of mathematical and physical models are radically different with the demonstrations of the former utilising mathematics, and the latter, physical properties such as the propagation of water waves. Similar remarks apply to the interpretation of these demonstrations, as well as to denotation. But as with Suárez's account, the definition sheds little light on the problem at hand as long as no concrete realisations of the abstract conditions are discussed. Despite Hughes' claims to the contrary, such an account could prove a viable answer to the ER-problem, and it seems to capture much of what is valuable about both the abstract minimalist version of INFERENTIALISM 2 as well as INTERPRETATION discussed above.

8.5 KEYS

A final account that can be described as inferential in flavour is Frigg's discussion of what he calls 't-representation' (2010a, 126-132). His discussion of modelling is tied up with discussing their ontology. He distinguishes between the notion of p-representation – i.e. the way in which model descriptions present model systems – and t-representation – i.e. the way in which models systems represent their target systems. The latter is my interest here. Frigg defines it as follows:

T-REPRESENTATION: 'A [*M*] t-represents [*T*] if and only if: (R1) [*M*] denotes [*T*] [and] (R2) [*M]* comes with a key *K* specifying how facts about [*M*] are to be translated into claims about [*T*]' (2010a, 126).

I don't want to discuss this account in huge amounts of detail here, since aspects of it also feature in my own account of representation presented in Part iii. However it is worth briefly sketching how it fares with respect to the problems introduced earlier.

The condition (R1) introduces a directionality in the account, since models denote their targets but not the other way around. (R2) allows for the possibility of surrogate reasoning, since investigating a model supplies a model-fact that can be translated into a purported target-fact by means of the key. Since the result of this process is explicitly a *claim* about the target, which could be true or false, the account allows for the possibility of misrepresentation. Relatedly, the standards of accuracy that naturally drop out of this account are that the resulting claims actually hold in the target system. Since Frigg illustrates the account with the example of a map, I suppose that he thinks that T-REPRESENTATION applies to non-scientific epistemic representations as well, thereby rejecting the demarcation problem. And it seems as though the account naturally allows for mathematical models, since facts about those can be translated into claims about the target in some way or another. With respect to the problem of ontology, like Levy and Toon, Frigg invokes Walton's account of make-believe to defend the idea that models are fictional (2010a; 2010b) But unlike those two, Frigg still thinks that models can be thought about as systems in their own right. Levy and Toon's criticisms of Frigg concern a tension between this ontological story and T-REPRESENTATION. According to Walton, fictions are just imaginative activities of players of games of make believe, and thus it remains unclear what the 'facts' about models keys are supposed to apply to.

Like Hughes' account, T-REPRESENTATION seems to capture much of what is valuable about the previously discussed accounts in this chapter. Frigg is explicit that these conditions are abstract in the sense that they need to be filled in in each instance of scientific representation (and thus looks like a version of abstract minimalism above), and moreover, it allows for a conventional association between facts about models and purported facts about the target (in a manner similar to INTERPRETATION), via the introduction of a key. The problem again is not that the account is false but that it leaves too much unsaid. Are all facts of the model translated (probably not); are there any constraints on the nature of a key? How are the facts chosen? And so on. Without further detail it is difficult to assess the account any more than this.

So to briefly summarise. In this section I have discussed accounts of scientific representation associated with Suárez, Contessa, Hughes, and (briefly) Frigg. For the most part I think each of these accounts gets something right about scientific representation. But none of them is completely successful at explicating the notion (because, for example, they are hampered by a commitment to deflationism, or alternatively leave too much unsaid). That being said, it will become clear in Part iii that each of these accounts has played a significant role in influencing my own account of scientific representation.

Part III

A POSITIVE PROPOSAL

9

DEKI

The remainder of this thesis is dedicated to detailing my preferred answer to the ER-problem. The account is inspired by Goodman (1976) and Elgin's (1983; 1996; 2009; 2010) discussions of *representation-as* in the context of pictorial, or more generally, aesthetic representation. In this chapter I revisit their discussions and argue that they need amending if representation-as is to underpin scientific representation. I then provide the details of the Phillips-Newlyn machine that functions as an economic model and use it to illustrate my account (the other models introduced in Chapter 1 are discussed in the following chapter).[94]

9.1 GOODMAN AND ELGIN

Many works of art represent their subjects as thus or so. A famous cartoon of Winston Churchill represents him as a bulldog. Less charitably, another represents Tony Blair as George Bush's lapdog. But this type of representation is not limited to caricature: Rembrandt's *Self-Portrait with Two Circles* represents the artist wearing a white hat. The statue of Augustus of Prima Porta represents the Roman emperor as raising his right hand in the *adlocutio* pose. The statue of Arnold Schwarzenegger in Columbus, Ohio represents the former bodybuilder as having a particular musculature. Goodman and Elgin term this sort of representational relationship *representation-as* (Elgin 2010, 3; Goodman 1976, 27). In its general form: a representational vehicle X (e.g. a picture or statue) represents a target or subject Y (e.g. a politician or bodybuilder) as Z (e.g. a bulldog or posing in a particular manner).

Representation-as is a complex form of representation that combines *denotation* and *exemplification* (Elgin, 1983, 141-142). It's worth outlining what Goodman and Elgin have to say about these notions before discussing how they can be combined.

94 This chapter is based on a manuscript in preparation co-authored with Roman Frigg.

9.1.1 *Denotation*

Denotation is the two-place relation between a symbol and the object to which it applies. Although usually restricted to proper names, one of Goodman and Elgin's crucial insights is to apply it more widely:

> 'A picture that represents [...] an object refers to and, more particularly *denotes* it. Denotation is the core of representation' (Goodman 1976, 5 *cf.* Elgin 1983, 19-35).

A number of qualifications about how Goodman and Elgin view denotation need to be added here. First, as noted, they do not restrict it to language. Pictures and statues denote their subjects, photographs denote what they are of, and as I discuss below, scientific models denote their target systems. Secondly, even in the linguistic context, denotation is often restricted to proper names; expressions denoting a singular object. The name 'Arnold Schwarzenegger' denotes the man. As such denotation is distinguished from predication, which deals with general terms. This restriction is unnecessary:

> 'A predicate denotes severally the objects in its extension. It does not denote the class that is its extension, but rather each of the members of that class' (Elgin 1983, 19 *cf.* Goodman 1976, 19).

The predicate 'red' denotes all red things and a picture of an apple in a child's colouring book denotes all apples.

Thirdly, notice that there can be a number of denotational relationships between a picture and its subject:

> 'What a picture is said to represent may be denoted by the picture as a whole or by a part of it [...] Consider an ordinary portrait of the Duke and Duchess of Wellington. The picture (as a whole) denotes the couple, and (in part) denotes the Duke' (Goodman, 1976, 28).

Presumably a part of the picture also denotes the Duchess, another part denotes the Duke's nose, yet another part denotes the Duchess's dress, and so on. In fact, there may, in principle, be an indefinitely large number of denotational relationships that hold between parts of the picture and parts of the situation it denotes.[95]

95 This is not to say that there must be part-part denotational relationships to establish the primary one that holds between the picture and its subject. Examples from modern art provide plausible instances where

What establishes denotation is a more vexing question. The question has a rich history in the philosophy of language, where there are two broad families of approaches.[96] According to the descriptivist approach (which goes back to Frege (1892/1952) and Russell (1905)) names function as disguised definite descriptions (or have senses attached to them that can be identified with a definite description), and as such denote whatever it is that satisfies them. According to so-called direct reference approach (which goes back to Mill (1843), Marcus (1961), and Kripke (1980)), names directly pick out their bearers without going via any descriptive content (where, at least in Kripke's formulation of the account, the mechanism of denotation is an initial baptism and a causal chain from this to every future use of the name). Which, if either, of these approaches is appropriate for establishing the denotation relation that holds between pictures and their subjects is an interesting question, which Goodman and Elgin do not discuss in detail. In Goodman's words: 'routes of reference [including, but not limited to, denotation] are quite independent of roots of reference' (1984, 55) and he is concerned with the former, rather than the latter. I take it that 'routes' of reference concern distinguishing between the different types of representational relationships I discuss in this chapter, and 'roots' of reference concern what actually establishes each of them. In the next section I revisit the roots of reference (denotation in particular) in the context of scientific representation.

Goodman and Elgin take X denoting Y as a necessary but insufficient condition on X representing Y as Z (Elgin 2010, 2; Goodman 1976, 28). It is necessary because it establishes that representational vehicles are about their subjects. Denotation picks out the subject and ensures that the vehicle points to it. The statue of Arnold Schwarzenegger represents Schwarzenegger because it denotes him. Following Goodman and Elgin I call this kind of representation representation-of: X is a representation-of Y if and only if X denotes Y.[97] But denotation is insufficient for representation-as because the sort of representational relationship involved there is richer than the relation that holds between a denoting symbol and its denotatum. The statue does not just denote Schwarzenegger; it represents him as having a particular musculature, and nothing in the concept of denotation would help explain how the statue manages to do so.

---

there is only one such relation. We can imagine a uniformly red canvas captioned 'Kierkegaard's Mood' which as a whole denotes Kierkegaard's mood (Danto, 1981). It's hard to imagine what it would take for a part of the canvas to denote a part of the philosopher's mood. So, whether or not there are such part-part relationships, and how many of them there are, can only be established on a case-to-case basis.

96 For discussions and surveys see Lycan (2000, Chapters 4,5) and Reimer and Michaelson (2014).

97 To sharply distinguish between different kinds of representation I use the hyphen where appropriate.

9.1.2 *Z-representation*

If denotation is a necessary condition on representation-as, then what are we to say about pictures that fail to denote? Are we to deny that they are representations at all? At first blush, it looks as if Böcklin's *Isle of the Dead* represents an islet dominated by cypress trees, with a boatman rowing a white figure into the cove. But there is no such islet, and thus, there is no such islet to be denoted. But if the painting fails to denote, and denotation is a necessary condition on representation, then why do we think it represents? Goodman and Elgin respond by distinguishing between being a picture of a soandso, and being a soandso-picture.

> 'Saying that a picture represents a soandso is thus highly ambiguous between saying that the picture denotes and saying what kind of picture it is. Some confusion can be avoided if in the latter case we speak rather of a 'Pickwick-representing-picture' of a 'unicorn-representing-picture' [...] or, for short, of a 'Pickwick-picture' or 'unicorn-picture' [...] Obviously a picture cannot, barring equivocation, both represent Pickwick and represent nothing. But a picture maybe of a certain kind – be a Pickwick-picture [...] – without representing anything' (Goodman, 1976, 22).

Just as a unicorn-picture is not a representation-of a unicorn, Böcklin's painting is an islet-picture despite not being a representation-of any islet. This observation goes beyond pictures. Statues of Greek gods, despite not being representations-of anything are still god-representations. The statue of Rocky outside the Philadelphia Museum of Art is not a representation-of anything, because Rocky doesn't exist and therefore cannot be denoted, but it is still boxer-representation. And although many maps are representations-of a target terrain, there are some that are not because the terrain does not exist. Maps of fantasy worlds, like Jon Roberts' maps depicting the countries and regions that feature in G.R.R Martin's Game of Thrones, provide illustrative examples. They are terrain-representations, they portray the geographic and sociopolitical make up of Westeros and Essos, but they are not representations-of anything.

This leads to the introduction of a Z-representation – e.g. Pickwick-representations, god-representations, terrain-representations – and the crucial distinction between being a Z-representation and being a representation-of a Z. The former is an unbreakable one-place predicate; the latter is a two-place relation that holds between a representational vehicle and its subject. This is a crucial distinction because there is a complete disconnect between what kind of representation X is and what X is a representation of: the kind of X does not determine what X denotes, and the denotation of X does not determine its kind. Not every islet-representation denotes an islet (Böcklin's), and islets can be denoted by representations that aren't islet-representations (for instance when a travel agency uses a palm-tree-picture to advertise a trip to the Maldives). Such representational practices are common in different contexts. In Dutch still life a snail-picture denotes humility and a skull-picture denotes mortality, and in Bollywood movies a two-intertwined-roses-representation denotes the couple being intimate.

What does it take to be a *Z*-representation? In the case of pictorial representation this is a much discussed issue. A widely held view is that a picture *X* is a *Z*-representation (in virtue of portraying a *Z*) if, under normal conditions, from the appropriate vantage point, an observer would see a *Z* in *X*. This idea is developed in so-called 'perceptual accounts' of pictorial representation, associated, among others, with Gombrich (1961), Schier (1986), Wollheim (1987, 1998), and Lopes (2004).[98]

Goodman and Elgin take a different route and explain *Z*-representation in terms of genres. A genre is a class of representation with the same ostensible subject matter.

> '[W]e readily classify pictures as landscapes without any acquaintance with the real estate – if any – that they represent. I suggest that each class of [Z]-representations constitutes a small genre, a genre composed of all and only representations with a common ostensible subject matter [...] And we learn to classify representations as belonging to such genres as we study those representations and the fields of inquiry that devise and deploy them' (Elgin, 2010, 3).

These genres are habitual ways of classifying and as such they are neither sharp nor historically stable, and they typically resist exact codification (Goodman, 1976, 23).

How pictures represent is a fascinating and important question, but it isn't one that I will discuss in any detail here. My concern is how scientific models work, and theories of pictorial representation do not directly carry over to the scientific case, irrespective of what these views are. I develop my own account of categorising *Z*-representations in science in the next section.

9.1.3 *Exemplification*

An item exemplifies a property if it at once instantiates the property and refers to it:

> 'Exemplification is possession plus reference. To have without symbolising is merely to possess, while to symbolise without having is to refer in some other way than by exemplifying' (Goodman, 1976, 53).

98 It bears noting that there are alternative accounts available; see Kulvicki (2006) for a review.

An item that exemplifies a property is an exemplar (Elgin, 1996, 171). The paradigmatic example of an exemplar is a sample. The swatches of cloth in a tailor's booklet of fabrics (Goodman, 1976, 53), the chip of paint on a manufacturer's sample card, and the bottle of shampoo we receive as a promotional gift (Elgin, 1983, 71) all refer to relevant properties – a pattern, a colour, a particular hair treatment – and instantiate them.

The formula 'exemplification is possession plus reference' stands in need of qualification. The point to emphasise is that the 'plus' ought not to be read as a conjunction of two logically distinct conditions. Goodman and Elgin do not use 'refer' as a synonym for 'denote'. Rather, denotation and exemplification are themselves basic *modes* of reference (Elgin, 1996, 71).[99] Reference is thus seen as a determinable for which denotation and exemplification are determinants. So, exemplification cannot literally be reference with something else added to it. Rather, exemplification is the kind of reference that employs instantiation of a property to achieve reference to that property. This can be encapsulated by altering the formula as follows: an item exemplifies a property *P* if and only if it instantiates *P* and *thereby* refers to *P*.

Exemplification requires instantiation: an item can exemplify a property only if it instantiates it (Elgin, 1996, 172). Only something that is red can exemplify *redness*. But the converse does not hold: not every property that is instantiated is also exemplified. Exemplification is selective (Elgin, 2010, 6). An exemplar typically instantiates a host of properties but it exemplifies only few of them. Consider the example of a chip of paint:

> 'a chip of paint on a manufacturer's sample card. This particular chip is blue, one-half inch long, one-quarter inch wide, and rectangular in shape. It is the third chip on the left on the top row of a card manufactured in Baltimore on a Tuesday. The chip then instantiates each of these predicates in the previous two sentences, and many others as well. But it clearly isn't a sample of all of them. Under the standard interpretation, it is a sample of "blue", but not of such predicates as "rectangular" and "made in Baltimore"' (Elgin, 1983, 71).[100]

99 Denotation and exemplification are not mutually exclusive. A symbol does not have to be either purely denotational or purely exemplificational. Indeed, some symbols combine denotational and exemplificational functions to procedure different kinds of 'complex reference', like, as I discussed below, representation-as.

100 Notice that Elgin talks about predicates being exemplified, whereas I prefer to talk about properties. Nothing hangs on this; the account of representation that emerges is compatible with nominalist or platonist accounts of properties (*cf.* Goodman, 1976, 46).

Which properties are exemplified and which properties are merely instantiated is not dictated by the object itself. Turning an instantiated property into an exemplified one requires a deliberate act of selection, and this depends on the context and is carried out against a certain set of background assumptions. The same sample card can exemplify *rectangularity* if used in geometry class, and it ceases to be an exemplar for colour if the painted wall fails to have the colour of the sample card. The specific details of how this works varies from case to case, but at the level of a general theory nothing depends on these details.

One aspect, however, is crucial: exemplars provide epistemic access to the properties they exemplify (Elgin, 1983, 93). So to be exemplified a property has to be selected and be epistemically accessible. A property that satisfies these criteria is salient. The paint chip makes a particular shade of red salient because the context of a paint shop selects red as the relevant property and the nature of the card is such that red is epistemically accessible (a sample card too small to see with the naked eye would not exemplify red). An exemplar is therefore not merely an instance of a property but a telling instance (Elgin, 2010, 5)

9.1.4 *Representation-as*

A key insight on the way to a definition of representation-as is that Z-representations can, and often do, exemplify properties associated with Zs. The Churchill caricature is a bulldog-picture and it exemplifies bulldog-properties like aggressiveness and relentlessness. Rembrandt's self-portrait is a man-representation and it exemplifies the man-property of wearing a white hat. The Schwarzenegger statue is a bodybuilder-representation and exemplifies bodybuilder-properties, like being in a particular pose or having well defined muscles.[101]

But for X to represent Y as Z it is not enough for X to denote Y and also be a Z-representation exemplifying certain Z-properties. To represent Churchill as a bulldog it is not sufficient that the caricature denotes Churchill and is bulldog-representation exemplifying certain bulldog properties. Elgin writes:

> 'Evidently, it takes more than being represented by a tree-picture to be represented as a tree. Some philosophy departments can be represented

101 One may worry that pictures and statues cannot instantiate properties like aggressiveness or having a particular musculature (since they don't, technically speaking, have any muscles whatsoever). Goodman and Elgin acknowledge this and say that these are examples of 'metaphorical exemplification', a notion that requires metaphorical instantiation. A painting may literally instantiate darkness while it metaphorically instantiates being disturbing (Elgin 1983, 81; Goodman 1976, 50). I discuss my preferred take on this issue in the next section.

> as trees. But to bring about such representation-as is not to arbitrarily stipulate that a tree picture shall denote the department' (2010, 4).

And she goes onto to identify a crucial last step involved in representation-as; the exemplified properties need to be *imputed* onto the subject:

> 'I said earlier that when [X] represents [Y] as [Z], [X] is a [Z]-representation that as such denotes [Y]. We are now in a position to cash out the "as such". It is because [X] is a [Z]-representation that [X] denotes [Y] as it does. [X] does not merely denote [Y] and happen to be a [Z]-representation. Rather in being a [Z]-representation, [X] exemplifies certain properties and imputes those properties or related ones to [Y] [...] The properties exemplified in the [Z]-representation thus serve as a bridge that connects [X] to [Y]. This enables [X] to provide an orientation to its target that affords epistemic access to the properties in question.' (2010, 10).

Thus we arrive at the following definition of representation-as:

REPRESENTATION-AS (RA) : X represents its subject Y as Z if and only if:

1. X denotes Y,
2. X is a Z-representation and exemplifies properties $P_1, ..., P_n$ associated with Z, and
3. $P_1, ..., P_n$ are imputed to Y.

Consider the bodybuilder statue as an example (Figure 13). The statue denotes Arnold. It is a bodybuilder-representation and exemplifies properties associated with bodybuilders, like its musculature and pose. These properties are imputed onto Arnold, and as a result, the statue *represents* him *as* having such a musculature (or being in such a pose).

Figure 13: A statue of Arnold Schwarzenegger (public domain) [102]

9.2 FROM ART TO SCIENCE

The question then, is whether RA would work as an answer to the ER-problem. *Prima facie* it seems that it would. Scientific models and pictures have in common that they represent their targets as being thus or so. Just as the statue of Schwarzenegger represents him as having a particular musculature, an ideal pendulum model (with a specific length, weight, and amplitude) might represent a clock's pendulum as having a particular period. Indeed pictures and statues meet at least some of the conditions of adequacy on scientific representation provided in Chapter 1: they can be used to reason about their subjects and they can misrepresent them.[103]

If there is no strict demarcation between how models represent, and how other epistemic representations, including statues and pictures, represent, then this would suggest that RA would double as an account of scientific representation if we take X to be a model, Y a target system and Z a specification of what kind of model X

102 Available here `https://roadtrippers.com/us/columbus-oh/points-of-interest/statue-of-arnold-schwarzenegger?lat=40.80972&lng=-96.67528&z=5`.

103 Indeed Hughes (1997), van Fraassen (2008) and Elgin herself (2010), have all urged that we think about scientific representation in a Goodmanian way (although with the exception of Elgin, they don't elaborate this in any detail).

is. Analysing the ideal pendulum in these terms yields: (1) the model denotes the clock pendulum (2) the model is a pendulum-representation that exemplifies properties like its length and period of oscillation; and (3) these properties, or related ones, are imputed onto the actual clock.

This is a good start. But each of the three conditions stands in need of either articulation or revision (or both) in order to operate successfully in the context of scientific modelling. Before doing so, it is worth discussing in a specific instance of scientific representation in terms of representation-as, and then thinking about these conditions with an eye on applying it to that instance.

9.2.1 *Introducing the Phillips-Newlyn machine*

The first example introduced in Chapter 1 is the Phillips-Newlyn machine, sometimes called the MONIAC (MOnetary National Income Analogue Computer). It is a hydraulic machine built by Walter Newlyn and Bill Phillips in the middle of the 20th Century. The machine consists of a system of reservoirs connected by pipes. Water flows circularly through the machine; being pumped up from a reservoir at the base of the machine, dividing and recombining at various places, and filling various tanks to certain depths as it does so. The water's flow through the pipes and tanks is determined by a series of valves whose behaviour is determined by 'slides' that a user of the machine can alter at will.

The machine is pictured below (Figure 14), but its workings are better appreciated through viewing diagrams that highlight how the different reservoirs are connected to each other. Let's start with a simplified version.

Figure 15 is a schematic representation of the machine. Starting at the the bottom of the machine, there is a tank labelled 'transaction balances'. Water flows out of that tank and is pumped to the top of the machine. Some water leaves the flow (via the 'tax' pipe) and rejoins the flow lower down the machine via the 'government spending' pipe. Further down, some water leaves the flow (through the 'savings' pipe) and enters the 'idle balances' tank. Water from that tank rejoins the central flow further down through the 'investment' pipe. The 'imports' pipe then takes some water out of the main flow into the 'foreign owned balances' tank, and from there water can rejoin the flow through the 'exports' pipe. The resulting flow goes back to the 'transaction balances' tank, from which the water gets recirculated through the machine.

The way that the water flows through the machine depends on numerous, some relatively complex, mechanisms. These are best appreciated by looking at a more complex diagram of the machine (Figure 16).

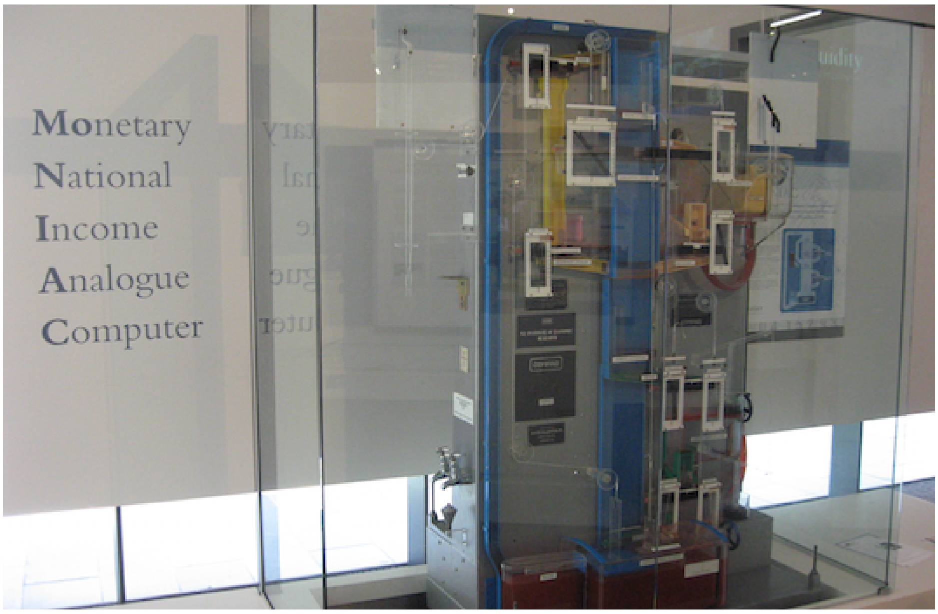

Figure 14: The Phillips-Newlyn machine (public domain)[104]

As the diagram illustrates, the flows through the pipes are determined by valves that open and close depending on the curves on the 'slides' that the model user inputs, and which specify relationships between various aspects of the machine. These slides are made out of perspex and determine how water levels in one reservoir affect how water flows in and out of the others via manipulating size of the valves. The underlying economic relationships are as follows (Barr, 2000, 102-103). Saving, at a given level of taxation, is determined by the level of income (the flow of water in the top of the machine) and the interest rate (the level of water in the 'idle balances' tank). Consumption is what is left of disposable income after saving has taken place; consumption and saving are thus determined simultaneously by the level of income and the interest rate. Investment is determined by the interest rate (level of water in the 'idle balances tank') and the rate of change of income (the flow of water at the top of the machine). Taxes and government spending are determined by the level of income. Imports and exports are determined by domestic expenditure and the exchange rate (the level of water in the 'foreign owned balances' tank).

The machine can be manipulated in various different ways. The relationships between the flows can be altered via changing the 'curves' (the square boxes in Figure 16) which specify the relationships between different aspects of the machine (and correspond to macroeconomic variables). The levels of water in the tank at the bottom

104 Available here `https://commons.wikimedia.org/wiki/File:MONIAC.jpg#/media/File:MONIAC.jpg`.

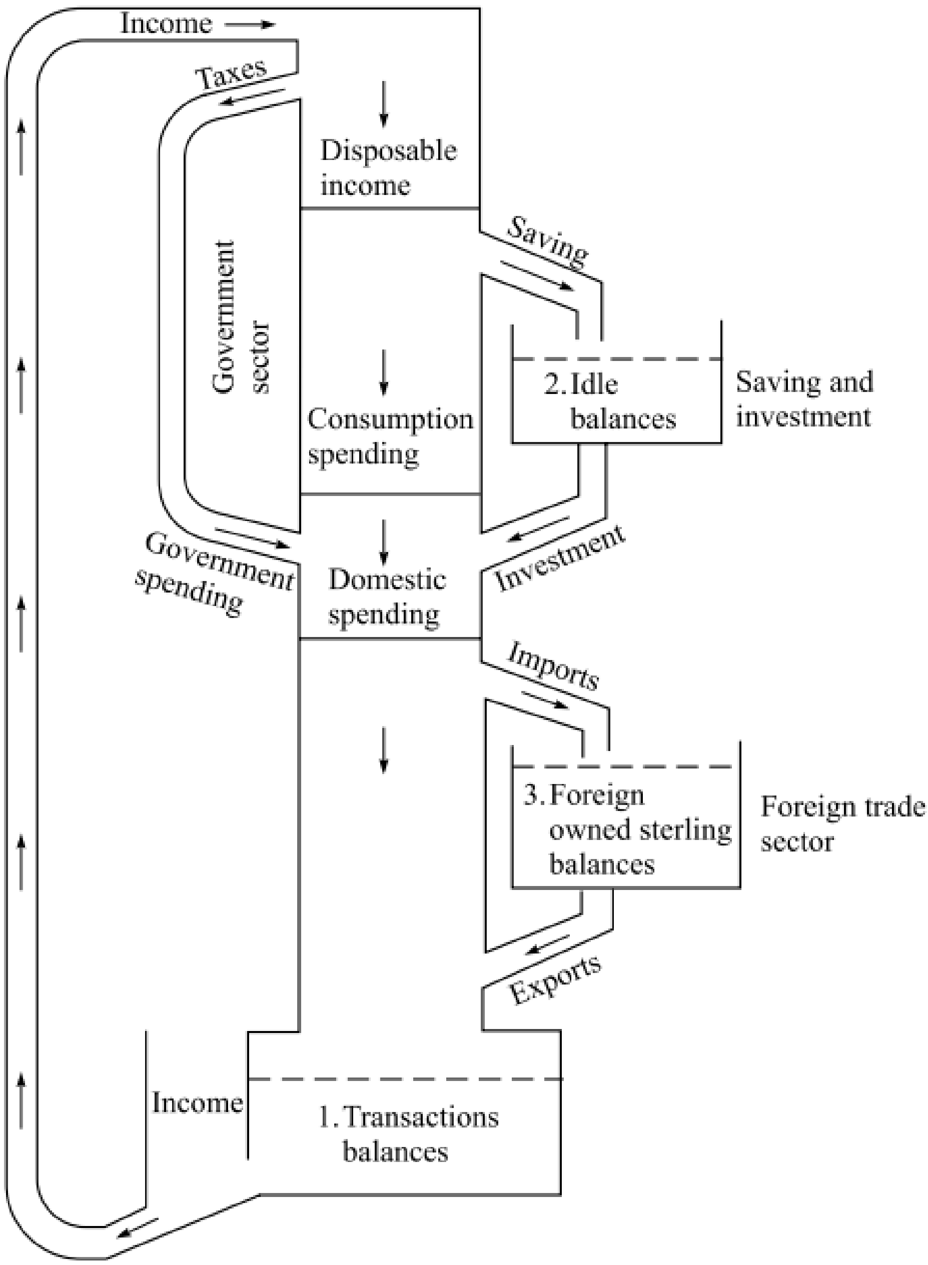

Figure 15: A simple diagram of the Phillips-Newlyn machine (Barr, 2000, 101)

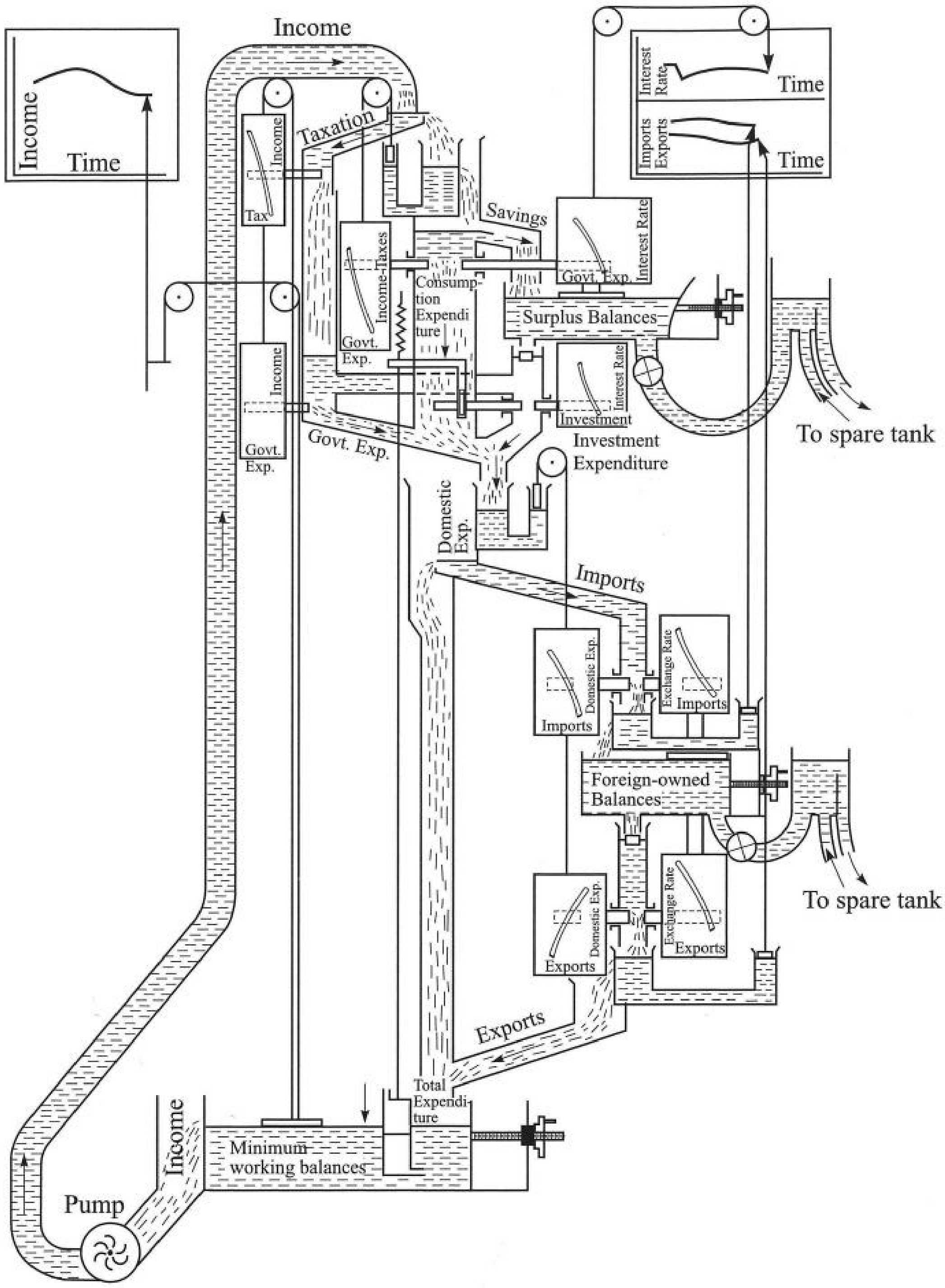

Figure 16: A detailed diagram of the Phillips-Newlyn machine (Barr, 2000, 102)

can be increased, leading to an increased 'income flow' and the knock-on effects that would have on the rest of the machine. The 'surplus balances' and 'foreign-owned balances' tanks are connected to a spare tank which allows a model user to keep the levels of water in those tanks constant (by water overflowing to, or in flowing from, the spare tank as appropriate). In turn, this means that the effect of expanding or restricting the flows out of those tanks on the rest of the machine can also be investigated.

In effect, the machine can be used to solve the IS-LM equations for an open economy.[105] The solutions are given quantitatively; the water levels in each of the three tanks are connected to various graphs that record their height through time. The higher the water in the 'transactions balances' tank, the more water flows through the income pipe and this is recorded in the graph at the top left of the diagram. The higher the level of water in the 'ideal balances' tank the lower the interest rate. And the 'foreign owned balances' tank records the difference between the import and export flows.

The machine is an 'economic model' insofar as it illustrates the (purported) relationship between various macroeconomic variables. But the machine was also customised to represent specific economies as well. In the original version the tank to the right was labelled 'sterling balances', indicating that it represented the British economy; in other versions of the machine the central bank was labelled the 'Federal Reserve' indicating that it represented the US economy (Morgan, 2012, 178). And the flexibility of the machine means that it can be personalised to represent very specific economic circumstances.

Let's now return to the example mentioned at the beginning of the thesis. It's 1953 and economists in the Central Bank of Guatemala are topping up the water tank in their Phillips-Newlyn machine. The economists are worried about a politically motivated decrease in foreign demand for Guatemalan goods, and want to know what effect such a decrease would have on the national economy. They adjust the machine to account for the macroeconomic conditions in Guatemala, and let the machine reach equilibrium. This requires setting the initial amount of water in the machine, specifying the relationships between each variable by means of a slide, and, since the Guatemalan quetzal is (was) pegged to the US dollar, opening up the valve to the spare tank in order to keep the water level in the tank marked 'foreign-owned balances' (which corresponds to the interest rate) constant.

105 The IS equation specifies the relationship between the interest rate and the level of income that arises in the goods and services market. The LM equation specifies the relationship between the interest rate and the level of income that arises in the market for money balances. An 'open economy' is one with imports and exports. Mankiw (2012, Chapters 11,12) provides an accessible introduction to the details of the IS-LM model.

They then close a valve marked 'exports' and watch what happens. To compensate for the fact that more water is flowing into the 'foreign-owned balances' tank than was leaving it, water flowed into the spare tank. Moreover, since less water flows out of the 'export' pipe, the flow marked 'income' falls as well. This provides the sought-after indication of the effects of falling exports on the Guatemalan economy; the economy would contract.[106]

The machine provided this indication in virtue of being an economic model that represented the Guatemalan economy. The question I've been addressing throughout this thesis is: what makes this the case? *Per se* it's a hydraulic machine; a collection of pipes, valves and reservoirs, with water flowing through it. What on earth has it got to do with economics, or the Guatemalan economy in particular? According to RA, the answer is the following. (1) The machine denotes the Guatemalan economy; (2) it is a economy-representation that exemplifies economic properties corresponding to the effect of a decrease in exports,such as leading to an economic contraction; and (3) these properties are imputed onto the Guatemalan economy. As in the case of the pendulum example above, this is a good start. But each of the conditions needs revision or further explication (or both) in order to explain how the machine represented the economy. How does it denote? What makes it an economy-representation? How can it exemplify these properties? And so on. Answering these questions is my next task here.

9.2.2 *Interpretation*

It is a crucial element of RA that the Phillips-Newlyn machine is an economy-representation just as Böcklin's *Isle of the Dead* is an islet-representation and the Schwarzenegger statue is a bodybuilder-representation. In the case of paintings, one could appeal to what a picture portrayed in establishing whether or not it was a *Z*-representation, or, following Goodman and Elgin, perhaps this categorisation could be done according to genre conventions. Irrespective of how these approaches fare in the case of pictorial representation, they are a non-starter in a scientific context. The Phillips-Newlyn machine does not look like an economy under normal conditions and from the appropriate vantage point (or indeed under any conditions and from any vantage point); and if hydraulic machines form an economy-representation 'genre' then

106 This example should be taken with a pinch of salt. There are records that the Guatemalan Central Bank did purchase a machine (Aldana, 2011; Stevenson, 2011). However there are no (publicly available) records of the exact uses the machine was put to. This is no detriment to my argument because none of the points I want to make about representation in general, and about the machine in particular, depend in any way on what exactly happened in Guatemala in 1953.

more needs to be said about what establishes this categorisation. And the machine is no exception. The problem is endemic to models. Caenorhabditis elegans, mice, oval containers, equations, and mathematical structures do not engender visual experiences that are similar to the experiences produced by their respective targets; and if they are members of scientific genres then we need to understand what establishes this. An alternative approach to understanding Z-representation is needed.

The first step towards understanding what makes models Z-representations is to get clear on what kind of object a model in itself is. In fact such specifications are often part of the naming of models. Economists refer to the Phillips-Newlyn machine as a hydraulic model. Likewise we speak of ball-and-stick models, electric-circuit models, and animal models. The material base of a model matters. To assign this a systematic place in an account of representation I introduce a term of art: O-objects. As used here, 'O' is simply a specification of what kind of thing an object is.[107] Derivatively, we can speak of O-properties to designate properties that X has qua O-object. The Phillips-Newlyn machine is a water-pipe-object and having a flow of one litre of water through its central hose per unit of time is one of its O-properties; the brain model is an electric-circuit-object and having a voltage of 1.5V at certain time over its condenser is one of its O-properties; and so on. It's important to note that X does not dictate O. The Philips-Newlyn machine could also be described as a metal and plastic object, or as post-war production object. Any kind of property that X instantiates could ground O. But once it's classified as an O object, then it's O-properties that are the relevant ones.

Classifying something as an O-object does not turn it into an O-representation. Categorising the Phillips-Newlyn machine as a pipe system does not turn it into a pipe-representation. This observation generalises: O-objects do not have to be O-representations; in fact O does not determine Z at all, and in general O and Z are distinct (although there are special cases where $O = Z$). There is nothing in water pipes or electric circuits that makes them economy-representations or brain-representations, and the cat object sleeping in the sun isn't a representation at all. Classifying something as an O-object does not regulate how, and indeed whether, the object functions symbolically.

What, then, turns an O-object into a Z-representation? One might reply that someone simply decides to use it as such. Phillips and Newlyn simply decided to use their machine as an economy-representation. This is not false, but merely pushes the question one step back: what does it take to use an O-object as a Z-representation? To answer this question it is illustrative to see how Phillips describes the workings of

107 There is no expectation that O be a natural kind, and the classification of objects into Os does not demand a rigid classification schema.

(technically a precursor of) his machine when introducing it to the wider economics community:

> 'the production flow of a commodity is represented by the flow of water into a tank. This flow is controlled by a valve, consisting of a flat plate sliding horizontally over a narrow parallel slot [...] The production flow goes into the tank containing stocks, from which is drawn the consumption flow, controlled and measured by a second valve similar to the first [...] Price is assumed to be determined at any instant by the quantity of stocks, represented by the quantity of liquid in the tank, and the demand schedule for them, represented by the capacity of the tank at different levels' (Phillips, 1950, 284).

And then later in the paper, when describing how, given the dimensions of the tanks and valves, scales for relevant quantities are constructed, he describes the relationship between the properties the model has qua O-object in terms of Z properties as follows:

> 'Assume that the price scale is so chosen that the required relation between stocks and price of a commodity is reproduced on the model when *one cubic inch of water is made equivalent to one hundred tons of the commodity*' (1950, 285, emphasis added).

So Phillips turns a pipe system into an economy-representation by taking properties of the machine to 'represent', or be 'equivalent to' economic properties. This is the crucial idea behind what I call an 'interpretation' (not to be confused with Contessa's (2007) 'analytic interpretations' discussed in the previous chapter): certain properties of the model are isolated, and they are appropriately related with some other set of properties. In the above, the properties of the model include the flow of water, the capacity of tanks, and so on. These are then associated with economic properties: the production flow of a commodity, and a quantity of stocks for example.[108] Once this interpretation is established, Phillips switches from talking about 'flow of water' to talking about 'the production flow goes into the tank containing stocks'. In fact, with the interpretation in place, he describes the entire machine explicitly in economic terms despite the fact that those properties don't literally apply to the model object.

The notion of ensuring that properties are 'appropriately related' with another set of properties can be made precise in the following way. Let $\mathcal{O}$ and $\mathcal{Z}$ be sets of relevant O properties and Z properties respectively. One could then define an interpretation

108 Morgan and Boumans (2004, 383) specify the physical properties of the machine that Phillips used to represent different economic elements.

as a bijective function $I : \mathcal{O} \to \mathcal{Z}$. Whilst correct in principle, this definition does not capture all that is important, because it doesn't explicitly distinguish between quantitative and qualitative properties.[109] Properties like 'being a reservoir' and 'being directly connected to the disposable income flow' are qualitative properties: they are all-or-nothing properties in that they either are or are not instantiated. In contrast, properties like 'the flow of water is x litres per minute' are quantitative properties: they come in certain degrees. In the case of such properties it's important to distinguish carefully between the property and its values. To make this distinction explicit I refer to the property 'itself' as the *variable* and to the specific quantity as the *value*. The former are denoted by upper-case letters and the latter by lower-case letters. I further adopt the convention that the members of $\mathcal{O}$ and $\mathcal{Z}$ are either qualitative properties or variables. So O_1 could be 'the flow of water' and o_1 would be a specification of how many litres exactly are flowing through at a given time. An interpretation can then be defined as follows:

INTERPRETATION: Let $\mathcal{O}$ and $\mathcal{Z}$ be sets of O properties and Z properties respectively, so that all members of either set are qualitative properties or variables. An *interpretation* is a bijection $I : \mathcal{O} \to \mathcal{Z}$ such that:

1. Properties are mapped on to properties of the same kind (that is, qualitative properties are mapped onto qualitative properties and variables onto variables).
2. For every variable O_i, and Z_i, there is a function f_i that associates every value of O_i with a value of Z_i.

In specific cases further restrictions on allowable functions could be imposed. In the case of the Phillips-Newlyn machine, for instance, the function associating the flow of water with flow of a commodity is assumed to be linear. However, such restrictions are idiosyncratic to the context and should not be built into a general definition.

An interpretation is what turns an O-object into a Z-representation.

Z-REPRESENTATION: A Z-representation is a pair $\langle X, I\rangle$ where X is an O-object and I and interpretation from $\mathcal{O}$ to some set of Z properties $\mathcal{Z}$.

Returning to the example, X is the Phillips-Newlyn machine; Z is an economy; I is an interpretation mapping qualitative pipe-system properties onto qualitative economy properties ('the reservoir on the right corresponds to the foreign trade sector')

109 A more detailed study of both kinds of properties is a worthwhile enterprise in many ways, but for my current purposes an intuitive characterisation is sufficient. See Eddon (2013) for a discussion of quantitative properties.

and quantitative pipe-system properties onto quantitative economic properties by associating variables ('the flow of water corresponds to the flow of commodity') and specifying a function associating quantities of those variables ('one cubic inch of water is equivalent to one hundred tons of the commodity').

Colloquially X itself might be called a Z-representation and as long as it is understood that this assumes that there is an interpretation in the background no harm is done. It is important, however, that in a final analysis a Z-representation is a pair $\langle X, I \rangle$. What kind of representation X is crucially depends on I, and different interpretations produce to different representations. One could, for instance, interpret the reservoirs as schools and universities and the flow of water as the movement of students through the system. Under that interpretation the same machine would be an education-system-representation.[110]

A model, then, is a Z-representation that is based on an object X that has been chosen by a modeller to be a model object. In the context of modelling X is the material substratum of the model, variously referred to as 'model-object', 'vehicle' and 'source'. When I refer to the object itself, rather than the object-interpretation pair, I use the term the 'base' of the model. A 'mere' object X is turned into the base of a model if an agent, or more generally a scientific discipline, chooses it as such and endows it with an interpretation.

MODEL: A model is a Z-representation: $M = \langle X, I \rangle$.

Sometimes the object that serves as the base of a representation is a 'ready made' object. Worms, mice, and electric circuits predate their use as models and came into existence independently of any modelling enterprise. Other times the base is tailor-made for the situation, as is the case with the Phillips-Newlyn machine. To understand representation the provenance of these objects is irrelevant. The choice of suitable object is the creative act of a modeller and although the choice may be informed by interpretations that one would like to impose on an object, it is in no way determined by it – economies can be modelled with things other than pipe systems.

The claim that everything can be a model of anything else drops out as a corollary of this definition.[111] There are no restrictions imposed on what X can be cho-

110 Indeed, an artistic reinterpretation of the machine happened in Michael Stevenson's 2006 art installation *Answers to some Questions about Bananas*. The installation showed a reconstruction of the machine that was deliberately left unattended throughout the duration of the exhibition. This lead to it rusting and falling apart, and, given the intentions of the artist, plausibly being interpreted as a failure-representation. See `http://www.michaelstevenson.info/projects/answers_to_some_questions_about_bananas/` for information about the exhibition.

111 This claim is found, for example, in Callender and Cohen (2006, 73), Frigg (2010a, 99), Giere (2010, 269), Suárez (2004, 773), Swoyer (1991, 452), and Teller (2001, 379).

sen as the base of a model. Another immediate corollary is that models need not be representations-of anything. Far from being an unwelcome eccentricity, this is an advantage of the account. It provides a natural answer to how models without targets represent: they are Z-representations that are not also representations of a Z (or indeed any target whatsoever). The Phillips-Newlyn machine would be an economy-representation even if it had never been used as a representation of an actual economy (Guatemalan or otherwise). Architectural models showing buildings that have never been built are not representations-of anything but they are building-representations nevertheless. And these models can be used to investigate how Z properties work, despite the fact that they are not representations of any particular target system.

A few clarifications about Z-representations are in order. Firstly, the definition does not require that all of the O object's properties are collected in $\mathcal{O}$, nor does it require that $\mathcal{Z}$ contain a complete list of Z properties. Compiling complete lists of all of a model object's, or target system's properties, is neither possible nor desirable. Nor is it necessary for an interpretation. All that is required for an interpretation is that there is least one property in each set. In fact, an interpretation can be highly selective in the properties that it considers. Which properties go into $\mathcal{O}$ and $\mathcal{Z}$ is determined by contextual factors, like the research question at hand, the purpose of the model, and the ability of the model user. Some interpretations are richer than others and the richness of an interpretation is at least in part a function of how many O properties are interpreted. But richness is not part of the notion of an interpretation.

Secondly, and in light of the above remarks, interpretations are not set in stone. In different contexts, where the model user has different purposes for example, the properties that feature in $\mathcal{O}$ and $\mathcal{Z}$ (and the interpretation function itself) may change. One nice example of this is that the Phillips-Newlyn machine often leaked water onto the floor when it was run. Originally this was seen as a technical problem of the machine that needed to be fixed, and as such it wasn't part of the interpretation given. However, at some point economists realised that it was actually an interesting feature and interpreted it as the flow of money from the regular economy into the black economy (Morgan and Boumans, 2004, 397, fn.14). This amounts to a property being added to $\mathcal{O}$, and a corresponding one to $\mathcal{Z}$, and the interpretation function being expanded in the appropriate way.

Finally, consider again the choice of O and Z. In the case of the Phillips-Newlyn machine they do not coincide, the machine is a hydraulic system, and it functions as an economy-representation. But this isn't the case in all scientific models; in some instances they coincide. The architect's cardboard house is a house object that is used as a house-representation; when studying ships floating-metal-vessel objects are used as floating-metal-vessel-representations; and when investigating the aerodynamic properties of cars, small car-shaped objects are put into wind tunnels and used as car-shaped-object-representations. These are usually considered to be iconic models (Black, 1962; Peirce, 1932). The definition of a Z-representation offered above affords a ready definition of an iconic model: a model M is an iconic model if and only if $O = Z$. Scale models, then, are a special kind of iconic models, namely ones in which some designated O_i are spatial dimensions and the f_i are linear functions with respect to those spatial dimensions. A small car, for instance is a scale model of a large car if measurements of features in the three spatial dimensions scale with the constant factors given by the scale, for instance 1:50.[112]

9.2.3 *Scientific exemplification*

Goodman and Elgin's notion of exemplification was introduced in the previous section. An item X exemplifies a property P and thereby refers to P. In order for this to occur, P must be selected as salient in the context under consideration, and this requires that it is epistemically accessible. This holds in the scientific context, as in art. In the case of the Phillip-Newlyn machine, the research context selects how the water flows through the machine, and the relative height of the liquid in its various tanks through time as salient properties. It doesn't select being made of Perspex, or being 2.5m tall. And it's not just properties that are irrelevant to the workings of the machine that are not selected as salient. In order for the water to move around the model at all, it requires a motor that pumps water from the floor level tank up to the top of the machine, and the force of gravity that draws the water downwards through the various pipes and reservoirs. Although these aspects of the machine are essential to its workings, they do not correspond to any economic features. Thus, they are not selected as relevant features of the machine in the context of using it as an economic model.[113]

These considerations motivate the following definition:

EXEMPLIFICATION: X exemplifies P in a context C if and only if

1. X instantiates P, and
2. P is salient in C.

where P is salient in C if and only if

112 Of course, a scale model is defined by linear scaling in designated dimensions. For many other quantities the scaling relations may be different. It is often not the case, for instance, that the air resistance of 1:50 car model also scales linearly with 1:50. See Sterrett (2006) for a useful discussion of scaling relations.

113 Morgan and Boumans (2004, 386) stress this, and emphasise that one shouldn't take these two to be the economic 'analogue' of the principle of effective demand, since this would negate the idea of a circular economic flow with no clear separation between the upstream (the pump) and downstream (gravity).

i. C selects P as a relevant property, and
ii. P is epistemically accessible in C.

The sample card exemplifies a certain shade of red because it instantiates that shade of red, and in a paint shop context, that shade is epistemically accessible and selected as a relevant property.

I am steering towards an account according to which a model $M = \langle X, I \rangle$ exemplifies certain properties. But M does not seem to be the right kind of object for exemplification: the instantiation condition is mostly unattainable, while salience is mostly trivial once an interpretation is in place. I consider both of these considerations in order.

Do scientific models instantiate the properties that they refer to? The problem is that if $O \neq Z$, then the model object X does not instantiate properties associated with Z, and thus cannot exemplify them. The Phillips-Newlyn machine instantiates water flows but not commodity flows, and so it can never exemplify the latter. But the intuition remains that the machine makes *economic* properties salient; it can be used to learn about the various relationships between macroeconomic variables. So what are we to make of sentences like $\psi =$ 'the Phillips-Newlyn machine exemplifies an economic contraction after a decline in exports'. Since the machine doesn't instantiate this property, and since, at least as it stands, by definition it must if the machine is to exemplify it, in what sense, if any, is that sentence true?

I can think of two strategies, broadly construed, to account for this using the notion of exemplification (and I discuss a third option later in this section). The first is to grant that sentences like ψ are, strictly speaking false. But this doesn't make them entirely useless; there is still a sense in which when Phillips described his machine in terms of 'consumption flow' and interest rate', he wasn't saying something which was blatantly false. Their truth value seems analogous to the truth value of sentences like 'Sherlock Holmes lives at 221b Baker Street'. There is still a sense in which they are true, despite the fact that, strictly speaking, they are false. Various different options are available for cashing out the sense in which those sentences could be true (see Sainsbury (2010) for a review). But two seem particularly promising from the perspective of making sense of sentences like ψ. Firstly, we might appeal to Sainsbury's notion of truth-relative-to-a-presupposition (2010, 143-151, 2011). The idea is that when a speaker utters a sentence like 'Sherlock Holmes lives at 221b Baker Street' they presuppose that Sherlock, and 221b Baker Street, exist. And then relative to this presupposition, the uttered sentence is true. Notice that this does not require that the speaker *believe* the presupposition. Sainsbury also takes his account to apply to conversational contexts like the following:

> '*Violinist*: A disturbed patient is recounting his (entirely fictitious) early history to his therapist:
>
> > When I was young, I played the violin. I performed Beethoven's sonata in E flat at the Wigmore Hall.
>
> The therapist knows this is false, but decides it's best to roll with her patient's delusions and says:
>
> > Did you play an encore?' (Sainsbury, 2011, 143-144).

The question presupposes that the patient played at the Wigmore Hall, but the therapist explicitly does not believe this. Instead, Sainsbury invokes Stalnaker's (1970; 1984; 2002) propositional attitude of *acceptance* to explain what's going on.[114] The therapist accepts that the patient played such a recital, and this makes sense of the question 'did you play an encore', relative to the presupposition, despite the fact that therapist does not believe that the patient played at Wigmore.

This approach could also work for cases like Phillips describing his machine in economic terms, uttering sentences like ψ for instance. Despite the fact that he could not plausibly believe that they were literally true, one could invoke the idea that he accepts such sentences, and takes them to be true-relative-to-an-interpretation. In this instance, the interpretation provides the presupposition that the machine is an economy, and specifies its properties in a rigid way (based upon the actual properties of the machine, and the specific interpretation in play).

An alternative approach would be to appeal to Walton's notion of pretense (1990). In a game of make believe we pretend that stumps are bears. And this allows us to make sense of sentences like 'there are four bears in the garden' despite the fact that such a sentence is not literally true. Relative to the props and rules of generation, players of the game are prescribed to imagine that there are that many bears in the garden (assuming, for example, that there were four stumps in the garden that were props for the principal of generation that prescribes the imagining of bears). Analogously, one could also treat interpretations as principles of generation, and model objects as props. In the case of the Phillips-Newlyn machine, the machine itself is a prop, and depending on its hydraulic properties, the interpretation prescribes us to imagine that foreign exports decrease and the economy contracts.[115] So although sentences like ψ

114 Notice that Stalknaker's *acceptance* is not the same as van Fraassen's, discussed in Chapter 6 previously. At least as Sainsbury uses it, *acceptance* won't generate any cases of Moore's paradox, since a presupposition can be granted and then retracted in the same sentence without pragmatic contradiction (Sainsbury, 2011, 146-148).

115 It should be emphasised that such an account would not be the same as the direct view accounts discussed in Chapter 7. According to this proposed use of Waltonian pretense, it explains how an object becomes a Z-representation, not an epistemic representation-of a target system.

are strictly speaking false, they are true-relative-to-an-interpretation in the sense that they are prescribed to be imagined by players in the specific game of make believe.

Although I think that the above strategies are worthy of further research, I adopt an alternative strategy here in order to aid exposition throughout the rest of the thesis. The key thing to notice is that what was important about the notion of exemplification was that when X exemplifies P it makes it salient, i.e. the context C selects P as a relevant property and P is epistemically accessibly in C. The question, then, is whether this can be the case despite the fact that the property is not, strictly speaking, instantiated. And it's plausible that it can. In the case of the Guatemalan economists using the Phillips-Newlyn machine, it is clear that a decline in foreign exports leading to an economic contraction is selected as a relevant property. It was precisely what they were trying to find out. Similarly, the property does seem to be epistemically accessible in this context, albeit in a more roundabout way than how *redness* is epistemically accessible in the context of using a sample card in a paint shop. By investigating the hydraulic properties of the machine, and considering them with respect to the interpretation offered, the economists could access the relevant economic property; not in a direct sensory manner, but access it nevertheless.

So it's the properties that the object actually instantiate qua O object, combined with the interpretation, that provides this epistemic access to Z properties. This insight allows the introduction of the idea of instantiating-under-an-interpretation, or I-instantiation for short:

I-INSTANTIATION: Let X be an O object and I an interpretation. A model $M = \langle X, I \rangle$ I-instantiates a property P if and only if X instantiates an O property P' which satisfies the following condition: $I(P') = P$, and if P and P' are variables then the value of P I-instantiated by M is equal to p, where $p = f(p')$.

The definition of exemplification can then be liberalised accordingly, by replacing the term 'instantiates' with 'instantiates or I-instantiates'. So the Phillips-Newlyn machine I-instantiates properties like a decline in exports leading to an economic contraction (under the appropriate interpretation), and sentences like ψ come out as true according to the new definition of exemplification.

The worry with instantiation was that it was too hard to come by; the opposite problem seems to beset saliency. Interpreting O properties in terms of Z properties seems to involve selecting these properties and selecting them seems to presuppose them being epistemically accessible. If this were true then all the Z properties that feature in the interpretation would automatically be exemplified, and this would trivialise the notion of exemplification. Fortunately, interpretations don't seem to work that way. An object can exemplify only those properties that are covered by an interpretation, but this does not imply that every property covered by an interpretation is *ipso facto* exemplified. This is because interpretations can, and often do, cover *O* properties we are unaware of, or not interested in, in a given context. Suppose there was a small pipe in the top right corner of the machine. We haven't paid any attention to that pipe, and the pipe is made from white plastic so that we cannot see how much water flows through it (or whether any water flows at all). The flow of water (if any) is covered by the interpretation, but it has not been made salient and so is not exemplified. Or consider again the case of the Guatemalan economists using the model to represent the impact of a decrease in foreign demand for Guatemalan goods; they may have been particularly interested in the change in the equilibrium values once the appropriate change had been made to the valve marked 'foreign exports'. This means that the machine would exemplify this property. But in other contexts, this property might not be exemplified at all. For example, when explaining the working of the machine, Phillips himself ignores the impact of foreign imports and exports until the end of his paper (1950, Section III). Which means that, although *I*-instantiated, the relevant *Z* properties would not be made epistemically salient and thereby would not be exemplified (in Section I and II) even though they have been covered by the interpretation all along.

Whether or not a *Z* property covered by the interpretation is also exemplified depends on whether we have epistemic access to the corresponding *O* property and on whether the context selects that *O* property as a focal point of investigation. The adoption of an interpretation in no way determines that this has to be the case. *X*, together with the interpretation, provides a 'menu' of *Z* properties that the model *I*-instantiates. Whether or not any of these properties is exemplified depends on the context in which the *Z* representation is used in a way that is sensitive to the purposes and interests of the model user.

9.2.4 *Imputation and Keys*

So far I have focused on what turns an ordinary object into a *Z*-representation. And Goodman and Elgin's observation that *Z*-representations do not have to represent any *Z*, or indeed any target system whatsoever, remains true in the context of scientific representations. I take this to be a significant attribute of my account. But at least some models do represent particular targets. RA stipulates two conditions for this to happen: *X* has to denote *Y* and *X* has to exemplify properties that are imputed to *Y*. I discuss this latter condition first, before turning to denotation in the next subsection.

Imputation can be analysed in terms of stipulation. The model user may simply stipulate that the properties exemplified by the model hold in the target system, and this is what establishes that the model represents the target as having those properties.[116] In this way models allow for surrogative reasoning; imputing a property onto a target generates a hypothesis: Y has whatever property was imputed to it. And notice that nothing in this discussion requires that the target actually has the property imputed; the result of the imputation can be right or wrong in a way that allows for models to misrepresent their targets.

However, in many cases of representation-as the properties exemplified by a Z-representation aren't transferred to a target unchanged. In fact in her discussion of imputation, Elgin posits that a representation imputes the exemplified properties 'or related ones' to Y (2010, 10). This observation is particularly pertinent in scientific contexts. The properties of a model are rarely, if ever, taken to hold directly in their target systems and so the properties imputed onto targets may diverge significantly from the properties exemplified in the model.

The problem with invoking 'related' properties is not its correctness, but its lack of specificity. Any property can be related to any other property in some way or another, and as long as nothing is said about what this way is, it remains unclear what properties X ascribes to Y. So what connects the properties exemplified by a Z-representation and those that are imputed to the target system? I do not think that there is a universal answer to this question. In some cases the connection could be described as 'de-idealisation'. In the Phillips-Newlyn machine – which was known to have a margin of error compared to Hick's equations on which it was based – the connection was to move from exact properties (like the interest rate being x) to intervals around those properties (like the interest rate being $x \pm 4$ %), or the property imputed could be even less specific, like an imputed positive correlation between foreign investment and total output, without any precise specification of the correlation. In other cases of scientific representation the connection might be quite arbitrary.

One could put faith into context and believe that it determines what properties are imputed to the target. I'd rather not. It remains unclear what a model says about its target as long as the relation between the properties exemplified by the model and the

116 One worry might be that this allows for radically non-standard uses of models, where arbitrary properties are imputed onto arbitrary target systems. This objection relates to the Humpty-Dumpty problem discussed in Chapter 2, as it seems as though, for example, a pendulum model could impute properties like having a particular period onto a target system like the human brain, which seems counter-intuitive. I think these sorts of cases are ruled out by the denotation condition discussed below, but for those with residual worries I would not be adverse to building some 'stability conditions' into the definition of imputation. Perhaps model properties are only imputed onto targets if they do so in a way that is stable across multiple uses of those models. This is a question deserving of further research.

properties imputed to the target remain unclear, and so it is crucial to make explicit in every case of modelling what that relationship is. I therefore prefer to write an explicit specification of the relation between the two sets of properties into my account of scientific representation. Let $P_1, ..., P_n$ be the Z properties exemplified by the model, and let $Q_1, ..., Q_m$ be the 'related' properties that the model imputes to Y (n and m are positive natural numbers which can, but need not, be equal to allow for the fact that not every exemplified property might correspond to an imputed one). Then the representation must come with a key K specifying how exactly $P_1, ..., P_n$ are converted into $Q_1, ..., Q_m$:

KEY: Let $M = \langle X, I \rangle$ be a model, and let $P_1, ..., P_n$ be Z properties exemplified by M. A key K associates $\{P_1, ..., P_n\}$ with a set $\{Q_1, ..., Q_m\}$ of properties that are candidates for imputation on the target system: K can thus be treated as a mapping (not necessarily a function[117]) from $\{P_1, ..., P_n\}$ to $\{Q_1, ..., Q_m\}$.

The third clause in RA then becomes: X exemplifies $P_1, ..., P_n$ and imputes some of the properties $Q_1, ..., Q_m$ to Y, where the two sets of properties are connected to each other by a key K.

The idea of a key, which is clearly influenced by Contessa (2007) and Frigg (2010a), comes from maps, paradigmatic examples of epistemic representation. Consider a map of the world. It exemplifies a distance of 29cm between the two points labelled 'Paris' and 'New York'. The map comes with a key, which includes a scale, 1:20,000,000 say, and this allows us to translate a property exemplified by the map (the 29cm distance between dots marked 'New York' and 'Paris') into a property of the world (that New York and Paris are 5,800km apart). Or consider the case of a scale model of a ship being used to represent the forces an actual ship faces when at sea. The exemplified property P in this instance is the resistance the model ship faces when moved through the water in a tank. But this doesn't translate into the resistance faced by the actual ship in the same way in which distances in a map translate into distances in reality. In fact, the relation between the resistance of the model and the resistance of the real ship stand in a complicated non-linear relationship because smaller models encounter disproportionate effects due to the viscosity of the fluid (*cf.* Sterrett, 2006). The exact form of the key is often highly non-trivial and emerges as the result of a thoroughgoing study of the situation. Determining how to move from properties exemplified by models to properties of their target systems can be a significant task, and should not go unrecognised in an account of scientific representation.

117 Using 'mapping' rather than 'function' allows for one-to-many relationships between P properties and Q properties.

K is a blank to be filled. What key a given model is based on depends on myriad factors: the scientific discipline, the context, the aims and purposes for which the model is used, the theoretical backdrop against which X operates, etc. Building K into the definition of scientific representation does not prejudge the nature of K, much less single out a particular key as the correct one. In some instances the key might be the identity key: the properties exemplified by the model are imputed unchanged onto the target. In such cases my definition of scientific representation captures all that the more nuanced similarity based accounts got right (and without the need to invoke the notion of similarity at all). The 'relevant' properties discussed there are, according to my account, relevant because they are exemplified. The hypothesis that M and T are similar (in the sense of partial identity) with respect to them amounts to the identity key. If similarity in the sense of likeness is preferred instead, then the key will map properties exemplified by the model to 'similar' properties as demanded. And since I place no restriction on the nature of the properties exemplified, or imputed, this allows for structural properties to be exemplified and imputed onto targets as well.[118]

But in keeping with my concerns about treating similarity, structural or otherwise, as a universal style of scientific representation, introducing the notion of a key allows the account to capture cases where similarity plays no role. In some cases the key might take the form of an 'ideal limit key' (*cf.* Laymon 1990; Frigg 2010a, 131-132), or 'approximation key' (Ducheyne, 2008). And keys might also associate exemplified properties with entirely different properties to be imputed onto the target (for example, colours with tube lines as is the case in the London Underground map). The requirement is merely that there must be some key for something to qualify as a scientific representation. The above examples show that introducing keys does not amount to smuggling in a mimetic conception of representation via the back door. On the contrary, keys can be as conventional as they like.

The introduction of a key raises a third possibility for answering the problem discussed above; that the Phillips-Newlyn machine, and models where $O \neq Z$ more generally, do not instantiate properties they are taken to exemplify. Recall the sentence ψ = 'the Phillips-Newlyn machine exemplifies an economic contraction after a decline in exports' and the concern that it could not be true, strictly speaking, since the machine doesn't instantiate any economic property whatsoever. Rather than appealing to a weaker notion of truth, or redefining exemplification, we could instead grant that the sentence is completely false, but acts as shorthand for 'the Phillips-Newlyn machine exemplifies a decrease in water in the machine after a close in the

118 Although one might want to rule these out for reasons strictly speaking external to the account of representation I am developing here, the questions concerning the 'structure' of models and targets for example. I discuss this in the next chapter.

valve marked "exports"' combined with a specification of a key translating hydraulic properties into economic ones. In general, the idea would be to ensure that models only exemplify properties that they instantiate, and the role of the interpretation gets shifted into the notion of a key. So the machine would exemplify a hydraulic property that gets translated by a key K into an economic property (either in the strict sense engendered by the interpretation with its function associating values of quantifies of water, or in a weaker sense in line with the idea that economists might only take them to be correlated without assigning a specific function to the correlation). Again, with further development this might serve as a successful alternative to my own preferred account, but I won't explore it in any more detail here.

9.2.5 *Denotation*

The final condition on a model M representing its target T as thus or so is that M denote T. As mentioned above, the use of denotation in this context follows Goodman and Elgin's liberalisation of the notion to allow it to include the relation that holds between pictures, or statues, and their subjects, as well as allowing a vehicle (picture, model, predicate) to denote multiple subjects, targets, or elements in its extension.

So what establishes this kind of relation? Recall the two broad approaches: the descriptivist approach, according to which the vehicle denotes whatever satisfies its 'content', and the direct reference approach, whereby the vehicle directly picks out the denotatum without going via any descriptive content whatsoever. How might these be developed in the case of scientific models? How do scientific models come to denote?

This question has received little to no attention in the literature on scientific modelling.[119] So it is useful to return to the context of pictorial representation. Lopes (2004, Chapter 5) develops both the approaches to denotation in that context. According to a descriptivist about pictorial denotation, just as a name denotes its bearer in virtue of the bearer satisfying description(s) associated with the name, a picture denotes its subject in virtue of the latter satisfying the picture's 'pictorial content', i.e. the visual properties it represents its subject as having. So, a descriptivist account of pictorial denotation would claim that a Z-picture denotes a Z (or Zs) in virtue of being a Z-picture. This is something that Goodman and Elgin explicitly deny, and is the target of Lopes' criticism.

I want to remain agnostic about how denotation in the context of scientific modelling is established (or at least leave it as a topic for future research). But some suggestions

119 As I have discussed, by and large philosophers of science have helped themselves to the notion of 'denotation' (Contessa, 2007), or the even harder to understand notion of 'representational force' (Suárez, 2004) without expanding on what establishes them.

are in order. Firstly, I think that in many cases, denotation is borrowed from language. In the London Tube map, we see a circle with 'Brixton' marked next to it. The circle denotes Brixton and it does so by borrowing the denotation of the proper name. Many models seem to work in this way as well. The Phillips-Newlyn machine denotes whatever the word 'Guatemalan economy' denotes. And parts of the machine denote parts of the economy in virtue of stickers printed next to them. This is to hand over the problem of uncovering the roots of denotation to the philosophy of language, and any account of linguistic denotation is compatible with the account of representation I am developing here (since, even if it is a descriptivist account, denotation isn't established via the 'content' of the model, but rather the 'content' of words used in the modelling context).

Secondly, it bears noting that there is an inherent tension between my account of representation and a descriptivist account that claims that every instance of a model M denoting a target T is established via T satisfying M's content. To see why, recall how important it is to distinguish between representation and accurate representation. The former is my primary interest throughout this thesis, and my account of representation is designed to allow for the possibility of misrepresentation. Now, we can define the 'content' of the model as everything that a model imputes, after the exemplification and keying up processes have occurred. And if the model denoted anything and everything that satisfied that content, and didn't denote anything else, then it's difficult to see how misrepresentation could enter the picture. If the model's target(s) included all and only those systems that satisfied what the model imputed this would preclude a model *mis*representing any target.

One could respond by invoking Searle's (1958) 'cluster theory' in the account of model denotation. The idea, which remains to be developed, would be that a model denotes any target that satisfies an appropriately weighted proportion of the model's content. And this would allow for the possibility of models misrepresenting targets, at least in some respects (i.e. as long as a system satisfied 'enough' of the properties that the key delivered, then the model would denote it, but this allows the system to fail to satisfy at least some of them). But the problem with such an approach is that it would preclude *total* misrepresentation; cases where everything that a model 'tells us' about its target system is false. Although such models would be complete failures in representational accuracy, my inclination is to treat them as representations nevertheless. With these observations in mind, I think that if it turns out that there are cases where what a model denotes is not parasitic on whatever the model users' words denote, then either a pluralistic account (i.e. one where model-target denotation relationships can be established in different ways in different cases), or an account developing the direct reference view is more appropriate in the context of scientific representation. This, of course, would require expanding on whether models can 'baptise' their targets, and whether the appropriate causal chains between different uses of models can be established. I think this is a viable avenue for future research.

9.3 THE DEKI ACCOUNT

It's now time to tie together the previous discussion into a fully fledged account of scientific representation. With the aforementioned notions in mind we can define what it takes for a scientific model to represent its target as thus or so in the following manner. In short, the claim is that a model represents its target as thus or so if and only if M is a Z-representation that denotes T and imputes properties $Q_1, ..., Q_m$ onto T. In more detail:

DEKI: Let $M = \langle X, I \rangle$ be a scientific model, where X is an O object and I is an interpretation from O properties to Z properties. Then M represents a target system T as thus or so if and only if:

1. M denotes T (and in some cases parts of M denotes parts of T),
2. M exemplifies Z properties $P_1, ..., P_n$,
3. M comes with a key associating the set $\{P_1, ..., P_n\}$ with a set of (possibly identical) properties $\{Q_1, ..., Q_m\}$, and
4. M imputes at least one of $\{Q_1, ..., Q_m\}$.

For obvious reasons, I call this the DEKI account of scientific representation. Figure 17 provides a schematic representation of the account.

So let's provide a complete analysis of how the Phillips-Newlyn machine represented the Guatemalan economy in 1953. The machine (X) is a hydraulic object (O) that was constructed by Phillips and Newlyn to be the base of a model. Z is an economy. They endowed the machine with an interpretation (I) which maps O properties onto Z properties. The machine so interpreted is an economy-representation, and as such is a model M. The economists in the Guatemalan Central bank used M as a representation-of the Guatemalan economy, by making it denote the economy. They did so by borrowing the denotation of the linguistic expression 'Guatemalan economy' (and parts of the machine denoted parts of the economy in virtue of the stickers next to the tanks). The machine instantiates a number of hydraulic machine properties, and via I it I-instantiates a number of economy properties. Some of them – the effect that a decrease in foreign exports had on income for instance – are exemplified because they were made salient by focusing on them and making sure they were epistemically

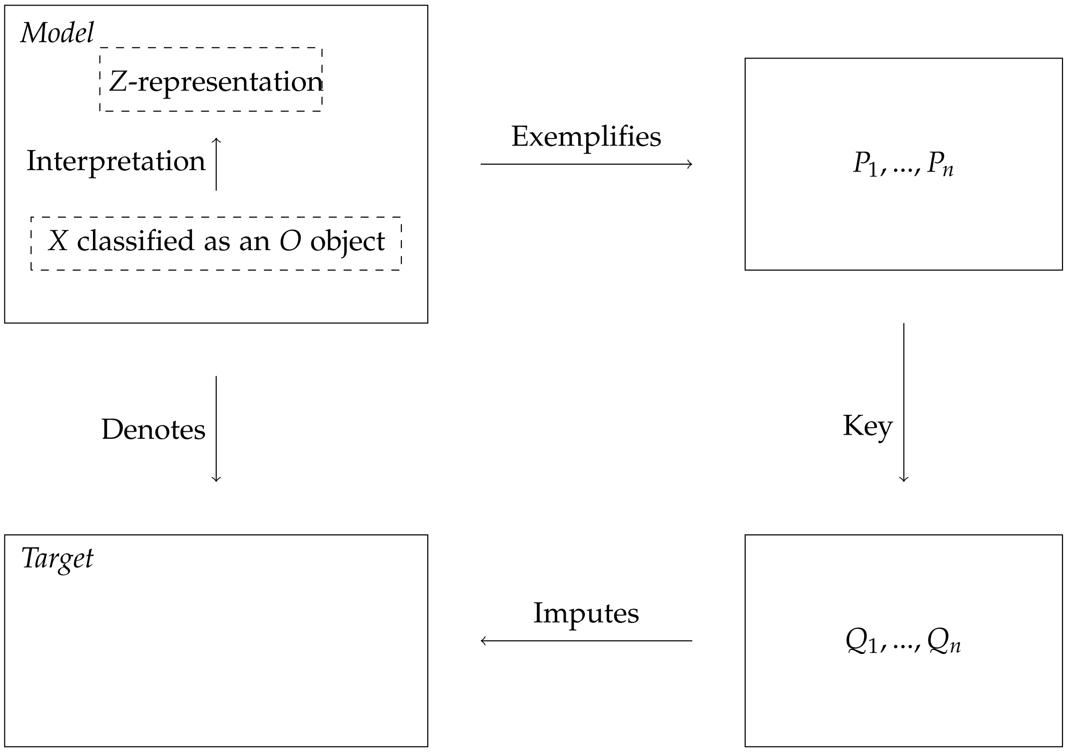

Figure 17: The *DEKI* account of scientific representation

accessible. In this instance we can suppose that the economists were so confident in the underlying mathematical theory that they used an interval-valued key, which moved from specific changes in value for income before and after the change in foreign exports to values 4% around them, and imputed the result onto the Guatemalan economy.[120]

I think that this is the correct analysis of how models represent. But it bears noting that the account is stated at such a level of generality that it needs to be concretised in every particular instance of representation. In every concrete case of a model representing a target one has to specify what the *O* object is, how it is interpreted, what sort of *Z*-representation it is and what properties it exemplifies, how denotation is established, what translation key is used, and how the imputation is taking place. Depending on what kind of representation we are dealing with, these 'blanks' will be filled differently. But far from being a defect, this degree of abstractness is an advantage. Scientific modelling is an umbrella term covering a vast array of different activities in different fields, and a view that sees representations in fields as diverse as elementary particle physics, evolutionary biology, hydrology, and rational choice theory work in exactly the same way is either mistaken or too coarse to make important features visible. My own account is stated at the right level of generality: it is general enough to cover a large array of cases and yet it highlights what all instances of scientific representation have in common.

Before moving on it is worth pointing out that the ordering of the conditions is not supposed to introduce a temporal element into either scientific representation or the process of constructing the model; nor is it meant to indicate logical priorities. It needn't be the case that a model user first establish denotation, before determining which properties are exemplified by the model and only then translating them and imputing them to the target system. None of the four conditions has to be established prior to the others, and the model could exemplify the properties even before being used to represent a target by the model user. The user could equally well start off with the target system and a set of properties of interest. She could then construct an 'inverse' key associating those properties with ones that we have firmer grasp on in the context of model building. She could then construct a model that exemplifies those properties, in the appropriate manner under the appropriate interpretation, before taking the model and establishing the denotation relation between it and the target. Such a process is not ruled out by the above conditions. The account does not function as a diachronic account of scientific representation. It is synchronic: as long as the

120 Alternatively, if the economists were more sceptical about the precise numbers to impute, they could just have imputed a directional correlation.

conditions are met, in whatever order, a model represents its target system as thus or so.

10

APPLICATION AND ANALYSIS

To further illustrate the account of scientific representation I am proposing it is worth applying it to further instances of scientific representation. In this chapter I investigate how the account explains how Kendrew's 'sausage model' represented myoglobin as having a particular tertiary structure, and how Volterra's model represented fish in the Adriatic Sea. I take it that Bak et al.'s sand pile model would receive an analogous treatment to Volterra's so I do not provide a detailed case study of SOC models here. I then explain how my account fares with respect to the problems and desiderata laid out in Chapter 1, and compare the account with those discussed previously.

10.1 MODELS OF MOLECULES

Recall the scenario. It's 1957 and John Kendrew threads a 'sausage' of plasticine through a system of vertical rods. He looks at the resulting shape and observes that it twists and turns back on itself forming a pattern that is highly difficult to explain in English. He concludes that myoglobin, a globular protein smaller than haemoglobin that is found in many animal cells, has a particular tertiary structure, and remarks that 'the arrangement seems to be almost totally lacking in the kind of regularities which one instinctively anticipates, and is more complicated than has been predicated by any theory of protein structure' (1958, 665).

Proteins are chains of amino acids covalently bonded together by peptide bonds. Specifying their primary structure simply requires specifying the sequence of amino acids. Specifying their secondary structure requires describing the three-dimensional form of local segments of the chain (a common example is an α-helix: a right handed spiral). Specifying their tertiary structure requires describing how the local segments of the chain are folded together in three-dimensional space.[121] The chemical and physical properties of a protein depend on all three of these types of structure.

121 Molecules made up of multiple polypeptide chains have a quaternary structure as well.

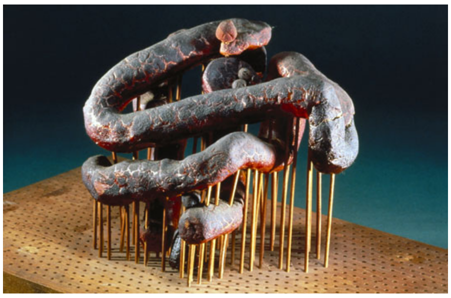

Figure 18: Kendrew's 'sausage model' of myoglobin (Science Museum)[122]

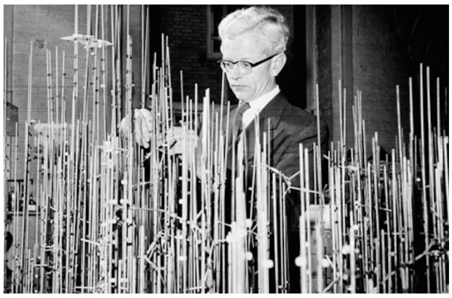

Figure 19: Kendrew with his 'forest of rods' model (MRC Lab of Molecular Biology)[123]

Determining tertiary structure is a difficult task, and Kendrew's work on myoglobins led him to win the 1962 Nobel Prize in chemistry. Kendrew's investigation contained two important elements. Firstly, through the process of X-ray diffraction and complex calculations on the results, he and his team in the Cavendish Laboratory at the University of Cambridge were able to determine the electron density throughout the molecule.

Using these data, Kendrew set about building a physical 'sausage' model of a myoglobin molecule. The first model was built at a resolution of 6Å (Figure 18). It consisted of a series of vertical supporting rods, around which was wrapped a sausage of plasticine, which twisted, turned, and folded back on itself. The sausage model was constructed from the electron density data. But it wasn't a simple summary of these data. The model provided epistemic access to the tertiary structure of the molecule in a way that the electron density alone could not (de Chadarevian, 2004, 344).

Describing this structure is very difficult. From observing his model, Kendrew states the following:

> '[The polypeptide chain] is folded to form a flat disk of dimensions about 43Å x 35Å x 23Å. Within the disk the chains pursue a complicated course turning at large angles and generally behaving so irregularly that it is difficult to describe the arrangement in simple terms; but we note the strong tendency for neighbouring chains to lie 8-10Å apart in spite of the irregularity. One might loosely say that the molecule consists of two layers of chains, the predominant directions of which are nearly at right angles in the two layers' (Kendrew et al., 1958, 665).

121 Available here http://www.sciencemuseum.org.uk/HoMImages/Components/799/79987_3.png.

122 Available here https://commons.wikimedia.org/wiki/File:KendrewMyoglobin.jpg#/media/File:KendrewMyoglobin.jpg.

The model played a serious role in both discovering this structure, and communicating it by means of photographs. Moreover, in the models he developed later, like the so-called 'forest of rods' model (Figure 19) representing myoglobin at a 2Å resolution, Kendrew used observations from plumb lines on the model to deduce specific atomic coordinates of the atoms in the polypeptide chain.[124,125] Although relatively simple, models such as this played a significant role in developing our understanding of the structure of proteins, and their importance should not be underestimated.

In virtue of what does the sausage model – a system of rods and a folded length of plasticine – represent its target: myoglobin, a protein molecule found in muscle tissue? Moreover, what is it about the model that allows us to learn about myoglobin by investigating the model? Here we can apply the account of representation discussed in the previous chapter to answer this question: the plasticine object is interpreted in such a way as to become a protein-representation, which represents myoglobin as having a particular shape in three-dimensional space. So, let's see how the conditions are met.

The first question to address is what makes the plasticine-wrapped-around-rods-system a *Z*-representation, in this instance a protein-representation (or 'protein model' for short)? What distinguishes it from the result of a child aimlessly playing with plasticine? As in the case of the Phillips-Newlyn machine, the point that needs emphasising is that the model is itself a physical object, with an associated set of properties: being such and such a size; being made out of such and such materials; being folded in such and such a way; and so on. The material properties of the model qua physical object are important to how the object functions as a model. The model can thus be characterised as an *O* object, in this case a plasticine-around-rods object with a particular shape and size in three dimensional space.

In order to turn the *O* object into a *Z*-representation, there needs to be an interpretation *I* that associates some properties of the object (collected into a set of *O* properties $\mathcal{O}$) with properties of proteins (collected into a set of protein properties $\mathcal{Z}$). For example, the plasticine sausage is associated with an amino acid chain, and the shape of the sausage is associated with the shape of the chain, and so on. Notice that not all properties of the object are collected into $\mathcal{O}$. Although the rods played a vital role in keeping the plasticine in the shape that it is, they do not correspond to any protein property. Similarly, not all protein properties are collected into $\mathcal{Z}$. The model, for instance, does not have the resources to tell us anything about the primary structure

124 For further examples about the use of these models as research tools see de Chadarevian (2004, 345-349).

125 A full specification of the tertiary structure of protein would require specifying an atomic coordinate for every atom in the polypeptide chain (and side chains). Although the 'sausage model' doesn't quite provide this, it was the closest anyone had come at the time.

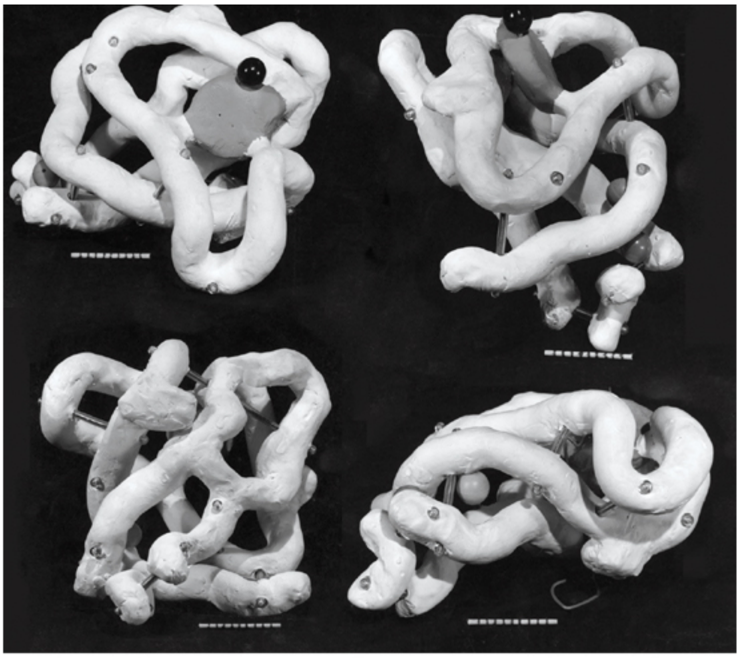

Figure 20: A photograph of the model in *Nature* (Kendrew et al., 1958, 665)

of myoglobin (since the plasticine rod is a continuous chain and thus lacks the ability to represent different amino acids in the chain).

Quantitative properties have to be dealt with carefully again here. In this example we have quantitative properties like the angles of curves in the length of plasticine and its overall length. Here, I has to include a function that associates the values of the appropriate quantitative properties in $\mathcal{O}$ with values of quantitative properties in $\mathcal{Z}$. In this case, the function associating the bend angles is the identity function. But the function associating lengths in the model with lengths in proteins is a scaling function determined by the size of the model and our knowledge about the size of proteins. Kendrew makes this scale explicit in his picture of the model, which includes a series of white dots that are specified to be 1Å apart (Figure 20). Obviously the white dots themselves are not 1Å apart, but with the interpretation in place we can understand what he means.

In the research context in question, where Kendrew was trying to determine the tertiary structure of myoglobin, the properties that the model exemplifies are those that relate to its shape in three dimensional space. It exemplifies being a 300Å long polypeptide chain folded to form a flat disk of dimensions about 43Å x 35Å x 23Å, whose chains within the disk turn at large angles, that neighbouring chains lie 8-10 Å apart, and consisting of two layers of chains with a heme group (the site where oxygen bonds to the molecule) connected in at least four distinct locations. The model only, strictly speaking, instantiates some of these properties (angles between the chains, number of connections...). The rest it *I*-instantiates.

As in the case of the Phillips-Newlyn machine, *I*-instantiation can be used to establish exemplification, since what's important about exemplification is that the exemplified properties are made epistemically salient. And in the case of the sausage model, the property of being a polypeptide chain, that is 300Å long, and is folded in such and such a way are both salient, and epistemically accessible.

The final thing to emphasise about how exemplification relates to being a *Z*-representation is that there is no reason to suppose that all properties in $\mathcal{Z}$ are exemplified. The interpretation provides a 'menu' of properties, which context selects as salient in certain cases. For example, in a context where Kendrew only cared about the protein length, rather than tertiary structure, the latter would not be exemplified despite being covered by the interpretation.

So, Kendrew's 'sausage' model is a pair $M = \langle O, I \rangle$, where O is a plasticine-around-rods model, and I an interpretation function mapping properties of of the model qua object to protein properties in a tightly constrained way. Notice that, other than the specification that it is protein properties that occur in the codomain of I, no reference here is made to any target system. The next step in establishing that the M represents myoglobin is that M denote myoglobin. The important thing to notice here is that the sausage model differs in an interesting way from the Phillips-Newlyn machine. In the latter case we had a *particular* target system; the Guatemalan economy. In the former case M features much more like a predicate; it denotes all and every myoglobin molecule. Here recall the way that I have been using denotation throughout this thesis. Following Goodman and Elgin, it doesn't need to be 'singular'; just as a name denotes its bearer, a predicate denotes everything in its extension, and the 'sausage' model denotes all myoglobin molecules. This better be the case in any account of scientific representation that includes denotation as a necessary condition, since numerous models denote multiple target systems.

What establishes denotation in the case of the 'sausage' model is a more troubling question. Again, in my opinion it is natural to think that the denotation of the model is parasitic on the denotation of the word 'myoglobin'; the model denotes whatever the word denotes. However, there is an interesting second option available as well. In this case, the model was constructed from the electron density data, which itself was the result of a measurement procedure on a sample of whale myoglobin (Kendrew et al., 1958).[126] If one were of a Kripkean bent regarding denotation then one could attempt

126 The measurement procedure and mathematical manipulation of the result was a complex process that played a highly important role in Kendrew's investigation.

to develop an account according to which the myoglobin model represents myoglobin in virtue of a causal chain that holds between myoglobin and the model; a causal chain that goes via the electron density data and Kendrew's use of that data in constructing his model. As before, I think this would be a useful avenue of future research (useful, not just from the perspective of the philosophy of science – 'how do models denote?' – but also from the perspective of philosophers more generally – 'once denotation is liberalised outside of the linguistic framework, can we learn anything about the roots of denotation?').

The final steps in establishing that the model represents the protein as thus or so, is that the exemplified properties of the model are keyed up with a a set of properties, and the latter are imputed onto the target system. In the case of the plasticine model, the key allows some flexibility between the properties directly exemplified by the protein model and those that are imputed onto myoglobin itself.

For example, although the length of plasticine in the model is a rope of uniform width throughout the model, Kendrew explicitly imputed a different property onto the molecule 'as it is at corners that the chain must lose the tightly packed configuration that makes it visible at this resolution' and proposed that perhaps 70% of the chain was an α-helix whilst the rest was fully extended (Kendrew et al., 1958, 665). Likewise, it is unlikely that Kendrew was confident that the 43Å x 35Å x 23Å dimensions exactly corresponded to the dimensions of the molecule. There were clear margins for error in the process leading to the construction of the molecule, so it is more likely that something like 'being a flat disk of 43Å$\pm$10% x 35Å$\pm$10% x 23Å$\pm$10% dimensions' was imputed.

With the aforementioned discussion in mind, how the plasticine-around-rods object came to represent myoglobin as having the particular tertiary structure that it does can now be tied together. The plasticine-around-rods system X is endowed with an interpretation I associating plasticine-properties with protein properties. The interpretation associates quantitative properties of the model, like angles between bends, or the length of the chain, with quantitative protein properties, like angles between bends, and the length of a polypeptide chain. X and I together form a model, a protein-representation. The model denotes myoglobin, which makes it a representation-of myoglobin. The model also exemplifies protein properties in virtue of the research context selecting them as salient, for instance consisting of two layers of chains (P_1), forming a flat disk of dimensions about 43Å x 35Å x 23Å (P_2), and having a uniform configuration throughout (P_3). These properties are related to other properties with key K: identity in case of P_1, applying with a tolerance threshold of around 10% in the case of P_2 and only applying to straight lengths of the polypeptide chain in the case of P_3. So the model imputes consisting of two layers of chains (Q_1); being a flat disk of dimensions 43Å±10% x 35Å±10% x 23Å±10%; and having a uniform configuration in only 70% of the chain (Q_3) to the target. These conditions establish how the plasticine model (M), represents myoglobin (T), as being a protein with such and such a tertiary structure (Z).

10.2 MODELS OF FISH

Recall the scenario. Volterra is interested in the change in relative fish populations in the Adriatic Sea. Before the First World War there was a certain proportion of predator to prey fish sold in the Italian fish markets, and he supposes that the proportions in the markets match the proportions in the sea itself. During the war fishing decreased significantly, which led to a higher proportion of predator to prey (i.e. light fishing favoured the predators). After the war, where fishing increased to pre-war levels, the proportion of predator to prey returned to its pre-war level.[127] He writes down the following coupled non-linear differential equations:

$$\frac{dV}{dt} = \alpha V - (\beta V)P \tag{1}$$

$$\frac{dP}{dt} = \gamma(\beta V)P - \delta P \tag{2}$$

He lets t denote time; V denote the size of the prey population; P the size of the predator population; α the intrinsic growth rate of the prey; δ the intrinsic death rate of the predators; and β and γ denote prey capture rate and rate at which each predator converts captured prey into more predator births respectively. He notices that the equations have two solutions where $\frac{dV}{dt} = \frac{dP}{dt} = 0$. First where $P = V = 0$, and second where $\alpha V = (\beta V)P$ and $\delta P = \gamma(\beta V)P$. The second is his primary interest. He calculates that although it is unstable, it corresponds to the mean values of P and V over an indefinitely large amounts of t. Letting $\hat{P}$ and $\hat{V}$ denote these values and $\rho = \frac{\hat{P}}{\hat{V}}$ he derives the following equation:

$$\rho = \frac{\alpha\gamma}{\delta} \tag{3}$$

He argues that heavy fishing is a general biocide, i.e. something that harms both predators and prey, and corresponds to lower values of α (it is harmful to the prey so it lowers their birth rate) and higher values of δ (it is harmful to the predators so it increases their death rate), which means, by equation 3, that heavy fishing corresponds

127 See Volterra (1926, 1928) for his original discussion. See Weisberg (2007); Weisberg and Reisman (2008); Weisberg (2013) for useful philosophical discussions that I draw upon here.

to lower values of ρ. Given that ρ is defined as the ratio of long term average size of of the predator population to the long term average size of the prey population, lower values of ρ mean higher relative size of prey population with respect to predator population. He thus concludes that pre-First World War fishing activity led to higher prey to predator ratios, and thus more prey in the Adriatic fish markets, and the decrease in fishing activity associated with the First World War led to higher predator to prey ratios.

The first thing to get clear in this instance is what the model is. In the previous cases the the model object, $\mathcal{O}$, was a physical object: a hydraulic machine or a length of plasticine wrapped around some rods. This isn't the case here. There was nothing that Volterra was 'holding in his hands' that represented the fish population. This example concerns a non-physical model. By and large, I want to remain non-committal about the ontological status of of such models. However there are two pertinent options available that I briefly outline, since the approach adopted has implications for how DEKI is to be understood as applied to this case. Firstly, the $\mathcal{O}$ object could be taken to be a *mathematical structure* that satisfies equations 1 and 2 above. Secondly, it could be taken to be a *fictional system* in some sense. I discuss both in turn, and then explore how the rest of DEKI's conditions could be met in each instance.

The equations above are coupled non-linear differential equations. Equation 1 tells us how the value of V changes with respect to the value of t, and equation 2 tells us how the value of P changes with respect to the value of t. So V and P are both real valued functions on t, which itself takes real values. The equation is thus satisfied by a mathematical structure $\mathcal{S}$.[128] Figure 21 shows how the functions V and P evolve through time.

The equations also have a phase space, which is constructed by eliminating the independent variable t from equations 1 and 2. This is done as follows. Since V is a function whose value depends only on the value of t and P, and P is a function whose value depends only on t and V we can use the chain rule to derive the following:

$$\frac{dP}{dV} = \frac{\frac{dP}{dt}}{\frac{dV}{dt}} = \frac{\gamma(\beta V)P - \delta P}{\alpha V - (\beta V)P} = -\frac{P}{V}\frac{(\gamma(\beta V) - \delta)}{(\beta P - \alpha)} \tag{4}$$

The solutions to equation 4 are the closed curves illustrated in figure 22. Each curve in the phase space corresponds to how the values of V and P change for each set of the initial conditions, i.e. initial values of V and P in equation 4 (keeping the parameters fixed). We can treat this phase space as a mathematical object, where points in the

128 I do not have the space to spell out the details of the sorts of structure that satisfy differential equations. For my current purposes it suffices that they are mathematical objects, but see Balzer et al. (1987) for useful examples of these sorts of structures.

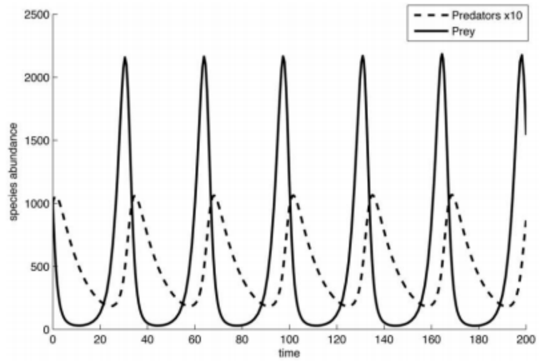

Figure 21: Oscillations in the Lokta-Volterra model (Weisberg and Reisman, 2008, 112)

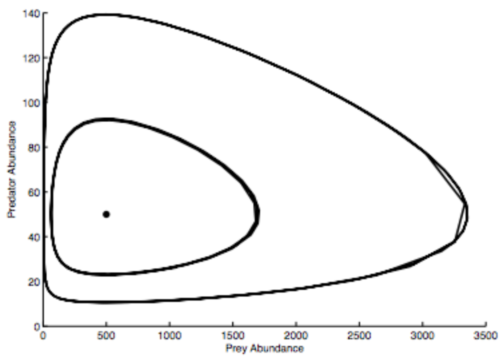

Figure 22: Phase space of the Lokta-Volterra model (Weisberg and Reisman, 2008, 133)

space are pairs of real numbers; points on the curves are pairs of values of V and P (for each value of t) compatible with equation 4; and a function is defined on the space relating each point with the 'next' one (as t increases). One option then, is to identify the model object X of the Lotka-Volterra model with the phase space represented in figure 22 (or the set theoretic structure $\mathcal{S}$above; the difference between the state-space, or phase-space, and set-theoretic approach doesn't matter for my current purposes). It is important to note here that, if this approach is taken, then the only properties instantiated by the model object X are, strictly speaking, mathematical ones, like 'a decrease in α and increase in δ entails a decrease in ρ, i.e. a decrease in the relative value of $\hat{P}$ with respect to $\hat{V}$'.[129]

An alternative option is suggested when we consider Volterra's own presentation of his model. Accompanying his presentation of the equations he writes things like the following:

> 'The first case I have considered is that of two associated species, of which one, finding sufficient food in its environment, would multiply indefinitely when left to itself, while the other would perish for lack of nourishment if left alone; but the second feed upon the first so the two species can co-exist together' (1926, 558).
>
> 'If we try to destroy individuals of both species uniformly and proportionally to their number, the average number of the *eaten* species grows and the average number of the eating species decreases ... But increased protection of the eaten species increases the average number of both' (1926, 558-559).[130]
>
> 'Let [V] and [P] be the numbers of individuals of the two species. Let [α] represent the coefficient of increase which the first would have if the other did not exist. Let us suppose that the second would die out because of lack of food if it were alone ... ' (1928, 9).

The equations are explicitly accompanied by a description of what we might take to be an 'imaginary' target environment: containing only two types of species; who multiply and die continuously; where there is unlimited food for the prey; and so on.

If we interpret these descriptions at face value, then rather than taking the model object to be a mathematical structure that satisfies the equations, we could instead

129 Strictly speaking, the mathematical object that instantiates this property will be a higher dimensional phase space corresponding to allowing the values of the parameters to vary, as well as the initial values of V and P.

130 This is a statement of the so-called 'Volterra Principle' as derived from equation 3 above.

take X to be an imaginary system, containing predators and prey such that the idealisation assumptions introduced by Volterra are 'true', in some sense, of them. This approach of identifying models with imaginary systems was originally suggested by Godfrey-Smith (2006) who compared models to works of fiction. Cashing out the details of the ontological status of such systems is a significant task. Frigg (2010a,b) provides a careful analysis of how Walton's account of fictions can be developed in the context of an indirect view of scientific representation (as opposed to Levy and Toon's use of Walton's framework in Direct View). According to Frigg, in the case of non-physical models, the equations and descriptions presented by scientists, like those written by Volterra above, should be seen as props in a Waltonian game of make-believe. Principles of generation prescribe the reader to imagine a fictional system such that statements like 'the intrinsic growth rate of the prey is α' are 'true' in the relevant game of make-believe. Under this approach, we can allow that the model system X instantiates *physical properties* like 'a general biocide favours the prey'.[131]

So we have at least two options, the model object X could be identified with a mathematical structure or with a fictional system. In the former case, the object instantiates purely mathematical properties; in the latter, the object instantiates physical ones as well. The next question is what turns the model object into a Z-representation. According to DEKI, the object must be imbued with an interpretation I that maps properties of the model system with, in the case under consideration, properties of predator-prey populations. Let's now consider what the interpretation would look like under each of the aforementioned ways of identifying $\mathcal{O}$.

If X is taken to be a mathematical object then the relevant properties it instantiates are mathematical. The interpretation then maps mathematical properties, like 'a decrease in α and increase in δ entails a decrease in ρ, i.e. a decrease in the relative value of $\hat{P}$ with respect to the value of $\hat{V}$' to properties of predator-prey systems like 'a general biocide favours the prey'. In this case, since the mathematical property is a *quantitative* one, specific changes to the value of α and δ will yield specific changes to the value of ρ, the interpretation also includes a function mapping these values onto values of the predator-prey population property, something for the form 'a biocide that reduces the prey birth rate and increases the predator death rate by y, will increase the ratio of prey to predator populations by x'. Specific changes to the values of α and δ will be mapped to specific changes to the value of y, and the implied changes to the value of ρ will be mapped to values of x. Similarly, the mathematical property of the

131 I don't mean to imply that Frigg's approach is the only way we can make sense of a non-physical model nevertheless instantiating physical properties. Contessa (2010) provides an account of the ontology of models in terms possible objects. Thomasson (forthcoming) provides an account where they are epistemic artifacts.

solutions to equations 1 and 2 being periodic, and out of phase, as displayed in figure 21, will be mapped to physical properties relating to how the size of the predator and prey fish populations change with respect to one another through time. Thus, under this construal of the model object, the interpretation function will map mathematical properties to physical ones.[132]

If, on the other hand, X is taken to be a fictional system that can itself instantiate physical properties, then the interpretation function will be very different. Here many of the physical properties of the fictional system will be mapped to themselves by the interpretation function. The role of the interpretation function will be to specify which of the many properties of the fictional system could be potentially relevant in some research scenario. For example, the property of 'being originally imagined by Volterra' will not be included as an argument of the interpretation function, but the property of 'a general biocide favours the prey' (along with more specific formulations of the principle) will be.

So the two options available are to define the Lotka-Volterra model as either the pair $\langle X, I \rangle$, where X is a mathematical object and I an interpretation function mapping mathematical properties to predator-prey system properties, or alternatively as a pair $\langle X', I' \rangle$, where X' is a fictional predator-prey system and I' highlights which properties of the X' could be potentially useful in a research context, and maps them to themselves as an 'identity interpretation'. The outcome of both is that the value of I and I' provides the 'menu' of Z properties available to be exemplified for the rest of the DEKI conditions to be met. So I don't think anything internal to DEKI favours one approach over the other.

External considerations might tell them apart however. For instance, if one hoped for a austere ontology – Quine's (1948) 'desert landscape' – then this may provide independent reasons to favour one approach over the other. I take it that philosophers of a Quinean bent are happier (even if reluctantly) to accept the existence of mathematical structures, than they are to accept the existence of fictional systems.

Moreover, one attribute that the mathematical approach has over the fictional one is that it provides a clear explanation of cases where the same mathematical model is been used to represent different types of target systems. For example, the mathematics that underpins the Lotka-Volterra model have been reinterpreted as an economic-model, where V denotes the employment and P denotes workers' share of national output (Goodwin, 1982). In the context of DEKI, this amounts to constructing an

132 It does bear nothing that there is an alternative option available. X could be taken to be a mathematical structure that, strictly speaking, exemplifies only mathematical properties. The key then translates these into physical proprieties to be imputed onto the target system. I discuss this option more in the following subsection.

alternative model that consists of the same mathematical object as that used in the predator-prey case, but introducing an alternative interpretation function. The mathematical construal of model objects allows us to clearly see how the same object can be used in a variety of different models, by varying the interpretation function, something that the fictional approach has trouble making sense of.

On the other hand, although the fictionalist approach has difficulty dealing with the applicability of mathematics, it's plausible that it better matches scientific practice in other respects. For example, it accords a systematic role for Volterra's descriptions accompanying the mathematical equations, and in addition, can readily accommodate non-mathematical models (Godfrey-Smith (2006) points out that many models in biology are not as mathematised as the Lotka-Volterra model). It could also be argued that there are some model 'reinterpretations', in contrast to the example in the previous paragraph, where the story accompanying the model stays the same whereas the mathematics changes. For example, we can think of the two-body model of a celestial orbit as a fictional system involving a planet orbiting the sun. The mathematics used to describe the orbit of the planet can change, Newtonian to relativistic mechanics for example, even though the system stays the same.

Regardless of how these details are worked out, we are left with a model that includes an interpretation function that delivers predator-prey properties as values. The next questions to address when establishing in virtue of what the model represents fish in the Adriatic Sea are the denotation, exemplification, keying up, and imputation, conditions in DEKI. Let's now turn to them.

I think that what establishes denotation in the Lotka-Volterra case can again be understood as parasitic on the denotation of the phrase 'fish populations in the Adriatic Sea', or if the model is taken to have multiple target systems – multiple predator-prey systems – then it is parasitic on the phrase 'predator-prey system'. Volterra makes it explicit that his model was originally developed to explain D'Ancona's observations about the proportion of predator to prey in the Adriatic fish markets, and states that his model demonstrates that 'that Man in fisheries, by disturbing the natural condition of proportion of two species, one of which feeds upon the other, causes diminution in the quantity of the species that eats the other, and an increase in the species fed upon' (1928, 4). And, as in the case of the Phillips-Newlyn machine denoting the Guatemalan economy, there are multiple denotation relationships to consider (for example, Volterra writes 'let t denote time', and so on with respect to the other variables and parameters). Again, I don't take these observations as firmly demonstrating what establishes the denotation relation between models and their targets, and I take it that this is an important avenue for future research.

The interpretation function delivers a number of predator-prey properties. The specific properties exemplified by the model are sensitive to the research question at hand. In the first instance, the model exemplifies the Volterra Principle, 'a biocide that reduces the prey birth rate and increases the predator death rate by y, will increase the ratio of prey to predator populations by x, for some specific values of x and y' (P_1), since this corresponds to the empirical observation that Volterra was trying to explain. But Volterra also notes multiple different properties exemplified by his model, including, for example, the specific oscillations of the predator and prey populations (P_2), and that, for fixed parameters, the average size of each of the populations tend towards specific values, regardless of their initial values (P_3). Notice that some of the properties covered by the interpretation function are not exemplified. For example, for a given set of initial conditions and values of the parameters in the equations 2 and 1, there exists specific values of V and P for a given value of t. This could prove useful if one wanted to accurately predict the actual size of a given population at a given time. But from Volterra's perspective, where he was not interested in predicting specific population sizes, this property is not exemplified.

The key involved in this case is similar to the key involved in the Phillips-Newlyn machine case. If Volterra were very confident in his model, he could have taken the exemplified property P_1, 'a biocide that reduces the prey birth rate and increases the predator death rate by y, will increase the ratio of prey to predator populations by x' and imputed that directly onto the target system (which would amount to a key that mapped P_1 to $Q_1 = P_1$), and similarly for the specific constant values that the average predator and prey populations sizes tended towards ($P_2 = Q_2$), and the specific oscillations of the populations ($P_3 = Q_3$). However, it is clear from his writing that Volterra didn't impute these specific properties, but rather imputed the more abstract properties, which he calls 'laws' (1928, 558), such as: 'a general biocide favours the prey' (Q_1); average predator and prey populations tend towards constant values regardless of their initial sizes (Q_2); and 'predator and prey populations oscillate out of phase' (Q_3). So we have a key which takes real-valued, specific, properties exemplified by the model, to more general correlations or tendency properties which were then imputed onto the target system(s).

We are now in a position to explain how the Lotka-Volterra model represented fish in the Adriatic Sea in terms of the DEKI conditions. Either the model consisted of a mathematical structure, with an interpretation function taking mathematical properties to predator-prey properties, or the model consisted of a fictional predator-prey system with the interpretation function highlighting ones that that might be useful in a representational context. Either way the the interpretation function yields a set of predator-prey properties instantiated, or I-instantiated, by the model object. In the context in which Volterra introduced his model he was interested in how different levels of fishing related to different proportions of predator to prey fish in the Adriatic Sea. As such, by stipulating that the model concerned these fish, he established that the model denoted them. In such a context, the property exemplified by the model was the following: 'a biocide that reduces the prey birth rate and increases the predator death rate by y, will increase the ratio of prey to predator populations by x' (P_1). Since Volterra wasn't too interested in the specific values of x and y, he used a key that took this to the more general property 'a general biocide favours the prey' (Q_1). He then imputed this property onto the fish in the Adriatic Sea. In this instance, since this matched D'Ancona's empirical data concerning the relative numbers of predator and prey fish sold in the Adriatic fisheries, the model was an *accurate* representation, at least with respect to the purpose it was used for.

10.3 analysing my account

I hope that I have demonstrated the value of my account of scientific representation in understanding its instances. I now turn to how it answers the questions discussed in Chapter 1 before comparing it to some of the accounts discussed in Part ii.

10.3.1 *Problems and desiderata*

Recall the problems of scientific representation I distinguished earlier:

er-problem: Provide necessary and sufficient conditions of a model representing its target.

demarcation problem: Demarcate scientific models from other epistemic representations (or motivate why they should not be so demarcated).

problem of style: Account for the fact that different scientific models seem to represent their targets in different ways.

standards of accuracy: Provide standards of accurate epistemic representation.

problem of ontology: Help us understand what (specifically non-physical) models are, ontologically speaking.

And the conditions of adequacy I provided:

surrogative reasoning condition: Account for the fact that models, and epistemic representations more generally, can be used to attempt to learn about their targets.

POSSIBILITY OF MISREPRESENTATION: Allow for the fact that a model may misrepresent its target system, i.e. represent it inaccurately.

APPLICABILITY OF MATHEMATICS: Help us understand how mathematical models represent their target systems.

The account is explicitly designed as an answer to the ER-problem. It meets the surrogative reasoning condition in a straightforward way. When a model user imputes a property Q_i onto the target the user provides a hypothesis concerning the target. The hypothesis can be coarse grained ('T has Q_i'), or more fine grained ('a specific part of T has Q_i') depending on the denotation relationships that hold (if a part of M, m, denotes a part of T, t, and m exemplifies a property that the key maps to Q_i, then Q_i can be imputed onto t, rather than T more generally.

The account, as stated, has little to say about the standards of accuracy, but I take it that these are pragmatic features in the context of using a model. Sometimes the hypotheses may be required to be perfectly accurate, at others, it may allow them to be 'true enough' (Elgin, 2004). I think this provides an interesting avenue for future research, but I want to point out a pertinent point here. Both Elgin (2004) and Teller (2001) have pushed to reorient what we require of our scientific representations. According to the former, as long as what they tell us is 'true enough', then we can arrive at understanding just fine. According to the latter, an expectation that models should be 'perfect' will never be met; the only perfect model of the world is the world itself.

I think that, by and large, I am in agreement with both of them. But I want to stress that as stated, my account of representation is independent of such a position. The introduction of a key allows for models that only provide hypotheses about their target systems that are 'true enough', but it is also compatible with further refining the key in such a way that everything a model tells us about its target comes out as accurate. In fact, keys provide a way of accounting for the motivations of Elgin and Teller's positions, without giving up the idea that models can be accurate representations in virtue of licencing truths (not Elgin's (2004) 'felicitous falsehoods') about their targets. The idea is the following. Teller's (2001) argument against the 'perfect model model' is that we should not expect our scientific models to tell us *everything* about target systems, in such a way that *all* of the properties of a model perfectly *match* the properties of its targets. This is clearly not required by DEKI. Not all of the properties of a target system are represented by the model, and those that are, needn't be related to model properties via identity. And although I agree with Teller that striving towards perfect models is a misguided enterprise, they are not ruled out by my account of scientific representation.

Elgin (2004) argues against the traditional epistemological position where understanding requires truth, and draws on examples from science where scientific models provide us with understanding about their target systems by licencing felicitous falsehoods about them. But if models also come with keys then there is another way that they can provide understanding without invoking falsehoods. What I have in mind is where the key takes a perfectly precise property of the model, and maps it to a vaguer property to be imputed onto a target. For example, if it is the case in the model that '[a] freely falling body falls at a rate of 32 ft./sec' (Elgin, 2004) the key needn't map this identically onto the claim that 'a freely falling body in the target system falls at a rate of 32.ft/second'. Rather, the key can map it to vaguer property, such as 'a freely falling body in the target system will fall at a rate of $32 \pm \epsilon$.ft/second', where ϵ is some number small enough not to matter, given the purposes of using the model. Whether or not the idea of relaxing precise properties of the model to vaguer properties to be imputed onto the target will work in all of the cases that Elgin has in mind remains an open question. Discussing this in any more detail goes beyond my current purposes, and as I noted above, keys need to be investigated on a case-by-case basis.

However this question is resolved, the account allows for misrepresentation in at least two places. It needn't be the case that target systems possess any of the imputed properties. M can represent T as possessing properties $Q_1, ..., Q_m$ and T might not instantiate a single of them. If M represents T as having properties that it doesn't have it misrepresents it. The other place where 'misrepresentation' can enter is denotation. Denotation can fail in various ways – a representation can purportedly denote a target that does not exist or it can denote the wrong target – such a failure would be an instance of missed (or failed) representation, rather than misrepresentation proper (which requires a representation relation, albeit an inaccurate one).

The abstract character also accounts for different styles of representation. Style, on this account, is not a monolithic concept; instead it has several dimensions. Firstly, different O objects can be chosen. So we can speak of, say, the checkerboard style (in the case of Schelling style models of social segregation) and of the cellular automaton style. In each case a specific kind of object has been chosen for various modelling purposes. Secondly, the notion of an interpretation allows us to talk about how closely connected the properties of the model are to those that the object I-instantiates. The Phillips-Newlyn machine is an economy-representation, despite the fact that hydraulic properties are completely unrelated to economy properties (prior the introduction of the interpretation). In other cases, scale models for example, the interpretations will connect much closely connected properties. Thirdly, different types of keys could be used to characterise different styles. In some instances the key might be the identity key, which would amount to a style of modelling that aims to construct replicas of target systems; in other cases the key incorporates different kinds of idealisations, approximations, or abstractions, which gives rise to idealisation, approximation, and abstraction keys. Different keys may be associated with entirely different representational styles.

With respect to the demarcation problem, it should be obvious that by invoking the notion of representation-as, my inclination is that there is no sharp distinction between scientific and non-scientific epistemic representations. Just as the Phillips-Newlyn machine represented the Guatemalan economy as having particular properties, a map represents its target terrain as having such and such features, and the statue of Schwarzenegger represents him as being in a particular pose with a particular musculature. However, if a distinction is to be made between scientific and non-scientific epistemic representations, then the account would allow for it in multiple places. Just like the problem of style, its abstract nature would allow for a distinction to be drawn between different types of epistemic representations in terms of the kinds of objects, interpretation functions, and keys used. It's plausible that the sorts of interpretation functions and keys used in the case of scientific representation are more constrained than those used in that case of non-scientific epistemic representation. For example, the requirement that interpretation functions associate quantitative properties with quantitative properties, and include a function specifying the relation between their values, seems distinctly scientific.

Many details in the account still need to be spelt out. But I think the most interesting way to develop the account into a complete philosophical package is to focus on the problem of ontology and the applicability of mathematics. The discussion of the Lotka-Volterra in the previous subsection indicated what I have in mind. The model object X, in the model $\langle X, I \rangle$, could be taken to be a mathematical structure. Then, according to DEKI, it could I-instantiate physical properties to be imputed onto the target system. Thus, the applicability of mathematics depends on model users being able to interpret the mathematical properties of mathematical objects as physical. This requires further investigation into how interpretation functions can map mathematical properties to physical properties. Alternatively, if X is taken to be a fictional system, then we need an account of the ontological status of fictional systems, and importantly, an account that allows us to make sense of the fact that they instantiate physical properties. I take it that Frigg's (2010a; 2010b) approach is a step in the right direction here, but it does bear noting that it is not without its problems (most pertinently, the deflationary aspect of the Waltonian account, that fictions are *just* products of scientists' imaginations, coupled with the requirement of DEKI that models instantiate properties (*cf.* Levy, 2015)).

10.3.2 *How it compares to others*

I mentioned in Chapter 9 that the use of an identity key would allow model users to impute a model's exemplified properties onto a target system. It's worth expanding on this claim in a little more detail, as well as examining how my account compares to the others discussed in Part ii.

Recall that according to SIMILARITY 3 a model M represents a target T if and only if a model user provides a theoretical hypotheses specifying that M and T are similar in the relevant respects and to the relevant degree. I argued that not all cases of scientific representation work in this way, and that in many cases, the accuracy of a scientific representation does not turn on such a similarity. This is not to say that none do however, and these need to be captured by my account. Fortunately, such cases are easily covered by DEKI. Consider a theoretical hypothesis specifying that M and T are similar in respects P_1 through to P_n. According to my account, the respects are relevant because those are the properties exemplified by the model, and the theoretical hypothesis that specifies this similarity corresponds to using an identity key, and then imputing properties P_1 through to P_n onto the target system. If likeness is to be preferred, then the key maps P_1 through to P_n onto properties Q_1 through to Q_n, where each P_i is 'similar' to Q_i, under whatever understanding of similarity between properties is offered. Of course an account of the ontology of scientific models that allows them to exemplify physical properties would still be required (which could be done through models I-instantiating them, or if non-physical models can somehow come to instantiate physical properties, then through instantiation alone).

A more interesting comparison arises when we compare DEKI with STRUCTURALISM 2. Recall that according to the latter, a scientific model M is a mathematical structure that represents its target T in virtue of a model user specifying an appropriate morphism from M to T. This approach could be captured in the DEKI framework in at least three distinct ways. Firstly, the model object X could be taken to be a mathematical structure, as discussed above, but which exemplifies its mathematical, structural, properties. The key would then take the form of a morphism (isomorphism, homomorphism, ...) which maps the structure to an appropriately morphic structure. The model imputes this structure onto the target system.[133]

Secondly, and I think more plausibly, the key could map purely mathematical properties exemplified by the model to physical ones, and these could then be imputed onto the target. Such an approach is suggested by Elgin when she writes:

133 Such an approach would have to make sense of how a physical system could enter into a morphism with X, *cf.* the discussion in Chapter 6.

> 'Since exemplification requires instantiation, if the model is to represent the pendulum as having a certain mass, the model must have that mass. But, not being a material object, the model does not have mass. So it cannot exemplify the mass of the pendulum. This is true. Strictly, the model does not exemplify mass. Rather it exemplifies an abstract mathematical property, the magnitude of the pendulum's mass. Where models are abstract, they exemplify abstract patterns, properties, and/or relations that may be instantiated by physical target systems. It does no harm to say that they exemplify physical magnitudes. But this is to speak loosely. Strictly speaking, they exemplify mathematical (or other abstract) properties that can be instantiated physically' (2010, 8-9).

Thirdly, and I think most plausibly, the interpretation function itself could map mathematical properties to physical ones. This was the approach outlined in the previous section, and it allowed us to make sense of claims like 'predator and prey populations oscillate out of phase' despite the fact that the model is a purely mathematical entity. The model *M* would exemplify physical properties, despite the fact that nothing in it instantiates such a property. Rather, the property is *I*-instantiated, but still relevant and salient in the research context under consideration, so still exemplified.[134]

Developing any of these approaches would make significant inroads in answering the applicability of mathematics question. The central idea is that *X* can be a mathematical structure, and mathematical structures can represent physical systems in virtue of exemplifying properties (mathematical structural properties, or physical properties that it instantiates under an interpretation), which can be keyed up with other properties (if all the model user imputes is 'structure' onto the target system then the key can be a morphism, if the model user imputes physical properties then the key can relate the exemplified properties with physical properties in some rigid way). Allowing the *X* in the definition of DEKI to range over mathematical objects thus opens up interesting questions for future research. But again, it is important to bear in mind that DEKI's flexibility does not require that all cases of scientific representation function in this way; it can make room for non-mathematical models as well.

Finally then, DEKI can be compared to the accounts discussed in Chapter 8. To start with INFERENTIALISM 2. I take it that denotation is what satisfies Suárez's 'representational force' condition, and the cases he worries about, where models lack targets and thus don't denote, are better explained by them being *Z*-representations but not representations-of any target system, rather than claiming that their 'representational

134 This account of the ontology of mathematical models shares some similarities with Weisberg's notion of an 'interpreted structure' (2013, 39-43).

force points' towards a non-existing target system. Moreover, although DEKI is stated at a high level of generality, it is not a deflationary account, and thus the fact that models allow competent and informed agents to draw specific inferences regarding their targets is better explained as a symptom of the fact its conditions are met, rather than building this into the definition of scientific representation itself (*cf.* Contessa, 2007, 67).

DEKI and Interpretation share some interesting features. Firstly, both select relevant properties of the model and associate them with properties that are purportedly in the target, and secondly, both take denotation as a necessary condition on scientific representation. The main difference, I take it, is that DEKI provides an elegant account of what happens in the case of target-less models; adds more detail about how the properties are selected as relevant; and is more liberal in its understanding of how properties of models are correlated with properties to be imputed onto the target. Contessa's analytic interpretations, suitably enriched in the DEKI framework, thus drop out as a specific type of key.

Compared to DDI-ER, DEKI again takes denotation as a necessary condition on scientific representation, and we can see how agents exploit the internal dynamics of models to make demonstrations which are interpreted to be about their targets. The demonstrations are in terms of investigating the properties exemplified by a model; finding the connections between them, determining their values in specific cases, and so on, and these are interpreted to be about target systems by means of a key. Compared to t-representation, again we find denotation in common, and now we know that the facts about models concern the properties they exemplify, and the key plays a similar role in transforming these into claims to be imputed onto their targets.

So in sum, I think that DEKI captures everything that the previous accounts of scientific representation got right, whilst at the same time avoiding the difficulties facing those accounts that were discussed in Part ii. Similarity, structural or otherwise, can play a role sometimes, but not in every instance of scientific representation. Mathematical structures can function as scientific representations. Target-less models are dealt with in a way that doesn't require invoking suspicious 'fictional' or 'generalised' targets. And the inferential capacity of models is explained in terms of the conditions. Moreover, I think its abstract character – the fact that each of the conditions requires further specification in each instance of scientific representation – makes it a highly useful framework in thinking about scientific representation, and epistemic representation more generally. Great insight could be gained by thinking about the use of mathematics, or different styles of scientific modelling, with reference to how each of the conditions are satisfied.

11

CONCLUSION

I want to conclude with some programmatic suggestions about what I take to be the most promising ways of further developing the DEKI account of scientific representation, and how the account can be put to use in further philosophical research.

As should be clear from my discussions in the previous section, there remain important questions to be addressed within the DEKI framework.

Denotation. The first question is what establishes the denotational relationships between models and their targets. I cautioned against taking a descriptivist approach to this sort of denotation, and suggested that, at least in some cases, it's plausible that it is parasitic on linguistic denotation: the Phillips-Newlyn machine denotes whatever 'economy' denotes; the Lotka-Volterra model denotes whatever 'predator-prey system' denotes; and so on. It remains to be seen whether this can be a general answer to model-target denotation, and whether it suffices to account for cases where models are used to denote target systems other than the ones they denoted when there were originally introduced. Cases where the denotation of words change through time are particularly relevant in the discussions of denotation in the philosophy of language, and I expect that similar concerns will arise in the context of model denotation. I hope that a thorough investigation into how models denote would also shed light on denotation in general, and thus should be of interest to philosophers of language as well as philosophers of science.

Ontology and the Applicability of Mathematics. The second question is to firm up what the *X*s are when it comes to non-physical models. I suggested two potential answers: take them to be mathematical objects; or take them to be fictional objects. Both approaches have their pros and cons. If the former route is taken then understanding the applicability of mathematics becomes more tractable, but questions remain whether it can account for models that are not as mathematised as the Lotka-Volterra model, and whether it captures scientific practice. If the latter is taken, then understanding how mathematics gets involved, and telling a full story about the ontology of scientific fictions is required. I take it that the DEKI framework provides a novel, and potentially highly useful, framework in which to explore the connections between scientific representation, the applicability of mathematics, and the ontology of fictions.

Realism vs. Anti-realism. I motivated this thesis by arguing that without an account of scientific representation, the notions of scientific realism and anti-realism, usually phrased in terms of the truth of descriptions, remain opaque in the context of model-based science. Since I've provided DEKI as an account of scientific representation, it's natural to ask what it entails about scientific realism and anti-realism. One suggestion would be to define it in terms of the truth of the imputed claims that result from applying a model. But whether or not this linguistic notion adequately captures model representation remains unclear.

Idealisation and Standards of Accuracy. Relatedly, as noted in the previous chapter, Elgin (2004) and Teller (2001) both champion an account of scientific understanding that does not require that our models are perfectly accurate representations of the world. Such an account of understanding is compatible with, but not entailed by, the DEKI account of scientific representation. However, the question remains: do the keys used by scientists deliver claims about target systems that are, strictly speaking, true? If so, must this be at the cost of precision with respect to what is imputed? If not, should we recalibrate what we expect from our scientific representations in terms of how they contribute to our understanding of the world?

Style and Demarcation. I suggested that there was no difference in kind between scientific epistemic representation and non-scientific epistemic representation. I think that both can be accounted for within the DEKI framework. However, a question remains as to whether there is anything particularly characteristic of scientific representation, and whether different styles of scientific representation can be distinguished from one another using the DEKI conditions. As discussed below, this would require a detailed analysis of further cases, which would prove fruitful in its own right.

I think that by further investigation into the types of keys associated with various different scientific models, we will be better positioned to address the questions raised in the previous three paragraphs.

Since the notion of representation is so central to our understanding of, and interaction with, the world, I take it that DEKI not only furthers our understanding of scientific representation, but could moreover be used to illuminate additional questions in the philosophy of science. The questions I have in mind are the following:

Further cases. The DEKI conditions are deliberately stated at a general level, and require further specification in each instance. I suggest that if we are to fully understand how scientists represent the world, then a detailed study of scientific practice is required. DEKI provides the relevant questions to ask in every instance: what is the model object? How is it interpreted? Which properties does it exemplify? How is denotation established? What is the key used? Investigating cases in the DEKI framework would provide greater understanding of the cases themselves, as well as deepening our understanding of how scientific representation works in general.

The Hole argument. A recent argument in the philosophy of physics concerns whether or not two isometric Lorentzian manifolds can represent different spacetimes (Weatherall, forthcoming). Weatherall attempts to diffuse the 'Hole argument', which is usually taken to show that substantivalism, the view that spacetime exists independently of its contents, introduces indeterminism into general relativity (Norton, 2015). Crucially, he does so by arguing that as long as we are careful in how we understand how mathematical structures like Lorentzian manifolds represent, we cannot make sense of the idea of two isometric manifolds such that one accurately represents some spacetime (constructed using a so-called 'hole transformation') and another doesn't. This turns on his premise that 'interpretations of our physical theories should be guided by the formalism of those theories ... [and] ... insofar as they are so guided, we need to be sure that we are using the formalism correctly, consistently, and according to our best understanding of the mathematics' (Weatherall, forthcoming, 3). This sounds plausible, but in using this premise to argue for his conclusion, he ends up arguing that there is *no* context in which isometric Lorentzian manifolds can be interpreted differently (Roberts, 2014). In the DEKI framework we can see how to resist Weatherall's premise, and therefore his conclusion. Even the very same object can be interpreted differently, so it should come as no surprise that the substantivalist is free to adopt different interpretations to isometric Lorentzian structures, such that one is an accurate representation of spacetime resulting from a hole transformation, and the other isn't.[135] Whether or not this is a useful position to take is a another question, but it's not ruled out by the nature of representation, or interpretation, themselves. Regardless, I think that this demonstrates that analysing how models represent will shed light on questions in the foundations of physics, as well as answering more general questions in the philosophy of science.

Theoretical equivalence. The previous point indicates a more general issue. There have been heated recent discussions regarding what makes two theories 'equivalent' to one another (Halvorson, 2012, 2013; Glymour, 2013; Barrett and Halvorson, 2015). But these have focused solely on the formal properties of theories and mathematical structures. In terms of the DEKI framework, the focus has been restricted to relationships between model objects. This misses important additional aspects of scientific representation: the interpretations; exemplified properties; keys used; and so on. Since it seems plausible that theoretical equivalence is related to, what might be called 'repre-

135 Compare this to the discussion of the colour inverted picture of Obama in Chapter 8.

sentational equivalence', investigating how DEKIs conditions are met in, for instance, Hamiltonian and Lagrangian formulations of mechanics, could shed light on theoretical equivalence.

All of this yields a hefty research programme to be developed from the ideas discussed in this thesis. I think that this tells in favour of the account. I hope to have provided an account of scientific representation that deepens our understanding of how scientific models represent. But the fact that this opens up avenues for future research just serves to highlight how central the notion of representation is to our philosophical projects, and our understanding of the world. As such, since DEKI provides a novel answer to the question of scientific representation, it provides a novel framework in which to think about philosophical questions that turn on the nature of scientific representation. Since these are many, the potential fruit of the account is significant.

BIBLIOGRAPHY

Abell, C. (2009). Canny resemblance. *Philosophical Review*, 118(2):183–223.

Achinstein, P. (1968). *Concepts of Science: A Philosophical Analysis*. Johns Hopkins Press, Baltimore.

Adams, E. W. (1959). The foundations of rigid body mechanics and the derivation of its laws from those of particle mechanics. In Henkin, L., Suppes, P., and Tarski, A., editors, *The Axiomatic Method: with Special Reference to Geometry and Physics*, pages 250–265. North-Holland, Amsterdam.

Ainsworth, P. (2009). Newman's objection. *British Journal for the Philosophy of Science*, 60(1):135–171.

Aldana, E. (2011). The moniac: Bill Phillips's machine. *Economia Politica*, XXVIII(1):167–170.

Ankeny, R. A. and Leonelli, S. (2011). What's so special about model organisms? *Studies in History and Philosophy of Science*, 42(2):313–323.

Armstrong, D. M. (1989). *Universals: An Opinionated Introduction*. Westview Press, London.

Bailer-Jones, D. M. (2003). When scientific models represent. *International Studies in the Philosophy of Science*, 17:59–74.

Bak, P. (1996). *How Nature Works: The Science of Self-Organized Criticality*. Springer-Verlag, New York.

Bak, P., Tang, C., and Wiesenfeld, K. (1987). Self-organized criticality: An explanation of the $1/f$ noise. *Phys. Rev. Lett.*, 59:381–384.

Balzer, W., Moulines, C. U., and Sneed, J. D. (1987). *An Architectonic for Science the Structuralist Program*. D. Reidel, Dordrecht.

Barr, N. (2000). The history of the Phillips machine. In Leeson, R., editor, *A. W. H. Phillips: Collected Works in Contemporary Perspective*, pages 89–114. Cambridge University Press, Cambridge.

Barrett, T. W. and Halvorson, H. (2015). Glymour and Quine on theoretical equivalence. *Journal of Philosophical Logic*, Online first. DOI:10.1007/s10992-015-9382-6.

Bartels, A. (2006). Defending the structural concept of representation. *Theoria*, 21(1):7–19.

Black, M. (1962). Models and archetypes. In Black, M., editor, *Models and Metaphors: Studies in Language and Philosophy*, pages 219–43. Cornell University Press, Ithaca and New York.

Bogen, J. and Woodward, J. (1988). Saving the phenomena. *Philosophical Review*, 97(3):303–52.

Bokulich, A. (2009). Explanatory fictions. In Suárez, M., editor, *Fictions in Science. Philosophical Essays on Modelling and Idealization*, pages 91–109. Routledge, London and New York.

Bolinska, A. (2013). Epistemic representation, informativeness and the aim of faithful representation. *Synthese*, 190(2):219–234.

Bolinska, A. (2015). *Epistemic Representation in Science and Beyond*. PhD: University of Toronto.

Boolos, G. S. and Jeffrey, R. C. (1989). *Computability and Logic*. Cambridge University Press, Cambridge, 3rd edition.

Brading, K. and Landry, E. (2006). Scientific structuralism: presentation and representation. *Philosophy of Science*, 73(5):571–581.

Bradley, R., Frigg, R., Steele, K., Thomson, E., and Werndl, C. (2016). The philosophy of climate science. In Carlos Galles, Pablo Lorenzano, E. O. and Rheinberger, H.-J., editors, *History and Philosophy of Science and Technology, Encyclopedia of Life Support Systems Volume 4*. Eolss, Isle of Man.

Brandom, R. B. (1994). *Making it Explicit: Reasoning, Representing and Discursive Commitment*. Harvard University Press, Cambridge MA.

Brandom, R. B. (2000). *Articulating Reasons: An Introduction to Inferentialism*. Harvard University Press, Cambridge MA.

Budd, M. (1993). How pictures look. In Knowles, D. and Skorupski, J., editors, *Virtue and Taste*, pages 154–175. Blackwell, Oxford.

Bueno, O. (2010). Models and scientific representations. In Magnus, P. and Busch, J., editors, *New Waves in Philosophy of Science*, pages 94–111. Pelgrave MacMillan, Hampshire.

Bueno, O. and Colyvan, M. (2011). An inferential conception of the application of mathematics. *Nous*, 45(2):345–374.

Bueno, O. and French, S. (2011). How theories represent. *British Journal for the Philosophy of Science*, 62(4):857–894.

Bueno, O., French, S., and Ladyman, J. (2002). On representing the relationship between the mathematical and the empirical. *Philosophy of Science*, 69(4):452–473.

Byerly, H. (1969). Model-structures and model-objects. *The British Journal for the Philosophy of Science*, 20(2):135–144.

Callender, C. and Cohen, J. (2006). There is no special problem about scientific representation. *Theoria*, 55:7–25.

Carnap, R. (1938). Foundations of logic and mathematics. In Neurath, O., Morris, C., and Carnap, R., editors, *International Encyclopaedia of Unified Science. Vol. 1.*, pages 139–213. University of Chicago Press, Chicago.

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

Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge University Press, Cambridge.

Cartwright, N., Shomar, T., and Suárez, M. (1995). The tool-box of science. In Herfel, W. E., Krajewski, W., Niiniluoto, I., and Wojcicki, R., editors, *Theories and Models in Scientific Processes (Poznan Studies in the Philosophy of Science and the Humanities 44)*, pages 137–150. Rodopi, Amsterdam and Atlanta.

Chakravartty, A. (2001). The semantic or model-theoretic view of theories and scientific realism. *Synthese*, 127(3):325–45.

Chakravartty, A. (2015). Scientific realism. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2015 edition.

Contessa, G. (2007). Scientific representation, interpretation, and surrogative reasoning. *Philosophy of Science*, 74(1):48–68.

Contessa, G. (2010). Scientific models and fictional objects. *Synthese*, 172(2):215–229.

Contessa, G. (2011). Scientific models and representation. In French, S. and Saatsi, J., editors, *The Continuum Companion to the Philosophy of Science*, page 120–137. Continuum Press, London.

Da Costa, N. C. A. and French, S. (1990). The model-theoretic approach to the philosophy of science. *Philosophy of Science*, 57:248–65.

Da Costa, N. C. A. and French, S. (2000). Models, theories, and structures: thirty years on. *Philosophy of Science*, 67(Supplement):S116–127.

Da Costa, N. C. A. and French, S. (2003). *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. Oxford University Press, Oxford.

Danto, A. (1981). *Transfiguration of the Commonplace: A Philosophy of Art*. Harvard University Press, Cambridge, MA and London.

Dardashti, R., Thébault, K. P. Y., and Winsberg, E. (2015). Confirmation via analogue simulation: what dumb holes could tell us about gravity. *The British Journal for the Philosophy of Science*, Online first. DOI:10.1093/bjps/axv010.

de Chadarevian, S. (2004). Models and the making of molecular biology. In de Chadarevian, S. and Hopwood, N., editors, *Models: The Third Dimension of Science*. Stanford University Press, Stanford.

de Donato Rodriguez, X. and Zamora Bonilla, J. (2009). Credibility, idealisation, and model building: an inferential approach. *Erkenntnis*, 70(1):101–118.

Decock, L. and Douven, I. (2011). Similarity after Goodman. *Review of Philosophy and Psychology*, 2(1):61–75.

Demopoulos, W. (2003). On the rational reconstruction of our theoretical knowledge. *British Journal for the Philosophy of Science*, 54(3):371–403.

Donnellan, K. S. (1968). Putting Humpty Dumpty together again. *Philosophical Review*, 77(2):203–215.

Downes, S. M. (2009). Models, pictures, and unified accounts of representation: lessons from aesthetics for philosophy of science. *Perspectives on Science*, 17(4):417–428.

Ducheyne, S. (2008). Towards an ontology of scientific models. *Metaphysica*, 9(1):119–127.

Ducheyne, S. (2012). Scientific representations as limiting cases. *Erkenntnis*, 76(1):73–89.

Dummett, M. (1991). *Frege: Philosophy of Mathematics*. Duckworth, London.

Eddon, M. (2013). Quantitative properties. *Philosophy Compass*, 8(7):633–645.

Elgin, C. Z. (1983). *With Reference to Reference*. Hackett, Indianapolis.

Elgin, C. Z. (1996). *Considered Judgement*. Princeton University Press, Princeton.

Elgin, C. Z. (2004). True enough. *Philosophical Issues*, 14(1):113–131.

Elgin, C. Z. (2009). Exemplification, idealization, and scientific understanding. In Suárez, M., editor, *Fictions in Science. Philosophical Essays on Modeling and Idealization*, pages 77–90. Routledge, New York and London.

Elgin, C. Z. (2010). Telling instances. In Frigg, R. and Hunter, M. C., editors, *Beyond Mimesis and Convention: Representation in Art and Science*, pages 1–18. Springer, Berlin and New York.

Enderton, H. B. (1972/2001). *A Mathematical Introduction to Logic*. Harcourt, San Diego and New York, 2nd edition.

Fodor, J. A. (1990). *A Theory of Content and Other Essays*. MIT Press.

Frege, G. (1892/1952). On sense and reference. In Geach, P. and Black, M., editors, *Translations from the Philosophical Writings of Gottlob Frege*. Blackwell, Oxford.

French, S. (2003). A model-theoretic account of representation (or, I don't know much about art ... but I know it involves isomorphism). *Philosophy of Science*, 70:1472–1483.

French, S. (2010). Keeping quiet on the ontology of models. *Synthese*, 172(2):231–249.

French, S. (2014). *The Structure of the World. Metaphysics and Representation*. Oxford University Press, Oxford.

French, S. and Ladyman, J. (1997). Superconductivity and structures: Revisiting the London account. *Studies in History and Philosophy of Modern Physics*, 28(3):363–393.

French, S. and Ladyman, J. (1999). Reinflating the semantic approach. *International Studies in the Philosophy of Science*, 13:103–121.

French, S. and Saatsi, J. (2006). Realism about structure: the semantic view and non-linguistic representations. *Philosophy of Science*, 73(5):548–559.

Friedman, M. (1953). The methodology of positive economics. In Friedman, M., editor, *Essays in Positive Economics*, pages 3–43. University of Chicago Press.

Frigg, R. (2002). Models and representation: why structures are not enough. *Measurement in Physics and Economics Project Discussion Paper Series, DP MEAS 25/02*.

Frigg, R. (2003). *Re-presenting Scientific Representation*. PhD: London School of Economics and Political Science.

Frigg, R. (2003). Self-organised criticality-what it is and what it isn't. *Studies in History and Philosophy of Science Part A*, 34(3):613–632.

Frigg, R. (2006). Scientific representation and the semantic view of theories. *Theoria*, 55(1):49–65.

Frigg, R. (2010a). Fiction and scientific representation. In Frigg, R. and Hunter, M., editors, *Beyond Mimesis and Convention: Representation in Art and Science*, pages 97–138. Springer, Berlin and New York.

Frigg, R. (2010b). Models and fiction. *Synthese*, 172(2):251–268.

Frigg, R. and Votsis, I. (2011). Everything you always wanted to know about structural realism but were afraid to ask. *European Journal for Philosophy of Science*, 1(2):227–276.

Gärdenfors, P. (2000). *Conceptual Spaces*. Bradford, Cambridge.

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

Giere, R. N. (1996). Visual models and scientific judgement. In Baigrie, B. S., editor, *Picturing Knowledge: Historical and Philosophical Problems Concerning the Use of Art in Science*, pages 269–302. University of Toronto Press, Toronto.

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science*, 71(4):742–752.

Giere, R. N. (2009). Why scientific models should not be regarded as works of fiction. In Suárez, M., editor, *Fictions in Science. Philosophical Essays on Modelling and Idealization*, pages 248–258. Routledge, London.

Giere, R. N. (2010). An agent-based conception of models and scientific representation. *Synthese*, 172(1):269 – 281.

Glymour, C. (2013). Theoretical equivalence and the semantic view of theories. *Philosophy of Science*, 80(2):286–297.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy*, 21(5):725–740.

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

Gombrich, E. (1961). *Art and Illusion*. Princeton University Press, Princeton.

Goodman, N. (1972). Seven strictures on similarity. In Goodman, N., editor, *Problems and Projects*, pages 437–446. Bobbs-Merrill, Indianapolis and New York.

Goodman, N. (1976). *Languages of Art*. Hacket, 2nd ed., Indianapolis and Cambridge.

Goodman, N. (1984). *Of Mind and Other Matters*. Harvard University Press, Cambridge MA.

Goodwin, R. M. (1982). A growth cycle. In Goodwin, R. M., editor, *Essays in Economic Dynamics*, pages 165–170. Palgrave Macmillan, London.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press, Cambridge.

Hale, S. (1988). Spacetime and the abstract/concrete distinction. *Philosophical Studies*, 53(1):85–102.

Halvorson, H. (2012). What scientific theories could not be. *Philosophy of Science*, 79(2):183–206.

Halvorson, H. (2013). The semantic view, if plausible, is syntactic. *Philosophy of Science*, 80(3):475–478.

Halvorson, H. (2016). Scientific theories. In Humphreys, P., editor, *The Oxford Handbook of Philosophy of Science*. Oxford University Press, Oxford.

Harris, T. (2003). Data models and the acquisition and manipulation of data. *Philosophy of Science*, 70(5):1508–1517.

Hartmann, S. (1995). Models as a tool for theory construction: some strategies of preliminary physics. In Herfel, W. E., Krajewski, W., Niiniluoto, I., and Wojcicki, R., editors, *Theories and Models in Scientific Processes (Poznan Studies in the Philosophy of Science and the Humanities 44)*, pages 49–67. Rodopi, Amsterdam and Atlanta.

Hellman, G. (1989). *Mathematics without Numbers: Towards a Modal-Structural Interpretation*. Oxford University Press, Oxford.

Hellman, G. (1996). Structuralism without structures. *Philosophia Mathematica*, 4(3):100–123.

Hempel, C. G. (1965). Aspects of scientific explanation. In Hempel, C. G., editor, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, pages 331–496. Free Press, New York.

Hendry, R. F. (1998). Models and approximations in quantum chemistry. In Shanks, N., editor, *Idealization IX: Idealization in Contemporary Physics*, volume 63 of *Poznan Studies in the Philosophy of the Sciences and the Humanities*, pages 123–142. Rodopi, Amsterdam.

Hesse, M. (1963). *Models and Analogies in Science*. Sheed and Ward, London.

Hughes, R. I. G. (1997). Models and representation. *Philosophy of Science*, 64(Supplement):S325–S336.

Hughes, R. I. G. (1998). Laws of nature, laws of physics, and the representational account of theories. *ProtoSociology*, 12:113–143.

Hughes, R. I. G. (2010). *The Theoretical Practises of Physics: Philosophical Essays*. Oxford Univeristy Press, Oxford.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181:662 – 666.

Kennedy, A. G. (2012). A non representationalist view of model explanation. *Studies in History and Philosophy of Science*, 43(2):326–332.

Kenny, A. (1973/2006). *Wittgenstein*. Blackwell, Cambridge MA, 2nd edition.

Ketland, J. (2004). Empirical adequacy and ramsification. *The British Journal for the Philosophy of Science*, 55(2):287–300.

Klein, C. (2013). Multiple realizability and the semantic view of theories. *Philosophical Studies*, 163(3):683–695.

Knuuttila, T. (2005). Models, representation, and mediation. *Philosophy of Science*, 72(5):1260–1271.

Kripke, S. (1980). *Naming and Necessity*. Harvard University Press, Cambridge MA.

Kroes, P. (1989). Structural analogies between physical systems. *British Journal for the Philosophy of Science*, 40(2):145–154.

Kulvicki, J. (2006). Pictorial representation. *Philosophy Compass*, 1(6):535–546.

Künne, W. (2003). *Conceptions of Truth*. Clarendon Press, Oxford.

Landry, E. (2007). Shared structure need not be shared set-structure. *Synthese*, 158(1):1–17.

Laurence, S. and Margolis, E. (1999). Concepts and cognitive science. In Laurence, S. and Margolis, E., editors, *Concepts: Core Readings*, pages 3–81. MIT Press, Cambridge MA.

Laymon, R. (1990). Computer simulations, idealizations and approximations. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2:519–534.

Levy, A. (2012). Models, fictions, and realism: two packages. *Philosophy of Science*, 79(5):738–748.

Levy, A. (2015). Modeling without models. *Philosophical Studies*, 152(3):781–798.

Liu, C. (1999). Explaining the emergence of cooperative phenomena. *Philosophy of Science*, 66(Supplement):S92–S106.

Liu, C. (2013). Deflationism on scientific representation. In Karakostas, V. and Dieks, D., editors, *EPSA11 Perspectives and Foundational Problems in Philosophy of Science*, pages 93–102. Springer.

Lloyd, E. (1994). *The Structure and Confirmation of Evolutionary Theory*. Princeton University Press, Princeton.

Lopes, D. (2004). *Understanding Pictures*. Oxford University Press, Oxford.

Lotka, A. (1925). *Elements of Physical Biology*. Williams & Wilkins Company.

Lutz, S. (2015). What was the syntax-semantics debate in the philosophy of science about? *Philosophy and Phenomenological Research*, Online first. DOI:10.1111/phpr.12221.

Lycan, W. G. (2000). *Philosophy of Language: a Contemporary Introduction*. Routledge Contemporary Introductions to Philosophy. Routledge, London, 2nd edition.

MacKay, A. F. (1968). Mr. Donnellan and Humpty Dumpty on referring. *Philosophical Review*, 77(2):197–202.

Mäki, U. (2009). Missing the world. models as isolations and credible surrogate systems. *Erkenntnis*, 70(1):29–43.

Mäki, U. (2011). Models and the locus of their truth. *Synthese*, 180(1):47–63.

Mankiw, N. G. (2012). *Macroeconomics*. Worth Publishers, New York, 7th edition.

Marcus, R. B. (1961). Modalities and intensional languages. *Synthese*, 13(4):303–322.

McAllister, J. W. (1997). Phenomena and patterns in data sets. *Erkenntnis*, 47(2):217–228.

Michaelson, E. (2013). *This and That: a Theory of Reference for Names, Demonstratives, and Things in Between*. PhD: University of California.

Mill, J. S. (1843). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. Number v. 1 in A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation. John W. Parker.

Morgan, M. (2012). *The World in the Model. How Economists Work and Think*. Cambridge University Press, Cambridge.

Morgan, M. and Boumans, M. (2004). Secrets hidden by two-dimensionality: the economy as a hydraulic machine. In de Chadarevian, S. and Hopwood, N., editors, *Models: The Third Dimension of Science*, page 369–401. Stanford University Press, Stanford.

Morgan, M. and Morrison, M. (1999). *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge University Press, Cambridge.

Morreau, M. (2010). It simply does not add up: the trouble with overall similarity. *Journal of Philosophy*, 107(9):469–490.

Muller, F. A. (2011). Reflections on the revolution at Stanford. *Synthese*, 183(1):87–114.

Muller, F. A. and van Fraassen, B. C. (2008). How to talk about unobservables. *Analysis*, 68(3):197–205.

Mundy, B. (1986). On the general theory of meaningful representation. *Synthese*, 67(3):391–437.

Nguyen, J. (2016). On the pragmatic equivalence between representing phenomena and data. *Philosophy of Science*, Online first. DOI:10.1086/684959.

Niiniluoto, I. (1988). Analogy and similarity in scientific reasoning. In Helman, D., editor, *In Analogical Reasoning: Perspectives of Artificial Intelligence, Cognitive Science, and Philosophy*, pages 271–298. Kluwer, Dordrecht.

Norton, J. (2012). Approximation and idealization: why the difference matters. *Philosophy of Science*, 79(2):207–232.

Norton, J. D. (2015). The hole argument. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2015 edition.

Parker, W. (2015). Getting (even more) serious about similarity. *Biology and Philosophy*, 30(2):267–276.

Peirce, C. (1932). *Collected Papers of Charles Sanders Peirce, Volumes I and II: Principles of Philosophy and Elements of Logic. Eds. Hartshorne, C and Weiss, P.* Harvard University Press, Cambridge MA.

Perini, L. (2005a). The truth in pictures. *Philosophy of Science*, 72(1):262–285.

Perini, L. (2005b). Visual representation and confirmation. *Philosophy of Science*, 72(5):913–926.

Perini, L. (2010). Scientific representation and the semiotics of pictures. In Magnus, P. and Busch, J., editors, *New Waves in the Philosophy of Science*, pages 131–154. Macmilan, New York.

Pero, F. and Suárez, M. (2015). Varieties of misrepresentation and homomorphism. *European Journal for Philosophy of Science*, 6(1):71–90.

Peschard, I. (2011). Making sense of modeling: beyond representation. *European Journal for Philosophy of Science*, 1(3):335–352.

Phillips, A. W. (1950). Mechanical models in economic dynamics. *Economica*, 17(67):283–305.

Pincock, C. (2012). *Mathematics and Scientific Representation*. Oxford University Press, Oxford.

Portides, D. (2005). Scientific models and the semantic view of theories. *Philosophy of Science*, 72(5):1287–1289.

Portides, D. (2010). Why the model-theoretic view of theories does not adequately depict the methodology of theory application. In Suárez, M., Dorato, M., and Rédei, M., editors, *EPSA Epistemology and Methodology of Science*, book section 18, pages 211–220. Springer, Dordrecht.

Poznic, M. (2015). Representation and similarity: Suárez on necessary and sufficient conditions of scientific representation. *Journal for General Philosophy of Science*, Online first. DOI:10.1007/s10838-015-9307-7.

Price, H. (2011). *Naturalism Without Mirrors*. Oxford University Press, Oxford.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. Philosophical issues in science. Routledge.

Putnam, H. (1981). *Reason, Truth, and History*. Cambridge University Press, Cambridge.

Putnam, H. (2002). *The Collapse of the Fact-Value Distinction*. Harvard University Press, Cambridge MA.

Quine, W. v. O. (1948). On what there is. *Review of Metaphysics*, 2(5):21–36.

Quine, W. V. O. (1969). *Ontological Relativity and Other Essays*. Columbia University Press, New York.

Ramsey, J. L. (2006). Approximation. In Sarkar, S. and Pfeifer, J., editors, *The Philosophy of Science: An Encyclopedia*, pages 24–27. Routledge, New York.

Redhead, M. (2001). The intelligibility of the universe. In O'Hear, A., editor, *Philosophy at the New Millennium*. Cambridge University Press, Cambridge.

Reimer, M. and Michaelson, E. (2014). Reference. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Winter 2014 edition.

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

Resnik, M. D. (1997). *Mathematics as a Science of Patterns*. Oxford University Press, Oxford.

Robbins, L. (1938). Interpersonal comparisons of utility: A comment. *The Economic Journal*, 48(192):635–641.

Roberts, B. W. (2014). Disregarding the 'hole argument'. Manuscript. http://philsci-archive.pitt.edu/11687/.

Rosen, G. (2014). Abstract objects. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2014 edition.

Russell, B. (1905). On denoting. *Mind*, 14(56):479–493.

Sainsbury, R. M. (2010). *Fiction and Fictionalism*. Routledge, Oxford.

Sainsbury, R. M. (2011). Fiction and acceptance-relative truth, belief and assertion. In Lihoreau, F., editor, *Truth in Fiction*, pages 137–152. Ontos Verlag.

Schier, F. (1986). *Deeper in Pictures: An Essay on Pictorial Representation*. Cambridge University Press, Cambridge.

Searle, J. R. (1958). Proper names. *Mind*, 67(266):166–173.

Shapiro, S. (2000). *Thinking About Mathematics*. Oxford University Press, Oxford.

Shech, E. (2014). Scientific misrepresentation and guides to ontology: the need for representational code and contents. *Synthese*, Online first. DOI:10.1007/s11229-014-0506-2.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210(4468):390–398.

Stalnaker, R. (1970). Pragmatics. *Synthese*, 22(1-2):272–289.

Stalnaker, R. (1984). *Inquiry*. Bradford Books MIT Press, Cambridge MA.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, 25(5-6):701–721.

Sterelny, K. and Griffiths, P. E. (1999). *Sex and Death: An Introduction to Philosophy of Biology*. University of Chicago Press, Chicago and London.

Sterrett, S. G. (2006). Models of machines and models of phenomena. *International Studies in the Philosophy of Science*, 20(1):69–80.

Stevenson, M. (2011). The search for the fountain of prosperity. *Economia Politica*, XXVIII(1):151–166.

Stich, S. and Warfield, T. (1994). *Mental Representation. A Reader*. Blackwell, Oxford.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M. (2013). Climate change 2013: The physical science basis. Technical report.

Suppe, F. (1989). *The Semantic Conception of Theories and Scientific Realism*. University of Illinois Press, Urbana and Chicago.

Suppes, P. (1960/1969a). A comparison of the meaning and uses of models in mathematics and the empirical sciences. In Suppes, P., editor, *Studies in the Methodology and Foundations of Science: Selected Papers from 1951 to 1969*, pages 10–23. Reidel, Dordrecht.

Suppes, P. (1962/1969b). Models of data. In Suppes, P., editor, *Studies in the Methodology and Foundations of Science: Selected Papers from 1951 to 1969*, pages 24–35. Reidel, Dordrecht.

Suppes, P. (2002). *Representation and Invariance of Scientific Structures*. CSLI Publications, Stanford.

Suárez, M. (2003). Scientific representation: against similarity and isomorphism. *International Studies in the Philosophy of Science*, 17(3):225–244.

Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science*, 71(Supplement):767–779.

Suárez, M. (2010). Scientific representation. *Philosophy Compass*, 5(1):91–101.

Suárez, M. (2015). Deflationary representation, inference, and practice. *Studies in History and Philosophy of Science*, 49:36–47.

Suárez, M. and Solé, A. (2006). On the analogy between cognitive representation and truth. *Theoria*, 55(1):39–48.

Swoyer, C. (1991). Structural representation and surrogative reasoning. *Synthese*, 87(3):449–508.

Tal, E. (2015). Measurement in science. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Summer 2015 edition.

Tegmark, M. (2008). The mathematical universe. *Foundations of Physics*, 38(2):101–150.

Teller, P. (2001). Twilight of the perfect model model. *Erkenntnis*, 55(3):393–415.

Thomasson, Amie, L. (forthcoming). If models were fictions, then what would they be? In Godfrey-Smith, P. and Levy, A., editors, *The Scientific Imagination*. Oxford University Press, Oxford.

Thomson-Jones, M. (2006). Models and the semantic view. *Philosophy of Science*, 73(5):524–535.

Thomson-Jones, M. (2010). Missing systems and face value practise. *Synthese*, 172(2):283 – 299.

Thomson-Jones, M. (2011a). Review of Scientific Representation: Paradoxes of Perspective. *Australasian Journal of Philosophy*, 89(3):567–570.

Thomson-Jones, M. (2011b). Structuralism about scientific representation. In Bokulich, A. and Bokulich, P., editors, *Scientific Structuralism*, volume 281 of *Boston Studies in the Philosophy of Science*, pages 119–141. Springer, Dordrecht.

Toon, A. (2010a). Models as make-believe. In Frigg, R. and Hunter, M., editors, *Beyond Mimesis and Convention: Representation in Art and Science*, pages 71–96. Springer, Berlin.

Toon, A. (2010b). The ontology of theoretical modelling: models as make-believe. *Synthese*, 172(2):301–315.

Toon, A. (2011). Playing with molecules. *Studies in History and Philosophy of Science*, 42:580–589.

Toon, A. (2012). *Models as Make-Believe. Imagination, Fiction and Scientific Representation*. Palgrave Macmillan, Basingstoke.

Tsui, K.-y. and Weymark, J. A. (1997). Social welfare orderings for ratio-scale measurable utilities. *Economic Theory*, 10(2):241–256.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4):327–352.

Ubbink, J. (1960). Model, description and knowledge. *Synthese*, 12(2-3):302–319.

van Fraassen, B. C. (1980). *The Scientific Image*. Oxford University Press, Oxford.

van Fraassen, B. C. (1981). Theory construction and experiment: an empiricist view. *Philosophy of Science (Proceedings)*, Vol. 2:663–677.

van Fraassen, B. C. (1985). Empricism in the philosophy of science. In Churchland, P. M. and Hooker, C. A., editors, *Images of Science: Essays on Realism and Empiricism with a Reply from Bas C. van Fraassen*, page 245–308. University of Chicago Press, Chicago and London.

van Fraassen, B. C. (1989). *Laws and Symmetry*. Clarendon Press, Oxford.

van Fraassen, B. C. (1991). *Quantum Mechanics: An Empiricist View*. Oxford University Press, Oxford.

van Fraassen, B. C. (1995). A philosophical approach to foundations of science. *Foundations of Science*, 1(1):5–9.

van Fraassen, B. C. (1997). Structure and perspective: philosophical perplexity and paradox. In Dalla Chiara, M. L., editor, *Logic and Scientific Methods*, pages 511–530. Kluwer, Dordrecht.

van Fraassen, B. C. (2002). *The Empirical Stance*. Yale University Press, New Haven and London.

van Fraassen, B. C. (2008). *Scientific Representation: Paradoxes of Perspective*. Oxford University Press.

van Fraassen, B. C. (2010). Reply to Contessa, Ghins, and Healey. *Analysis Reviews*, 70(3):547–556.

van Fraassen, B. C. (2014). One or two gentle remarks about Hans Halvorson's critique of the semantic view. *Philosophy of Science*, 81(2):276–283.

Volterra, V. (1926). Fluctuations in the abundance of a species considered mathematically. *Nature*, 118:558 – 560.

Volterra, V. (1928). Variations and fluctuations of the number of individuals in animal species living together. *Journal du Conseil*, 3(1):3–51.

Walton, K. L. (1990). *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard University Press, Cambridge MA.

Weatherall, J. O. (forthcoming). Regarding the "hole argument". *The British Journal for the Philosophy of Science*.

Weisberg, M. (2007). Who is a modeler? *The British Journal for the Philosophy of Science*, 58:207–233.

Weisberg, M. (2012). Getting serious about similarity. *Philosophy of Science*, 79(5):785 – 794.

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

Weisberg, M. (2015). Biology and philosophy symposium on Simulation and Similarity: Using Models to Understand the World: response to critics. *Biology and Philosophy*, 30(2):299–310.

Weisberg, M. and Reisman, K. (2008). The robust Volterra principle. *Philosophy of Science*, 75(1):106–131.

Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. Richard Courant lecture in mathematical sciences delivered at New York University, May 11, 1959. *Communications on Pure and Applied Mathematics*, 13(1):1–14.

Winther, R. G. (2015). The structure of scientific theories. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2015 edition.

Wittgenstein, L. (1953). *Philosophical Investigations. (Translated by Anscombe, G.E.M.)*. Oxford.

Wollheim, R. (1987). *Painting as an Art*. Princeton University Press, Princeton.

Wollheim, R. (1998). On pictorial representation. *The Journal of Aesthetics and Art Criticism*, 56(3):217–226.

Woodward, J. (1989). Data and phenomena. *Synthese*, 79(3):393–472.

Woody, A. I. (2004). More telltale signs: what attention to representation reveals about scientific explanation. *Philosophy of Science*, 71(5):780 – 793.

Yablo, S. (2014). *Aboutness*. Princeton University Press, Princeton.