Part A

Theoretic

Part A Theoretical Issues in Models

Ed. by Demetris Portides

It is not hard to notice the lack of attention paid to scientific models in mid-twentieth century philosophy of science. Models were, for instance, absent from philosophical theories of scientific explanation; they were also absent from attempts to understand how theoretical concepts relate to experimental results. In the last few decades, however, this has changed, and philosophers of science are increasingly turning their attention to scientific models. Models and their relation to other parts of the scientific apparatus are now under philosophical scrutiny; at the same time, they are instrumental parts of approaches that aim to address certain philosophical questions.

After recognizing the significance of models in scientific inquiry and in particular the significance of models in linking theoretical concepts to experimental reports, philosophers have begun to explore a number of questions about the nature and function of models. There are several philosophically interesting questions that could fit very well into the theme of this set of chapters. For example, *what is the function of models?* and *what is the role of idealization and abstraction in modeling?*. It is, however, not the objective of this set of chapters to address every detail about models that has gained philosophical interest over time. In this part of the book five model-related philosophical questions are isolated from others and are explored in separate chapters:

1. What is a scientific model?
2. How do models and theories relate?
3. How do models represent phenomena?
4. How do models function in scientific explanation?
5. How do models and other modes of scientific theorizing, such as simulations, relate?

Of course, the authors of these chapters are all aware that isolating these questions is only done in order to reach an intelligible exposition of the explored problems concerning models, and not because different questions have been kept systematically apart in the philosophical literature that preceded this work. In fact, the very nature of some of these questions dictates an interrelation with others and attempts to address one leads to overlaps with attempts to address others. For example, how one addresses the question *what sort of entities are models?* or how one conceives the theory–model relation affects the understanding of their scientific representation and scientific explanation, and vice versa. Although this point becomes evident in the subsequent chapters, a conscious attempt was made by each author to focus on the one question of concern of their chapter and to attempt to extrapolate and explicate the different proposed philosophical accounts that have been offered in the quest to answer that particular question. We hope that the final outcome is helpful and illuminating to the reader.

Axel Gelfert in his contribution, Chap. 1: *The Ontology of Models*, explicates the different ways in which philosophers have addressed the issue of what a scientific model is. For historical reasons, he begins by examining the view that was foremost almost a century ago, which held that models could be understood as analogies. He then quickly turns his attention to a debate that took place in the second half of the twentieth century between advocates of logical positivism, who held that models are interpretations of a formal calculus, and advocates of the semantic view, which maintained that models are directly defined mathematical structures. He continues by examining the more recent view, which identifies models with fictional entities. He closes his chapter with an explication of what he calls the more pragmatic accounts, which hold that models can best be understood with the use of a mixed ontology.

In Chap. 2: *Models and Theories*, *Demetris Portides* explicates the two main conceptions of the structure of scientific theories (and subsequently the two main conceptions of the theory–model relation) in the history of the philosophy of science, the received and the semantic views. He takes the reader through the main arguments that led to the collapse of the received view and gives the reader a lens by which to distinguish the different versions of the semantic view. He finally presents the main arguments against the semantic view and in doing so he explicates a more recent philosophical trend that conceives the theory–model relation as too complex to sufficiently capture with formal tools.

Roman Frigg and *James Nguyen*, in Chap. 3: *Models and Representation* begin by analyzing the concept of representation and clarifying its main characteristics and the conditions of adequacy any theory of representation should meet. They then proceed to explain the main theories of representation that have been proposed in the literature and explain with reference to their proposed set of characteristics and conditions of adequacy where each theory is found wanting. The similarity, the structuralist, the inferential, the fictionalist, and the denotational accounts of representation are all thoroughly explained and critically assessed. By doing this the authors expose and explicate many of the weaknesses of the different accounts of representation.

In Chap. 4: *Models and Explanation*, *Alisa Bokulich* explains that by recognizing the extensive use of models in science and by realizing that models are more often than not highly idealized and incomplete descriptions of phenomena that frequently incorporate fictional elements, philosophers have been led to revise previous philosophical accounts of scientific explanation. By scrutinizing different model-based accounts of scientific explanation offered in the literature and exposing the problems involved, she highlights the difficulties involved in resolving the issue of whether or not the falsehoods present in models are operative in scientific explanation.

Finally, *Nancy Nersessian* and *Miles McLeod*, in Chap. 5: *Models and Simulations*, explicate a more recent issue that is increasingly gaining the interest of philosophers: how scientific models, i.e., the mathematical entities that scientists traditionally use to represent phenomena, relate to simulations, particularly computational simulations. They give a flavor of the characteristics of computational simulations both in the context of well-developed overarching theories and in the context where an overarching theory is absent. The authors also highlight the epistemological significance of simulations for all such contexts by elaborating on how simulations introduce novel problems that should concern philosophers. Finally, they elaborate on the relation between simulations and other constructs of human cognition such as thought experiments.

In most cases, in all chapters the technical aspects of the philosophical arguments have been kept to a minimum in order to make them accessible even to readers working outside the sphere of the philosophy of science. Suppressing the technical aspects has not, however, introduced misrepresentation or distortion to philosophical arguments.

1. The Ontology of Models

Axel Gelfert

The term *scientific model* picks out a great many things, including scale models, physical models, sets of mathematical equations, theoretical models, toy models, and so forth. This raises the question of whether a general answer to the question *What is a model*? is even possible. This chapter surveys a number of philosophical approaches that bear on the question of what, in general, a scientific model is. While some approaches aim for a unitary account that would apply to models in general, regardless of their specific features, others take as their basic starting point the manifest heterogeneity of models in scientific practice. This chapter first motivates the ontological question of what models are by reflecting on the diversity of different kinds of models and arguing that models are best understood as *functional entities*. It then provides some historical background regarding the use of analogy in science as a precursor to contemporary notions of *scientific model*. This is followed by a contrast between the syntactic and the semantic views of theories and models and their different stances toward the question of what a model is. Scientists, too, typically operate with tacit assumptions about the ontological status of models: this gives rise to what has been called the *folk ontology* of models, according to which models may be thought of as descriptions of missing (i. e., uninstantiated) systems. There is a close affinity between this view and recent philosophical positions (to be discussed in the penultimate section) according to which models are fictions. This chapter concludes by considering various pragmatic conceptions of models, which are typically associated with what may be called *mixed ontologies*, that is, with the view that any quest for a unitary account of the nature of models is bound to be fruitless.

The philosophical discussion about models has emerged from a cluster of concerns, which span a range of theoretical, formal, and practical questions across disciplines ranging from logic and mathematics to aesthetics and artistic representations. In what follows, the term *models* will normally be taken as synonymous to *scientific models*, and any departure from this usage – for example, when discussing the use of models in nonscientific settings – will either be indicated explicitly or will be clear from context. Focusing on scientific models helps to clarify matters, but still leaves a wide range of competing philosophical approaches for discussion. This chapter will summarize and critically discuss a number of such approaches, especially those that shed light on the question *what is a model*?; these will range from views that, by now, are of largely historical interest to recent proposals at the cutting edge of the philosophy of science. While the emphasis throughout will be on the ontology of models, it will often be necessary to also reflect on their function, use, and construction. This is not meant to duplicate the discussion provided in other chapters of this handbook; rather, it is the natural result of scientific models having traditionally been defined either in terms of their function (e.g., to provide representations of target systems) or via their relation to other (purportedly) better understood entities, such as scientific theories.

The rest of this chapter is organized as follows: Sect. 1.1 will set the scene by introducing a number of examples of scientific models, thereby raising the question of what degree of unity any philosophical account of scientific models can reasonably aspire to. Section 1.2 will characterize models as functional entities and will provide a general taxonomy for how to classify various possible philosophical approaches. A first important class of specific accounts, going back to nineteenth-century scientists and philosophers, will be discussed in Sect. 1.3, which focuses on models as analogies. Section 1.4 is devoted to formal approaches that dominated much of twentieth-century discussion of scientific models. In particular, it will survey the syntactic view of theories and models and its main competitor, the semantic view, along with recent formal approaches (such as the partial structures approach) which aim to address the shortcomings of their predecessors. Section 1.5 provides a sketch of what has been called the *folk ontology* of models – that is, a commonly shared set of assumptions that inform the views of scientific practitioners. On this view, models are place-holders for *imaginary concrete systems* and as such are not unlike fictions. The implications of fictionalism about models are discussed in Sect. 1.6. Finally, in Sect. 1.7, recent pragmatic accounts are discussed, which give rise to what may be called a *mixed ontology*, according to which models are best conceived of as a heterogeneous mixture of elements.

1.1 Kinds of Models: Examples from Scientific Practice

Models can be found across a wide range of scientific contexts and disciplines. Examples include the Bohr model of the atom (still used today in the context of science education), the billiard ball model of gases, the DNA double helix model, scale models in engineering, the Lotka–Volterra model of predator–prey dynamics in population biology, agent-based models in economics, the Mississippi River Basin model (which is a 200 acres hydraulic model of the waterways in the entire Mississippi River Basin), and general circulation models (GCMs), which allow scientists to run simulations of Earth's climate system. The list could be continued indefinitely, with the number of models across the natural and social sciences growing day by day.

In philosophical discussions of scientific models, the situation is hardly any different. The *Stanford Encyclopedia of Philosophy* gives the following list of model types that have been discussed by philosophers of science [1.1]:

> "Probing models, phenomenological models, computational models, developmental models, explanatory models, impoverished models, testing models, idealized models, theoretical models, scale models, heuristic models, caricature models, didactic models, fantasy models, toy models, imaginary models, mathematical models, substitute models, iconic models, formal models, analogue models and instrumental models."

The proliferation of models and model types, in the sciences as well as in the philosophical literature, led *Goodman* to lament in his 1968 *Languages of Art* [1.2, p. 171]: "Few terms are used in popular and scientific discourse more promiscuously than *model*." If this was true of science and popular discourse in the late 1960s, it is all the more true of the twenty-first century philosophy of science.

As an example of a physics-based model, consider the *Ising model*, proposed in 1925 by the German physicist Ernst Ising as a model of ferromagnetism in certain metals. The model starts from the idea that a macroscopic magnet can be thought of as a collection of elementary magnets, whose orientation determines the overall magnetization. If all the elementary magnets are aligned along the same axis, then the system will be perfectly ordered and will display a maximum value of the magnetization. In the simplest one-dimensional (1-D) case, such a state can be visualized as a chain of *elementary magnets*, all pointing the same way

$$\cdots\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\cdots$$

The alignment of elementary magnets can be brought about either by a sufficiently strong external magnetic field or it can occur spontaneously, as will happen below a critical temperature, when certain substances (such as iron and nickel) undergo a ferromagnetic phase transition. Whether or not a system will undergo a phase transition, according to thermodynamics, depends on its energy function, which in turn is determined by the interactions between the component parts of the system. For example, if neighboring *elementary magnets* interact in such a way as to favor alignment, there is a good chance that a spontaneous phase transition may occur below a certain temperature. The energy function, then, is crucial to the model and, in the case of the Ising model, is defined as

$$E = -\sum_{i,j} J_{ij} S_i S_j \ ,$$

with the variable S_i representing the orientation ($+1$ or -1) of an elementary magnet at site i in the crystal lattice and J_{ij} representing the strength of interaction between two such elementary magnets at different lattice sites i and j.

Contrast this with *model organisms* in biology, the most famous example of which is the fruit fly *Drosophila melanogaster*. Model organisms are real organisms – actual plants and animals that are alive and can reproduce – yet they are used as representations either of another organism (e.g., when rats are used in place of humans in medical research) or of a biological phenomenon that is more universal (e.g., when fruit flies are used to study the effects of crossover between homologous chromosomes). Model organisms are often bred for specific purposes and are subject to artificial selection pressures, so as to purify and *standardize* certain features (e.g., genetic defects or variants) that would not normally occur, or would occur only occasionally, in populations in the wild. As *Ankeny* and *Leonelli* put it, in their ideal form "model organisms are thought to be a relatively simplified form of the class of organism of interest" [1.3, p. 318]; yet it often takes considerable effort to work out the actual relationships between the model organism and its target system (whether it be a certain biological phenomenon or a specific class of target organisms). Tractability and various experimental desiderata – for example, a short life cycle (to allow for quick breeding) and a relatively small and compact genome (to allow for the quick identification of variants) – take precedence over theoretical questions in the choice of model organisms; unlike for the Ising model, there is no simple mathematical formula that one can rely on to study how one's model behaves, only the messy world of real, living systems.

The Ising model of ferromagnetism and model organisms such as *Drosophila melanogaster* may be at opposite ends of the spectrum of scientific models. Yet the diversity among those models that occupy the middle ground between theoretical description and experimental system is no less bewildering. How, one might wonder, can a philosophical account of scientific models aspire to any degree of unity or generality in the light of such variety? One obvious strategy is to begin by drawing distinctions between different overarching types of models. Thus, *Black* [1.4] distinguishes between four such types:

1. Scale models
2. Analog models
3. Mathematical models
4. Theoretical models.

The basic idea of scale and analog models is straightforward: a scale model increases or decreases certain (e.g., spatial) features of the target system, so as to render them more manageable in the model; an analog model also involves the change of medium (as in once popular hydraulic models of the economy, where the flow of money was represented by the flow of liquids through a system of pumps and valves). Mathematical models are constructed by first identifying a number of relevant variables and then developing empirical hypotheses concerning the relations that may hold between the variables; through (often drastic) simplification, a set of mathematical equations is derived, which may then be evaluated analytically or numerically and tested against novel observations. Theoretical models, finally, begin usually by extrapolating imaginatively from a set of observed facts and regularities, positing new entities and mechanisms, which may be integrated into a possible theoretical account of a phenomenon; comparison with empirical data usually comes only at a later stage, once the model has been formulated in a coherent way.

Achinstein [1.5] includes mathematical models in his definition of *theoretical model*, and proposes an analysis in terms of sets of assumptions about a model's target system. This allows him to include Bohr's model of the atom, the DNA double-helix model (considered as a set of structural hypotheses rather than as a physical ball-and-stick model), the Ising model, and the Lotka–Volterra model among the class of theoretical systems. Typically, when a scientist constructs a theoretical model, she will help herself to certain established principles of a more fundamental theory to which she is committed. These will then be adapted or modified, notably by introducing various new assumptions specific to the case at hand. Typically, an inner structure or mechanism is posited which is thought to explain the features of the target system. At the same time, there is the (often explicit) acknowledgment that the target system is far more complex than the model is able to capture: in this sense, a theoretical model is believed by the practitioner to be false as a description of the target system. However, this acknowledgment of the limits of applicability of models also allows researchers to simultaneously use different models of the same target system alongside each other. Thus understood, theoretical models usually involve the combination of general theoretical principles and specific auxiliary assumptions, which may only be valid for a narrow range of parameters.

1.2 The Nature and Function of Models

The great variety of models employed in scientific practice, as illustrated by the long list given in the preceding section, suggests two things. First, it makes vivid just how central the use of models is to the scientific enterprise and to the self-image of scientists. As *von Neumann* put it, with some hyperbole [1.6, p. 492]: "The sciences do not try to explain, they hardly even try to interpret, they mainly make models." Whatever shape and form the scientific enterprise might take without the use of models, it seems safe to say that it would not look anything like science as we presently know it. Second, one might wonder whether it is at all reasonable to look for a unitary philosophical account of models. Given the range of things we call *models*, and the diversity of uses to which they are being put, it may simply not be possible to give a one-size-fits-all answer to the question *what is a model?* This has led some commentators to propose quietism as the only viable attitude toward ontological questions concerning models and theories. As *French* puts it [1.7, p. 245],

> "whereas positing the reality of quarks or genes may contribute to the explanation of certain features of the physical world, adopting a similar approach toward theories and models – that is, reifying them as entities for which a single unificatory account can be given – does nothing to explain the features of scientific practice."

While there are good grounds for thinking that quietism should only be a position of last resort in philosophy, the sentiment expressed by French may go some way toward explaining why there has been a relative dearth of philosophical work concerning the ontology of models. The neglect of ontological questions concerning models has been remarked upon by a number of contributors, many of whom, like *Contessa*, find it [1.8, p. 194]

> "surprising if one considers the amount of interest raised by analogous questions about the ontology and epistemology of mathematical objects in the philosophy of mathematics."

A partial explanation of this discrepancy lies in the arguably greater heterogeneity in what the term *scientific models* is commonly thought to refer to, namely, anything from physical ball-and-stick models of chemical molecules to mathematical models formulated in terms of differential equations. (If we routinely included dividers, compasses, set squares, and other technical drawing tools among, say, the class of *geometrical entities*, the ontology of mathematical entities, too, would quickly become rather unwieldy!)

In the absence of any widely accepted unified account of models – let alone one that would provide a conclusive answer to ontological questions arising from models – it may be natural to assume, as indeed many contributors to the debate have done, that "if all scientific models have something in common, this is not their *nature* but their *function*" [1.8, p. 194]. One option would be to follow the quietist strategy concerning the ontology of models and "refuse to engage with this issue and ask, instead, how can we best represent these features [and functions of models] in order that we can understand" [1.7, p. 245] the practice of scientific modeling. Alternatively, however, one might simply accept that the function of models in scientific inquiry is our best – and perhaps only – guide when exploring answers to the question *what is a model?*. At the very least, it is not obvious that an exploration of the ontological aspects of models is necessarily fruitless or misguided. *Ducheyne* puts this nicely when he argues that [1.9, p. 120],

> "if we accept that models are functional entities, it should come as no surprise that when we deal with scientific models ontologically, we cannot remain silent on how such models function as carriers of scientific knowledge."

As a working assumption, then, let us treat scientific models as *functional entities* and explore how much ontological unity – over and above their *mere* functional role – we can give to the notion of *scientific model*.

Two broad classes of functional characterizations of models can be distinguished, according to which it is either *instantiation* or *representation* that lie at the heart of how models function. As *Giere* [1.10] sees it, on the *instantial view*, models instantiate the axioms of a theory, where the latter is understood as being comprised of linguistic statements, including mathematical statements and equations. (For an elaboration of how such an account might turn out, see Sect. 1.4.) By contrast, on the *representational view*, "language connects not directly with the world, but rather with a model, whose characteristics may be precisely defined"; the model then connects with the world "by way of similarity between a model and the designated parts of the world" [1.10, p. 156]. Other proponents of the representational view have de-emphasized the role of similarity, while still endorsing representation as one of the key functions of scientific models. Generally speaking, proponents of the representational view consider models to be "tools for *representing the world*," whereas those who favor the instantial view regard them primarily as "providing a means for interpreting formal systems" [1.10, p. 44].

Within the class of representational views, one can further distinguish between views that emphasize the *informational* aspects of models and those that take their *pragmatic* aspects to be more central. *Chakravartty* nicely characterizes the informational variety of the representational view as follows [1.11, p. 198]:

> "The idea here is that a scientific representation is something that bears an objective relation to the thing it represents, on the basis of which it contains information regarding that aspect of the world."

The term *objective* here simply means that the requisite relation obtains independently of the model user's beliefs or intentions as well as independently of the specific representational conventions he or she might be employing. Giere's similarity-based view of representation – according to which scientific models represent in virtue of their being similar to their target systems in certain specifiable ways – would be an example of such an informational view similarity, as construed by Giere, is a relation that holds between the model and its target, irrespective of a model user's beliefs or intentions, and regardless of the cognitive uses to which he or she might put the model. Other philosophical positions that are closely aligned with the informational approach might posit that, for a model to represent its target, the two must stand in a relation of isomorphism, partial isomorphism, or homomorphism to one another.

By contrast, the *pragmatic* variety of the representational view of models posits that models function as representations of their targets in virtue of the cognitive uses to which human reasoners put them. The basic idea is that a scientific model facilitates certain cognitive activities – such as the drawing of inferences about a target system, the derivation of predictions, or perhaps a deepening of the scientific understanding – on the part of its user and, therefore, necessarily involves the latter's cognitive interests, beliefs, or intentions. *Hughes* [1.12], for example, emphasizes the interplay of three cognitive–theoretical processes – denotation, demonstration, and interpretation – which jointly give rise to the representational capacity of (theoretical) models in science. On Hughes' (aptly named) *DDI account* of model-based representation, *denotation* accounts for the fact that theoretical elements of a model purport to refer to elements in the physical world. The possibility of *demonstration* from within a model – in particular, the successful mathematical derivation of results for models that lend themselves to mathematical derivation techniques – attests both to the models having a nontrivial internal dynamic and to its being a viable object of fruitful theoretical investigation. Through successful *interpretation*, a model user then relates the theoretically derived results back to the physical world, including the model's target system. Clearly, the DDI account depends crucially on there being someone who engages in the activities of interpreting and demonstrating – that is, it depends on the cognitive activities of human agents, who will inevitably draw on their background knowledge, cognitive interests, and derivational skills in establishing the requisite relations for bringing about representation.

The contrast between informational and pragmatic approaches to model-based representation roughly maps onto another contrast, between what *Knuuttila* has dubbed *dyadic* and *triadic* approaches. The former takes "the model–target dyad as a basic unit of analysis concerning models and their epistemic values" [1.13, p. 142]. This coheres well with the informational approach which, as discussed, tends to regard models as (often abstract) structures that stand in a relation of isomorphism, or partial isomorphism, to a target system. By contrast, triadic accounts – in line with pragmatic views of model-based representation – based representation shift attention away from models and the abstract relations they stand in, toward modeling as a theoretical activity pursued by human agents with cognitive interests, intentions, and beliefs. On this account, model-based representation cannot simply be a matter of any abstract relationship between the model and a target system since one cannot, as *Suárez* puts it, "reduce the essentially intentional judgments of representation users to facts about the source and target object or systems and their properties" [1.14, p. 768]. Therefore, so the suggestion goes, the model–target dyad needs to be replaced by a three-place relation between the model, its target, and the model user. Suárez, for example, proposes an inferentialist account of model-based representation, according to which a successful model must allow "competent and informed agents to draw specific inferences regarding" [1.14, p. 773] the target system – thereby making the representational success of a model dependent on the qualities of a (putative) model user.

1.3 Models as Analogies and Metaphors

Some scholars trace the emergence of the concept of a *scientific model* to the second half of the nineteenth century [1.15]. Applying our contemporary concept of *model* to past episodes in the history of science, we can of course identify prior instances of models being employed in science; however, until the nineteenth century scientists were engaged in little systematic self-reflection on the uses and limitations of models. Philosophy of science took even longer to pay attention to models in science, focusing instead on the role and significance of scientific theories. Only from the middle of the twentieth century onward did philosophical interest in models acquire the requisite momentum to carry the debate forward. Yet in both science and philosophy, the term *model* underwent important transformations, so it will be important to identify some of these shifts, in order to avoid unnecessary ambiguity and confusion in our exploration of the question *What is a model?*.

Take, for example, *Duhem*'s dismissal, in 1914, of what he takes to be the excessive use of models in Maxwell's theory of electromagnetism, as presented in an English textbook published at the end of the nineteenth century [1.16, p. 7]:

> "Here is a book intended to expound the modern theories of electricity and to expound a new theory. In it there are nothing but strings which move round pulleys which roll around drums, which go through pearl beads, which carry weights; and tubes which pump water while others swell and contract; toothed wheels which are geared to one another and engage hooks. We thought we were entering the tranquil and neatly ordered abode of reason, but we find ourselves in a factory."

What Duhem is mocking in this passage, which is taken from a chapter titled *Abstract Theories and Mechanical Models*, is a style of reasoning that is dominated by the desire to *visualize* physical processes in purely mechanical terms. His hostility is thus directed at *mechanical* models only – as the implied contrast in the chapter title makes clear – and does not extend to the more liberal understanding of the term *scientific model* in philosophy of science today.

Indeed, when it comes to the use of *analogy* in science, Duhem is much more forgiving. The term *analogy*, which derives from the Greek expression for *proportion*, itself has multiple uses, depending on whether one considers its use as a rhetorical device or as a tool for scientific understanding. Its general form is that of "pointing to a resemblance between relations in two different domains, that is, *A* is related to *B* like *C* is related to *D*" [1.17, p. 110]. An analogy may be considered merely *formal*, when only the relations (but not the relata) resemble another, or it may be *material*, when the relata from the two domains (i. e., *A* and *B* on one side, *C* and *D* on the other) have certain attributes or characteristics in common. *Duhem*'s understanding of *analogy* is more specific, in that he conceives of analogy as being a relation between two sets of statements, such as between one theory and another [1.16, p. 97]:

> "Analogies consist in bringing together two abstract systems; either one of them already known serves to help us guess the form of the other not yet known, or both being formulated, they clarify the other. There is nothing here that can astonish the most rigorous logician, but there is nothing either that recalls the procedures dear to ample but shallow minds."

Consider the following example: When Christiaan Huygens (1629–1695) proposed his theory of light, he did so on the basis of *analogy* with the theory of sound waves: the relations between the various attributes and characteristics of light are similar to those described by acoustic theory for the rather different domain of sound. Thus understood, analogy becomes a legitimate instrument for learning about one domain on the basis of what we know about another. In modern parlance, we might want to say that sound waves provided Huygens with a good *theoretical model* – at least given what was known at the time – for the behavior of light.

There is, however, a risk of ambiguity in that last sentence – an ambiguity which, as *Mellor* [1.18, p. 283] has argued, it would be wrong to consider harmless. Saying that *sound waves provide a good model for the theory of light* appears to equate the model *with the sound waves* – as though one physical object (sound waves) could be identified with the model. At first sight, this might seem unproblematic, given that, as far as wave-like behavior is concerned, we do take light and sound to be relevantly analogous. However, while it is indeed the case that "some of the constructs called *analogy* in the nineteenth century would today be routinely referred to as *models*" [1.19, p. 46], it is important to distinguish between, on the one hand, *analogy* as the similarity relation that exists between a theory and another set of statements and, on the other hand, the latter set of statements as the *analog of the theory*. Furthermore, we need to distinguish between the analog (e.g., the theory of sound waves, in Huygens's case) and the set of entities *of which the analog is true* (e.g., the sound waves themselves). (On this point, see [1.18, p. 283].) What Duhem resents about the naïve use of what he refers to as *mechanical models* is the hasty conflation of the visualized entities – (imaginary) pulleys, drums, pearl beads, and toothed wheels – with what is *in fact* scientifically valuable, namely the relation of analogy that exists between, say, the theory of light and the theory of sound.

This interpretation resolves an often mentioned tension – partly perpetuated by Duhem himself, through his identification of different styles of reasoning (the *English* style of physics with its emphasis on mechanical models, and the *Continental* style which prizes mathematical principles above all) – between Duhem's account of models and that of the English physicist Norman Campbell. Thus, *Hesse*, in her seminal essay *Models and Analogies in Science* [1.20], imagines a dialogue between a *Campbellian* and a *Duhemist*. At the start of the dialogue, the Campbellian attributes to the Duhemist the following view: "I imagine that along with most contemporary philosophers of science, you would wish to say that the use of models or analogs is not essential to scientific theorizing and that [. . .] the theory as a whole does not require to be interpreted by means of any model." To this, the Duhemist, who admits that "models may be useful guides in suggesting theories," replies: "When we have found an acceptable theory, any model that may have led us to it can be thrown away. Kekulé is said to have arrived at the structure of the benzene ring after dreaming of a snake with its tail in its mouth, but no account of the snake appears in the textbooks of organic chemistry." The Campbellian's rejoinder is as follows: "I, on the other hand, want to argue that models in some sense are essential to the logic of scientific theories" [1.20, pp. 8–9]. The quoted part of Hesse's dialogue has often been interpreted as suggesting that the bone of contention between Duhem and Campbell is the status of *models in general* (in the modern sense that includes theoretical models), with Campbell arguing in favor and Duhem arguing against. But we have already seen that Duhem, using the language of *analogy*, *does* allow for theoretical models to play an important role in science. This apparent tension can be resolved by being more precise about the target of Duhem's criticism: "Kekulé's snake dream might illustrate the use of a visualizable model, but it certainly does not illustrate the use of an analogy, in Duhem and Campbell's sense" [1.18, p. 285]. In other words, Duhem is not opposed to scientific models in general, but to its mechanical variety in particular. And, on the point of over-reliance on mechanical models, *Campbell*, too, recognizes that dogmatic attachment to such a style of reasoning is *open to criticism*. Such a dogmatic view would hold "that theories are completely satisfactory only if the analogy on which they are based is mechanical, that is to say, if the analogy is with the laws of mechanics" [1.21, p. 154]. Campbell is clearly more sympathetic than Duhem toward our "craving for mechanical theories," which he takes to be firmly rooted in our psychology. But he insists that [1.21, p. 156]

> "we should notice that the considerations which have been offered justify only the attempt to adopt some form of theory involving ideas closely related to those of force and motion; it does not justify the attempt to force all such theories into the Newtonian mold."

To be sure, significant differences between Duhem and Campbell remain, notably concerning what *kinds* of uses of analogies in science (or, in today's terminology, of scientific – including theoretical – models) are appropriate. For Duhem, such uses are limited to a heuristic role in the discovery of scientific theories. By contrast, *Campbell* claims that "in order that a theory may be valuable [. . .] it must display analogy" [1.21, p. 129] – though it should be emphasized again, not necessarily analogy *of the mechanical sort*. (As *Mellor* argues, Duhem and Campbell differ chiefly in their views of scientific theories and less so in their take on analogy, with Duhem adopting a more *static* perspective regarding theories and Campbell taking a more realist perspective [1.18].)

It should be said, though, that Hesse's *Campbellian* and *Duhemist* are at least partly intended as caricatures and serve as a foil for Hesse's own account of models as analogies. The account hinges on a three-part distinction between *positive*, *negative*, and *neutral* analogies [1.20]. Using the billiard ball model of gases as her primary example, Hesse notes that some characteristics are shared between the billiard balls and the gas atoms (or, rather, are ascribed by the billiard ball model to the gas atoms); these include velocity, momentum, and collision. Together, these constitute the *positive* analogy. Those properties we know to belong to billiard balls, but not to gas atoms – such as color – constitute the *negative* analogy of the model. However, there will typically be properties of the model (i. e., the billiard ball system) of which we do not (yet) know whether they also apply to its target (in this case, the gas atoms). These form the *neutral* analogy of the model. Far from being unimportant, the neutral analogy is crucial to the fruitful use of models in scientific inquiry, since it holds out the promise of acquiring new knowledge about the target system by studying the model in its place [1.20, p. 10]:

> "If gases are really like collections of billiard balls, except in regard to the known negative analogy, then from our knowledge of the mechanics of billiard balls, we may be able to make new predictions about the expected behavior of gases."

In dealing with scientific models we may choose to disregard the negative analogy (which results in what Hesse calls *model*$_1$) and consider only the known positive and neutral analogies – that is, only those properties that are shared, or for all we know may turn out to be shared, between the target system and its analog. (On the terminology discussed in Sect. 1.1, due to Black and Achinstein, model$_1$ would qualify as a *theoretical model*.) This, Hesse argues, typically describes our use of models for the purpose of explanation: we resolve to treat model$_1$ as taking the place of the phenomena themselves. Alternatively, we may actively include the negative analogy in our considerations, resulting in what Hesse calls *model*$_2$ or a form of analog model. Given that, let us assume, the model system (e.g., the billiard balls) was chosen because it was observable – or, at any rate, more accessible than the target system (e.g., the gas) – model$_2$ allows us to study the similarities and dissimilarities between the two analogous domains; model$_2$, qua being a model for its target, thus has a deeper structure than the system of billiard balls considered in isolation – and, like model$_1$, importantly includes the neutral analogy, which holds out the promise of novel insights and predictions. As *Hesse* puts it, in the voice of her Campbellian interlocutor [1.20, pp. 12–13]:

> "My whole argument is going to depend on these features [of the neutral analogy] and so I want to make it clear that I am not dealing with static and formalized theories, corresponding only to the known positive analogy, but with theories in the process of growth."

Models have been discussed not only in terms of analogy, but also in terms of metaphor. *Metaphor*, more explicitly than *analogy*, refers to the linguistic realm: a metaphor is a linguistic expression that involves at least one part that is being transferred from a domain of discourse where it is common to another – the target domain – where it is uncommon. The existence of an analogy may facilitate such a transfer of linguistic expression; at the same time, it is entirely possible that "it is the metaphor that prompts the recognition of analogy" [1.17, p. 114] – both are compatible with one another and neither is obviously prior to the other. Metaphorical language is widespread in science, not just in connection with models: for example, physicists routinely speak of *black holes* and *quantum tunneling* as important predictions of general relativity theory and quantum theory, respectively. Yet, as *Soskice* and *Harré* note, there is a special affinity between models and metaphor [1.22, p. 302]:

> "The relationship of model and metaphor is this: if we use the image of a fluid to explicate the supposed action of the electrical energy, we say that the fluid is functioning as a model for our conception of the nature of electricity. If, however, we then go on to speak of the *rate of flow* of an *electrical current*, we are using metaphorical language based on the fluid model."

In spite of this affinity, it would not be fruitful to simply equate the two – let alone jump to the conclusion that, in the notion of *metaphor*, we have found an answer to the question *What is a model?*. Models and metaphors both issue in descriptions, and as such they may draw on analogies we have identified between two otherwise distinct domains; more, however, needs to be said about the nature of the relations that need to be in place for something to be considered a (successful) model of its target system or phenomenon.

1.4 Models Versus the Received View: Sentences and Structures

Much of the philosophical debate about models is indebted to model theory as a branch of (first-order) mathematical logic. Two philosophical frameworks for thinking about scientific models and theories – the *syntactic view* of models and theories and its main competitor, the *semantic view* – can be traced back to these origins; they are the topic of this section. (For a more extensive discussion, see also other chapters in this handbook.) The syntactic view (Sect. 1.4.2) is closely aligned with logical positivism, which dominated much anglophone philosophy of science until the mid-1960s, and is sometimes referred to as *the received view*. Given that less rigid approaches and an overarching movement toward pluralism have reshaped the philosophy of science over the past half-century or so, this expression is somewhat dated; to make matters worse, other contributors to the debate have, over time, come to apply the same label to the syntactic view's main competitor, the semantic view of models and theories. Instead of adjudicating which position deserves this dubious honor, the present section will discuss how each view conceives of models. Before doing so, however, a few preliminaries are in order concerning the competing views' joint origins in logical model theory.

1.4.1 Models and the Study of Formal Languages

Model theory originated as the study of formal languages and their interpretations, starting from a Tarski-style truth theory based only on notions from syntax and set theory. On a broader understanding, the restriction to formal languages may be dropped, so as to include scientific languages (which are often closer to natural language than to logic), or even natural languages. However, the distinction between the syntax and the semantics of a language, which is sharpest in logic, also provides a useful framework for studying scientific languages and has guided the development of both the syntactic and the semantic views of theories and models. The *syntax* of a language L is made up of the vocabulary of L, along with the rules that determine which sequence of symbols counts as a well-formed expression in L; in turn, the *semantics* of L provides interpretations of the symbolic expressions in L, by mapping them onto another relational structure R, such that all well-formed expressions in L are rendered intelligible (e.g., via rules of composition) and can be assessed in terms of their truth or falsity in R.

The contrast between the syntax and the semantics of a language allows for two different approaches to the notion of a *theory*. A theory T may either be defined syntactically, as the set of all those sentences that can be derived, through a proper application of the syntactic rules, from a set of axioms (i. e., statements that are taken to be fundamental); or it may be defined semantically, as all those (first-order) sentences that a particular structure, M, satisfies. An example of the former would be Euclidean geometry, which consists of five axioms and all the theorems derivable from them using geometrical rules; an example of the latter would be group theory, which simply consists of all those first-order sentences that a set of groups – definable in terms of set-theoretic entities – satisfies. (This example, and much of the short summary in this section, is owed to [1.23]; for further discussion, see references therein.) The syntactic and semantic definitions of what a theory is are closely related: starting from the semantic definition, to see whether a particular structure M is a model of an axiomatizable first-order theory T, all that one needs to show is that M satisfies the axioms.

1.4.2 The Syntactic View of Theories

The syntactic view of theories originated from the combination of the insights – or, to put it a little more cautiously, fundamental tenets – of two research programs: the philosophical program, aligned with Pierre Duhem (Sect. 1.3) and Henri Poincaré, of treating (physical) theories as systems of hypotheses designed to *save the phenomena*, and the mathematical program, pioneered by David Hilbert, which sought to formalize (mathematical) theories as axiomatic systems. By combining the two, it seemed possible to identify a theory with the set of logical consequences that could be derived from its fundamental principles (which were to be treated as axioms), using only the rules of the language in which the theory was formulated. In spite of its emphasis on syntax, the syntactic view is not entirely divorced from questions of semantics. When it comes to scientific theories, we are almost always dealing with *interpreted* sets of sentences, some of which – the fundamental principles or axioms – are more basic than others, with the rest derivable using syntactic rules. The question then arises at which level interpretation of the various elements of a theory is to take place. This is where the slogan *to save the phenomena* points us in the right direction: on the syntactic view, interpretation only properly enters at the level of matching singular theoretical predictions, formulated in strictly observational terms, with the observable phenomena. Higher level interpretations – for example, pertaining to purely theoretical terms of a theory (such as posited unobservable entities, causal mechanisms, laws, etc.) – would be addressed through *correspondence rules*, which offered at least a partial interpretation, so that *some* of the meaning of such higher level terms of a theory could be linked up with observational sentences.

As an example, consider the example of classical mechanics. Similar to how Euclidean geometry can be fully derived from a set of five axioms, classical mechanics is fully determined by Newton's laws of mechanics. At a purely formal level, it is possible to provide a fully syntactic axiomatization in terms of the relevant symbols, variables, and rules for their manipulation – that is, in terms of what Rudolf Carnap calls the *calculus of mechanics*. If one takes the latter as one's starting point, it requires interpretation of the results derived from within this formal framework, in order for the calculus to be recognizable as a theory of mechanics, that is, of physical phenomena. In the case of mechanics, we may have no difficulty stating the axioms in the form of the (physically interpreted) *Newtonian laws of mechanics*, but in other cases – perhaps in quantum mechanics – making this connection with observables may not be so straightforward. As *Carnap* notes [1.24, p. 57]:

> "[t]he relation of this theory [= the physically interpreted theory of mechanics] to the calculus of mechanics is entirely analogous to the relation of physical to mathematical geometry. "

As in the Euclidean case, the syntactic view identifies the theory with a formal language or calculus (including, in the case of scientific theories, relevant correspondence rules), "whose interpretation – what the calculus is a theory *of* – is fixed at the point of application" [1.25, p. 125].

On the syntactic view of theories, models play at best a very marginal role as limiting cases or approximations. This is for two reasons. First, since the nonobservational part of the theory – that is, the *theory proper*, as one might put it – does not admit of direct interpretation, the route to constructing theoretical models on the basis of our directly interpreting the core ingredients of the theory is obstructed. Interpretation at the level of observational statements, while still available to us, is insufficient to imbue models with anything other than a purely *one-off* auxiliary role. Second, as *Cartwright* has pointedly argued in criticism directed at both the syntactic and the semantic views, there is a shared – mistaken – assumption that theories are a bit like vending machines [1.26, p. 247]:

> "[Y]ou feed it input in certain prescribed forms for the desired output; it gurgitates for a while; then it drops out the sought-for-representation, plonk, on the tray, fully formed, as Athena from the brain of Zeus."

This limits what we can do with models, in that there are only two stages [1.26, p. 247]:

> "First, eyeballing the phenomenon, measuring it up, trying to see what can be abstracted from it that has the right form and combination that the vending machine can take as input; secondly, [...] we do either tedious deduction or clever approximation to get a facsimile of the output the vending machine would produce."

Even if this caricature seems a little too extreme, the fact remains that, by modeling theories after first-order formal languages, the syntactic view limits our understanding of what theories and models are and what we can do with them.

1.4.3 The Semantic View

One standard criticism of the syntactic view is that it conflates scientific theories with their linguistic formulations. Proponents of the semantic view argue that by adding a layer of (nonlinguistic) structures between the linguistic formulations of theories and our assessment of them, one can side-step many of the problems faced by the syntactic view. According to the semantic view, a theory should be thought of as the set of set-theoretic structures that satisfy the different linguistic formulations of the theory. A structure that provides an interpretation for, and makes true, the set of sentences associated with a specific linguistic formulation of the theory is called a *model of the theory*. Hence, the semantic view is often characterized as conceiving of theories as *collections of models*. This not only puts models – where these are to be understood in the logical sense outlined earlier – center stage in our account of scientific theories, but also renders the latter fundamentally *extra-linguistic* entities.

An apt characterization of the semantic view is given by *Suppe* as follows [1.27, pp. 82–83]:

> "This suggests that theories be construed as propounded abstract *structures* serving as models for sets of interpreted sentences that constitute the linguistic formulations. [...] [W]hat the theory does is directly describe the behavior of abstract systems, known as *physical systems*, whose behaviors depend only on the selected parameters. However, physical systems are abstract replicas of actual phenomena, being what the phenomena *would have been* if no other parameters exerted an influence."

According to a much-quoted remark by one of the main early proponents of the semantic view, *Suppes*, "the meaning of the concept of model is the same in mathematics and in the empirical sciences." However, as *Suppe*'s quote above makes clear, models in science have additional roles to play, and it is perhaps worth noting that *Suppes* himself immediately continues: "The difference to be found in these disciplines is to be found in their use of the concept" [1.28, p. 289]. Supporters of the semantic view often claim that it is closer to the scientific practices of modeling and theorizing than the syntactic view. On this view, according to *van Fraassen* [1.29, p. 64],

> "[t]o present a theory is to specify a family of structures, its *models*; and secondly, to specify certain parts of those models (the *empirical substructures*) as candidates for the direct representation of observable phenomena."

Unlike what the syntactic view suggests, scientists do not typically formulate abstract theoretical axioms and only interpret them at the point of their application to observable phenomena; rather, "scientists build in their mind's eye systems of abstract objects whose properties or behavior satisfy certain constraint (including law)" [1.23, p. 154] – that is, they engage in the construction of theoretical models.

Unlike the syntactic view, then, the semantic view appears to give a more definite answer to the question *what is a model?* In line with the account sketched so far, *a model of a theory is simply a (typically extra-*

linguistic) structure that provides an interpretation for, and makes true, the set of axioms associated with the theory (assuming that the theory is axiomatizable). Yet it is not clear that, in applying their view to actual scientific theories, the semanticists always heed their own advice to treat models as both *giving an interpretation*, and *ensuring the truth*, of a set of statements. More importantly, the model-theoretic account demands that, in a manner of speaking, a model should fulfil its truth-making function *in virtue of* providing an interpretation for a set of sentences. Other ways of ensuring truth – for example by limiting the domain of discourse for a set of fully interpreted sentences, thereby ensuring that the latter will happen to be true – should not qualify. Yet, as *Thomson-Jones* [1.30] has argued, purported applications of the semantic view often stray from the original model-theoretic motivation. As an example, consider Suppes' *axiomatization* of Newtonian particle physics. (The rest of this subsection follows [1.30, pp. 530–531].) *Suppes* [1.31] begins with the following definition (in slightly modified form)

Definition 1.1
A system $\beta = \langle P, T, s, m, f, g \rangle$ is a model of particle mechanics if and only if the following seven axioms are satisfied:

Kinematical axioms:

1 The set P is finite and nonempty
2 The set T is an interval of real numbers
3 For p in P, s_p is twice differentiable.

Dynamical axioms:

4 For p in P, $m(p)$ is a positive real number
5 For p and q in P and t in T,

$$f(p,q,t) = -f(q,p,t)\ .$$

6 For p and q in P and t in T,

$$s(p,t) \times f(p,q,t) = -s(q,t) \times f(q,p,t)\ .$$

7 For p in P and t in T,

$$m(p)D^2 s_p(t) = \sum_{q \in P} f(p,q,t) + g(p,t)\ .$$

At first sight, this presentation adheres to core ideas that motivate the semantic view. It sets out to define an extra-linguistic entity, β, in terms of a set-theoretical predicate; the entities to which the predicate applies are then to be singled out on the basis of the seven axioms. But as *Thomson-Jones* points out, a specific model S defined in this way "is not a serious interpreter of the predicate or the *axioms* that compose it" [1.30, p. 531]; it merely fits a structure to the description provided by the fully interpreted axioms (1)–(7), and in this way ensures that they are satisfied, but it does not make them come out true in virtue of providing an interpretation (i. e., by invoking semantic theory). To *Thomson-Jones*, this suggests that identifying scientific models with truth-making structures in the model-theoretic sense may, at least in the sciences, be an unfulfilled promise of the semantic view; instead, he argues, we should settle for a less ambitious (but still informative) definition of a model as "a mathematical structure used to represent a (type of) system under study" [1.30, p. 525].

1.4.4 Partial Structures

Part of the motivation for the semantic view was its perceived greater ability to account for how scientists actually go about developing models and theories. Even so, critics have claimed that the semantic view is unable to accommodate the great diversity of scientific models and faces special challenges from, for example, the use of inconsistency in many models. In response to such criticisms, a philosophical research program has emerged over the past two decades, which seeks to establish a *middle ground* between the classical semantic view of models discussed in the previous section and those who are sceptical about the prospects of formal approaches altogether. This research program is often called the *partial structures approach*, which was pioneered by *Newton da Costa* and *Steven French* and whose vocal proponents include Otávio Bueno, James Ladyman, and others; see [1.32] and references therein.

Like many adherents of the semantic view, partial structures theorists hold that models are to be reconstructed in set-theoretic terms, as ordered n-tuples of sets: a set of objects with (sets of) properties, quantities and relations, and functions defined over the quantities. A *partial structure* may then be defined as $\mathfrak{A} = \langle D, R_i \rangle_{i \in \mathrm{I}}$, where D is a nonempty set of n-tuples of just this kind and each R_i is a n-ary relation. Unlike on the traditional semantic view, the relations R_i need not be complete isomorphisms, but crucially are *partial relations*: that is, they need not be defined for all n-tuples of elements of D. More specifically, for each partial relation R_i, in addition to the set of n-tuples for which the relation holds and the set of n-tuples for which it does not hold, there is also a third set of n-tuples for which it is underdetermined whether or not it holds. (There is a clear parallel here with Hesse's notion of positive, negative, and neutral analogies which, as *da Costa* and *French* put it, "finds a natural home in the context of partial structures" [1.32, p. 48].) A total structure is said to *extend* a partial structure, if it subsumes the first two sets without change (i. e., includes all those objects and definite relations that exist in the partial structures) and renders each extended relation well defined for every n-tuple of objects in its domain. This gives rise to a hierarchy of structures and substructures, which together with the notion of partial isomorphism loosens the requirements on representation, since all that is needed for two partial models A and A' to be *partially* isomorphic is that a partial substructure of A be isomorphic to a partial substructure in A'.

Proponents of the partial structures approach claim that it "widens the framework of the model-theoretic approach and allows various features of models and theories – such as analogies, iconic models, and so on – to be represented," [1.33, p. 306] that it can successfully contain the difficulties arising from inconsistencies in models, and that it is able to capture "the existence of a hierarchy of models stretching from the data up to the level of theory" [1.33]. Some critics have voiced criticism about such sweeping claims. One frequent criticism concerns the proliferation of partial isomorphisms, many of which will trivially obtain; however, if partial relations are so easy to come by, how can one tell the interesting from the vast majority of irrelevant ones? (*Pincock* speaks in this connection of the "danger of trivializing our representational relationships" [1.34, p. 1254].) *Suárez* and *Cartwright* add further urgency to this criticism, by noting that the focus on set-theoretical structures obliterates all those uses of models and aspects of scientific practice that do not amount to the making of claims [1.35, p. 72]:

> "So all of scientific practice that does not consist in the making of claims gets left out. [...] Again, we maintain that this inevitably leaves out a great deal of the very scientific practice that we are interested in."

It is perhaps an indication of the limitations of the partial structures approach that, in response to such criticism, its proponents need to again invoke heuristic factors, which cannot themselves be subsumed under the proposed formal framework of models as set-theoretic structures with partial relations.

1.5 The Folk Ontology of Models

If we accept that scientific models are best thought of as functional entities (Sect. 1.2), perhaps something can be learnt about the ontology of scientific models from looking at their functional role in scientific inquiry. What one finds across a range of different kinds of models is the practice of taking models as stand-ins for systems that are not, in fact, instantiated. As *Godfrey-Smith* puts it, "modelers often *take* themselves to be describing imaginary biological populations, imaginary neural networks, or imaginary economies" [1.36, p. 735] – that is, they are aware that due to idealization and abstraction, model systems will differ in their descriptions from a full account of the actual world. A model, thus understood, may be thought of as a "description of a missing system," and the corresponding research practice of describing and characterizing model systems *as though* they were real instantiated systems (even though they are not) may be called, following *Thomson-Jones*, the "face-value practice" of scientific modeling [1.37, pp. 285–286].

On the heels of the face-value practice of scientific modeling, it has been argued, comes a common – though perhaps not universally shared – understanding of *what models are* [1.36, p. 735]:

> "[...] to use a phrase suggested by Deena Skolnick, the treatment of model systems as comprising imagined concrete things is the *folk ontology* of at least many scientific modelers. It is the ontology embodied in many scientists' unreflective habits of talking about the objects of their study-talk about what a certain kind of population will do, about whether a certain kind of market will clear. [...O]ne kind of understanding of model-based science requires that we take this *folk ontology* seriously, as part of the scientific strategy."

The ontology of *imagined concrete things* – that is, of entities that, *if real*, would be on a par with concrete objects in the actual world – leads quickly into the thorny territory of fictionalism. *Godfrey-Smith* is explicit about this when he likens models to "something we are all familiar with, the imagined objects of literary fiction" [1.36] – such as Sherlock Holmes, J.R.R. Tolkien's Middle Earth, and so on. Implicit in this suggestion is, of course, a partial answer to our question *What is a model?* – namely, that the ontological status of scientific models is *just like* that of literary (or other) fictions. The advantages and disadvantages of such a position will be discussed in detail in Sect. 1.6 of this chapter.

There is, however, another direction into which a closer analysis of the face-value practice can take us. Instead of focusing on the ontological status of the entities we are imagining when we contemplate models as imagined concrete things, we can focus on the conscious processes that attend such imaginings (or, if one prefers a different way of putting it, the *phenomenology* of interacting with models). Foremost among these is the mental imagery that is conjured up by the descriptions of models. (Indeed, as we shall see in the next section, on certain versions of the fictionalist view, a model *prescribes* imaginings about its target system.) How much significance one should attach to the mental pictures that attend our conscious consideration of models has been a matter of much controversy: recall Duhem's dismissal of mechanical imagery as a way of conceptualizing electromagnetic phenomena (Sect. 1.3).

Focusing on the mental processes that accompany the use of scientific models might lead one to propose an analysis of models in terms of their cognitive foundations. Nancy Nersessian has developed just such an analysis, which ties the notion of models in science closely to the cognitive processes involved in mental modeling. Whereas the traditional approach in psychology had been to think of reasoning as consisting of the mental application of logical rules to propositional representations, mounting empirical evidence of the role of heuristics and biases suggested that much of human reasoning proceeds via *mental models* [1.38], that is, by carrying out thought experiments on internal models. A *mental model*, on this account, is "a structural analog of a real-world or imaginary situation, event, or process" as constructed by the mind in reasoning (and, presumably, realized by certain underlying brain processes) [1.39, pp. 11–12]:

> "What it means for a mental model to be a structural analog is that it embodies a representation of the spatial and temporal relations among, and the causal structures connecting the events and entities depicted and whatever other information that is relevant to the problem-solving talks. [...] The essential points are that a mental model can be nonlinguistic in form and the mental mechanisms are such that they can satisfy the model-building and simulative constraints necessary for the activity of mental modeling."

While this characterization of mental models may have an air of circularity, in that it essentially defines mental models as place-holders for *whatever it takes* to support *the activity of mental modeling*, it nonetheless suggests a place to look for the materials from which models are constructed: the mind itself, with its various types of content and mental representation. As *Nersessian* puts it: "Whatever the format of the model itself, information in various formats, including linguistic, formulaic, visual, auditory, kinesthetic, can be used in its construction" [1.39, p. 12].

How does this apply to the case of *scientific* models? As an example, Nersessian considers James Clerk Maxwell's famous molecular vortex model, which visualized the lines of magnetic force around a magnet as though they were vortices within a continuous fluid (Fig. 1.1).

As Nersessian sees it, Maxwell's drawing "is a *visual* representation of an *analogical* model that is accompanied with instructions for *animating* it correctly in thought" [1.39, p. 13]. And indeed *Maxwell* gives detailed instructions regarding how to interpret, and bring to life, the model of which the reader is only given a momentary *snapshot* [1.40, p. 477]:

> "Let the current from left to right commence in *AB*. The row of vortices *gh* above *AB* will be set in motion in the opposite direction to a watch [...]. We shall suppose the two of vortices *kl* still at rest, then the layer of particles between these rows will be acted on by the row *gh*,"

and so forth. It does seem plausible to say that such instructions are intended to prescribe certain mental models on the part of the reader. Convincing though this example may be, it still begs the question of what, *in general*, a mental model is. At the same time, it illustrates what is involved in conjuring up a mental model and which materials – in this case, spatial representations, along with intuitions about the mechanical motion of parts in a larger system – are involved in its constitution.

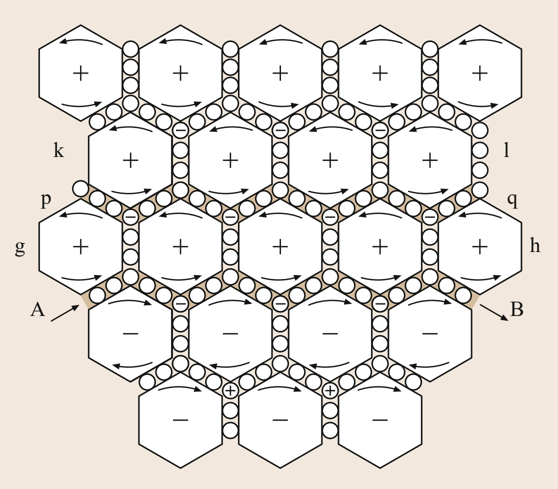

Fig. 1.1 Maxwell's drawing of the molecular vortex model (after [1.40])

1.6 Models and Fiction

As noted in the previous section, the face-value practice of scientific modeling and its concomitant folk ontology, according to which models are imagined concrete things, have a natural affinity to the way we think about fictions. As one proponent of models as fictions puts it [1.41, p. 253]:

> "The view of model systems that I advocate regards them as imagined physical systems, that is, as hypothetical entities that, as a matter of fact, do not exist spatiotemporally but are nevertheless not purely mathematical or structural in that they would be physical things if they were real."

Plausible though this may sound, the devil is in the details. A first – perhaps trivial – caveat concerns the restriction that model systems *would be physical things if they were real*. In order to allow for the notion of model to be properly applied to the social and cognitive sciences, such as economics and psychology, it is best to drop this restriction to physical systems. (On this point, see [1.30, p. 528].) This leaves as the gist of the folk-ontological view the thought that model systems, *if* they were real, would be *just as we imagine them* (or, more carefully, *just as the model instructs us to imagine them*).

In order to sharpen our intuitions about fictions, let us introduce an example of a literary fiction, such as the following statement from *Doyle*'s *The Adventure of the Three Garridebs* (1924) [1.42]: "Holmes had lit his pipe, and he sat for some time with a curious smile upon his face." There is, of course, no actual human being that this statement represents: no one is sitting smilingly at 221B Baker Street, filling up the room with smoke from their pipe. (Indeed, until the 1930s, the address itself had no real-world referent, as the highest number on Baker Street then was No. 85.) And yet there is a sense in which this passage does seem to represent Sherlock Holmes and, within the context of the story, tells us something informative about him. In particular, it seems to lend support to certain statements about Sherlock Holmes as opposed to others. If we say *Holmes is a pipe smoker*, we seem to be asserting something true about him, whereas if we say *Holmes is a nonsmoker*, we appear to be asserting something false. One goal of the ontology of fictions is to make sense of this puzzle.

Broadly speaking, there are two kinds of philosophical approaches – realist and antirealist – regarding fictions. On the realist approach, even though Sherlock Holmes is not an actual human being, we must grant that he *does* exist in some sense. Following *Meinong* [1.43], we might, for example, distinguish between *being* and *existence* and consider Sherlock Holmes to be an object that has all the requisite properties we normally attribute to him, except for the property of existence. Or we might take fictions to have existence, but only as abstract entities, not as objects in space and time. By contrast, antirealists about fictions deny that they have independent being or existence and instead settle for other ways of making sense of how we interpret fictional discourse. Following Bertrand Russell, we might paraphrase the statement *Sherlock Holmes is a pipe smoker and resides at 221B Baker Street* without the use of a singular term (*Sherlock Holmes*), solely in terms of a suitably quantified existence claim: *There exists one and only one x such that x is a pipe smoker and x resides at 221B Baker Street*. However, while this might allow us to parse the meaning of further statements about Sherlock Holmes more effectively, it does not address the puzzle that certain claims (such as *He is a pipe smoker*) ring true, whereas others do not – since it renders each part of the explicated statement false. This might not seem like a major worry for the case of literary fictions, but it casts doubt on whether we can fruitfully think about scientific models in those terms, given the epistemic role of scientific models as contributors to scientific knowledge.

In recent years, an alternative approach to fictions has garnered the attention of philosophers of science, which takes *Walton*'s notion of "games of make-believe" as its starting point. *Walton* introduces this notion in the context of his philosophy of art, where he characterizes (artistic) representations as "things possessing the social function of serving as props in games of make-believe" [1.44, p. 69]. In games of make-believe, participants engage in behavior akin to children's pretend play: when a child uses a banana as a telephone *to call grandpa*, this action does not amount to actually calling her grandfather (and perhaps not even *attempting* to call him); rather, it is a move within the context of play – where the usual standards of realism are suspended – whereby the child resolves to treat the situation *as if* it were one of speaking to her grandfather on the phone.

The banana is simply a prop in this game of make-believe. The use of the banana as a make-believe telephone may be inspired by some physical similarity between the two objects (e.g., their elongated shape, or the way that each can be conveniently held to one's ear and mouth at the same time), but it is clear that props can go beyond material objects to include, for example, linguistic representations (as would be the case with the literary figure of Sherlock Holmes). While the rules governing individual pretend play may be ad hoc, communal games of make-believe are structured by shared normative principles which *authorize* certain moves as legitimate, while excluding other moves as illegitimate. It is in virtue of such principles that fictional truths can be generated: for example, a toy model of a bridge at the scale of 1 : 1000 prescribes that, "if part of the model has a certain length, then, fictionally, the corresponding part of the bridge is a thousand times that length" [1.45, p. 38] – in other words, even though the model itself is only a meter long, it *represents* the bridge *as* a thousand meters long. Note that the scale model could be a model of a bridge that is yet to be built – in which case it would still be true that, fictionally, the bridge is a thousand meters long: props, via the rules that govern them, *create* fictional truths.

One issue of contention has been what kinds of metaphysical commitments such a view of models entails. Talk of *imagined concrete things* as the material from which models are built has been criticized for amounting to an indirect account of modeling, by which [1.46, pp. 308, fn. 14]

> "prepared descriptions and equations of motion ask us to imagine an *imagined concrete system* which then bears some other form of representation relation to the system being modelled."

A more thoroughgoing direct view of models as fictions is put forward by *Toon*, who considers the following sentence from *Wells*'s *The War of the Worlds*: "The dome of St. Paul's was dark against the sunrise, and injured, I saw for the first time, by a huge gaping cavity on its western side" [1.47, p. 229]. As *Toon* argues [1.46, p. 307]:

> "There is no pressure on us to postulate a fictional, damaged, St. Paul's for this passage to represent; the passage simply represents the actual St. Paul's. Similarly, on my account, our prepared description and equation of motion do not give rise to a fictional, idealised bouncing spring since they represent the actual bouncing spring."

By treating models as prescribing imaginings about *the actual objects* (where these exist and are the model's target system), we may resolve to imagine all sorts of things that are, as a matter of fact, false; however, so the direct view holds, this is nonetheless preferable to the alternative option of positing *independently existing* fictional entities [1.45, p. 42]. Why might one be tempted to posit, as the indirect view does, that fictional objects fitting the model descriptions must exist? An important motivation has to do with the assertoric force of our model-based claims. As *Giere* puts it: "If we insist on regarding principles as genuine statements, we have to find something that they describe, something to which they refer" [1.48, p. 745]. In response, proponents of the direct view have disputed the need "to regard theoretical principles formulated in modeling as genuine statements"; instead, as *Toon* puts it, "they are prescriptions to imagine" [1.45, p. 44].

One potential criticism the models as fictions view needs to address is the worry that, by focusing on the user's imaginings, what a model is becomes an entirely subjective matter. A similar worry may be raised with respect to the mental models view discussed in Sect. 1.5: if a model is merely a place-holder for whatever is needed to sustain the activity of mental modeling (or imagining) on the part of an agent, how can one be certain that the same kinds of models (or props) reliably give rise to the same kinds of mental modeling (or imaginings)? In this respect, at least, the models as fictions view appears to be in a stronger position. Recall that, unlike in individual pretend play (or unconstrained imagining), in games of make-believe certain imaginations are sanctioned by the prop itself and the – public, shared – rules of the game. As a result, "someone's imaginings are governed by intersubjective rules, which guarantee that, as long as the rules are respected, everybody involved in the game has the same imaginings" [1.41, p. 264] – though it should be added, not necessarily the same *mental images*.

In his 1963 book, *Models and Metaphors*, *Black* expressed his hope that an "exercise of the imagination, with all its promise and its dangers" may help pave the way for an "understanding of scientific models and archetypes" as "a reputable part of scientific culture" [1.4, p. 243]. Even though Black was writing in general terms (and perhaps for rhetorical effect), his characterization would surely be considered apt by the proponents of the models as fictions view, who believe that models allow us to imagine their targets to be a certain way, and that, by engaging in such imaginings, we can gain new scientific insights.

1.7 Mixed Ontologies: Models as Mediators and Epistemic Artifacts

In Sect. 1.1, a distinction was drawn between *informational* views of models, which emphasize the objective, two-place relation between the model and what it represents, and *pragmatic* views, according to which a model depends at least in part on the user's beliefs or intentions, thereby rendering model-based representation a three-place relation between model, target, and user. Unsurprisingly, which side one comes down on in this debate will also have an effect on one's take on the ontology of scientific models. Hence, structuralist approaches (e.g., the partial structures approach discussed in Sect. 1.4.4) are a direct manifestation of the informational view, whereas the models as fictions approach – especially insofar as it considers models to be props for the user's imagination – would be a good example of the pragmatic view. The pragmatic dimension of scientific representation has received growing attention in the philosophical literature, and while this is not the place for a detailed survey of pragmatic accounts of model-based representation in particular, the remainder of this section will be devoted to a discussion of the ontological consequences of several alternative pragmatic accounts of models. Particular emphasis will be placed on what I shall call *mixed ontologies*, that is, accounts of models that emphasize the heterogeneity and diversity of their components.

1.7.1 Models as Mediators

Proponents of pragmatic accounts of models usually take scientific practice as the starting point of their analysis. This often directly informs how they think about models; in particular, it predisposes them to treat models as the outcome of a process of model construction. On this view, it is not only the *function* of models – for example, their capacity to represent target systems – which depends on the beliefs, intentions, and cognitive interests of a model user, but also the very *nature* of models which is dependent on human agents in this way. In other words, what models are is crucially determined by their being the result of a deliberate process of model construction. Model construction, most pragmatic theorists of models insist, is marked by "piecemeal borrowing" [1.35, p. 63] from a range of different domains. Such conjoining of heterogeneous components to form a model cannot easily be accommodated by structuralist accounts, or so it has been claimed; at the very least, there is considerable tension between, say, the way that the partial structures approach allows for a nested *hierarchy* of models (connected with one another via partial isomorphisms) and the much more ad hoc manner in which modelers piece together models from a variety of ingredients. (On this point, see especially [1.35, p. 76].)

A number of such accounts have coalesced into what has come to be called the *models as mediators* view (see [1.49] for a collection of case studies). According to this view, models are to be regarded neither as a merely auxiliary intermediate step in applying or interpreting scientific theories, nor as constructed purely from data. Rather, they are thought of as mediating between our theories and the world in a partly autonomous manner. As *Morrison* and *Morgan* put it, models "are *not* situated in the middle of an hierarchical structure between theory and the world," but operate outside the hierarchical "theory-world axis" [1.50, pp. 17–18]. A central tenet of the models as mediators view is the thesis that models "are made up from a *mixture* of elements, including those from outside the domain of investigation"; indeed, it is thought to be precisely in virtue of this heterogeneity that they are able to retain "an element of independence from both theory and data (or phenomena)" [1.50, p. 23].

At one level, the models as mediators view appears to be making a descriptive point about scientific practice. As *Morrison* and *Morgan* [1.50] point out, there is "no *logical* reason why models should be constructed to have these qualities of partial independence" [1.50, p. 17], though in practice they do exhibit them, and examples that involve the integration of heterogeneous elements beyond theory and data "are not the exception but the rule" [1.50, p. 15]. Yet, there is also the further claim that models could not fulfil their epistemic function *unless* they are partially autonomous entities: "we can only expect to use models to learn about our theories or our world if there is at least partial independence of the model from both" [1.50, p. 17]. Given that models are functional entities (in the sense discussed in Sect. 1.2), this has repercussions for the ontological question of what kind of entities models are. More often than not, models will integrate – perhaps imperfectly, but in irreducible ways – heterogeneous components from disparate sources, including (but not limited to) "elements of theories and empirical evidence, as well as stories and objects which could form the basis for modeling decisions" [1.50, p. 15]. As proponents of the models as mediators view are at pains to show, even in cases where models initially seem to derive straightforwardly from fundamental theory or empirical data, closer inspection reveals the presence of other elements – such as "simplifications and approximations which have to be decided independently of the theoretical requirements or of data conditions" [1.50, p. 16].

For the models as mediators approach, any answer to the question *what is a model?* must be tailored to the specific case at hand: models in high-energy physics will have a very different composition, and will consist of an admixture of different elements, than, say, models in psychology. However, as a general rule, no model – or, at any rate, no *interesting* model – will ever be fully reducible to theory and data; attempts to *clean up* the ontology of scientific models so as to render them either purely theoretical or entirely empirical, according to the models as mediators view, misconstrue the very nature and function of models in science.

1.7.2 Models as Epistemic Artifacts

A number of recent pragmatic approaches take the models as mediators view as their starting point, but suggest that it should be extended in various ways. Thus, *Knuuttila* acknowledges the importance of mediation between theory and data, but a richer account of models is needed to account for how this partial independence comes about. For *Knuuttila*, *materiality* is the key enabling factor that imbues models with such autonomy: it is "the material dimension, and not just *additional elements*, that makes models able to mediate" [1.51, p. 48]. Materiality is also seen as explaining the various epistemic functions that models have in inquiry, not least by way of analogy with scientific experiments. For example, just as in experimentation much effort is devoted to minimizing unwanted external factors (such as noise), in scientific models certain methods of approximation and idealization serve the purpose of neutralizing undesirable influences. Models typically draw on variety of formats and representations, in a way that *enables* certain specific uses, but at the same time *constrains* them; this breaks with the traditional assumption that we can "clearly tell apart those features of our scientific representations that are attributable to the phenomena described from the conventions used to describe them" [1.52, p. 268].

On the account sketched thus far, attempting to characterize the nature and function of models in the language of theories and data would, in the vast majority of cases, give a misleading impression; instead, models are seen as *epistemic tools* [1.52, p. 267]:

> "Concrete artifacts, which are built by various representational means, and are constrained by their design in such a way that they enable the study of certain scientific questions and learning through constructing and manipulating them."

This links the philosophical debate about models to questions in the philosophy of technology, for example concerning the ontology of artifacts, which are likewise construed as both material bodies and functional objects. It also highlights the constitutive role of design and construction, which applies equally to models with a salient material dimension – such as scale models in engineering or ball-and-stick models in chemistry – and to largely theoretical models. For example, it has been argued that mathematical models (e.g., in many-body physics) may be fruitfully characterized not only in theoretical terms (say, as a Hamiltonian) or as mathematical entities (as an operator equation), but also as the output of a *mature mathematical formalism* (in this case, the formalism of second quantization) – that is, a physically interpreted set of notational rules that, while embodying various theoretical assumptions, is not usually reducible to fundamental theory [1.53].

As in the case of the models as mediators approach, the ontological picture that emerges from the artifactual approach to models is decidedly mixed: models will typically consist of a combination of different materials, media and formats, and deploy different representational means (such as pictorial, symbolic, and diagrammatic notations) as well as empirical data and theoretical assumptions. Beyond merely acknowledging the heterogeneity of such a *mixture of elements*, however, the artifactual approach insists that it is *in virtue of their material dimension* that the various elements of a model, taken together, enable and constrain its representational and other epistemic functions.

1.8 Summary

As the survey in this chapter demonstrates, the term *model* in science refers to a great variety of things: physical objects such as scale models in engineering, descriptions and sets of sentences, set-theoretic structures, fictional objects, or an assortment of all of the above. This makes it difficult to arrive at a uniform characterization of models *in general*. However, by paying close attention to philosophical accounts of model-based representation, it is possible to discern certain clusters of positions. At a general level, it is useful to think of models as functional entities, as this allows one to explore how different functional perspectives lead to different conceptions of the ontology of models. Hence, with respect to the representational function of models, it is possible to distinguish between *informational* views, which we found to be closely associated with structuralist accounts of models, and *pragmatic* views, which tend to give rise to more heterogeneous accounts, according to which models may be thought of as *props for the imagination*, as partly autonomous mediators between theory and data, or as epistemic artifacts consisting of an admixture of heterogeneous elements.

When nineteenth century physicists began to reflect systematically on the role of *analogy* in science, they did so out of a realization that it may not always be possible to apply fundamental theory directly to reality, either because any attempt to do so faces insurmountable complexities, or because no such fundamental theory is as yet available. At the beginning of the twenty-first century, these challenges have not diminished, and scientists find themselves turning to an ever greater diversity of scientific models, a unified philosophical theory of which is still outstanding.

References

1.1 R. Frigg: Models in science. In: *Stanford Encyclopedia of Philosophy*, ed. by E.N. Zalta http://plato.stanford.edu/entries/models-science/ (Spring 2012 Edition)

1.2 N. Goodman: *Languages of Art* (Bobbs-Merrill, Indianapolis 1968)

1.3 R. Ankeny, S. Leonelli: What's so special about model organisms?, Stud. Hist. Philos. Sci. **42**(2), 313–323 (2011)

1.4 M. Black: *Models and Metaphors: Studies in Language and Philosophy* (Cornell Univ. Press, Ithaca 1962)

1.5 P. Achinstein: *Concepts of Science: A Philosophical Analysis* (Johns Hopkins, Baltimore 1968)

1.6 J. von Neumann: Method in the physical sciences. In: *Collected Works Vol. VI. Theory of Games, Astrophysics, Hydrodynamics and Meteorology*, ed. by A.H. Taub (Pergamon, Oxford 1961) pp. 491–498

1.7 S. French: Keeping quiet on the ontology of models, Synthese **172**(2), 231–249 (2010)

1.8 G. Contessa: Editorial introduction to special issue, Synthese **2010**(2), 193–195 (2010)

1.9 S. Ducheyne: Towards an ontology of scientific models, Metaphysica **9**(1), 119–127 (2008)

1.10 R. Giere: Using models to represent reality. In: *Model-Based Reasoning in Scientific Discovery*, ed. by L. Magnani, N. Nersessian, P. Thagard (Plenum, New York 1999) pp. 41–57

1.11 A. Chakravartty: Informational versus functional theories of scientific representation, Synthese **217**(2), 197–213 (2010)

1.12 R.I.G. Hughes: Models and representation, Proc. Philos. Sci., Vol. 64 (1997) pp. S325–226

1.13 T. Knuuttila: Some consequences of the pragmatist approach to representation. In: *EPSA Epistemology and Methodology of Science*, ed. by M. Suárez, M. Dorato, M. Rédei (Springer, Dordrecht 2010) pp. 139–148

1.14 M. Suárez: An inferential conception of scientific representation, Proc. Philosophy of Science, Vol. 71 (2004) pp. 67–779

1.15 M. Jammer: Die Entwicklung des Modellbegriffs in den physikalischen Wissenschaften, Stud. Gen. **18**(3), 166–173 (1965)

1.16 P. Duhem: *The Aim and Structure of Physical Theory* (Princeton Univ. Press, Princeton 1954), Transl. by P.P. Wiener

1.17 D. Bailer-Jones: Models, metaphors and analogies. In: *The Blackwell Guide to the Philosophy of Science*, ed. by P. Machamer, M. Silberstein (Blackwell, Oxford 2002) pp. 108–127

1.18 D.H. Mellor: Models and analogies in science: Duhem versus Campbell?, Isis **59**(3), 282–290 (1968)

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

1.20 M. Hesse: *Models and Analogies in Science* (Sheed Ward, London 1963)

1.21 N.R. Campbell: *Physics: The Elements* (Cambridge Univ. Press, Cambridge 1920)

1.22 J.M. Soskice, R. Harré: Metaphor in science. In: *From a Metaphorical Point of View: A Multidisciplinary Approach to the Cognitive Content of Metaphor*, ed. by Z. Radman (de Gruyter, Berlin 1995) pp. 289–308

1.23 C. Liu: Models and theories I: The semantic view revisited, Int. Stud. Philos. Sci. **11**(2), 147–164 (1997)

1.24 R. Carnap: *Foundations of Logic and Mathematics* (Univ. Chicago Press, Chicago 1939)

1.25 R. Hendry, S. Psillos: How to do things with theories: An interactive view of language and models in science. In: *The Courage of Doing Philosophy: Essays Presented to Leszek Nowak*, ed. by J. Brzeziński, A. Klawiter, T.A.F. Kuipers, K. Lastowksi, K. Paprzycka, P. Przybyzs (Rodopi, Amsterdam 2007) pp. 123–158

1.26 N. Cartwright: Models and the limits of theory: Quantum hamiltonians and the BCS model of superconductivity. In: *Models as Mediators*, ed. by M. Morrison, M. Morgan (Cambridge Univ. Press, Cambridge 1999) pp. 241–281

1.27 F. Suppe: *The Semantic Conception of Theories and Scientific Realism* (Univ. Illinois Press, Urbana 1989)

1.28 P. Suppes: A comparison of the meaning and uses of models in mathematics and the empirical sciences, Synthese **12**(2/3), 287–301 (1960)

1.29 B. van Fraassen: *The Scientific Image* (Oxford Univ. Press, Oxford 1980)

1.30 M. Thomson-Jones: Models and the semantic view, Philos. Sci. **73**(4), 524–535 (2006)

1.31 P. Suppes: *Introduction to Logic* (Van Nostrand, Princeton 1957)

1.32 N. da Costa, S. French: *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning* (Oxford Univ. Press, New York 2003)

1.33 S. French: The structure of theories. In: *The Routledge Companion to Philosophy of Science*, 2nd edn., ed. by M. Curd, S. Psillos (Routledge, London 2013) pp. 301–312

1.34 C. Pincock: Overextending partial structures: Idealization and abstraction, Philos. Sci. **72**(4), 1248–1259 (2005)

1.35 M. Suárez, N. Cartwright: Theories: Tools versus models, Stud. Hist. Philos. Mod. Phys. **39**(1), 62–81 (2008)

1.36 P. Godfrey-Smith: The strategy of model-based science, Biol. Philos. **21**(5), 725–740 (2006)

1.37 M. Thomson-Jones: Missing systems and the face value practice, Synthese **172**(2), 283–299 (2010)

1.38 P.N. Johnson-Laird: *Mental Models* (Harvard Univ. Press, Cambridge 1983)

1.39 N. Nersessian: Model-based reasoning in conceptual change. In: *Model-Based Reasoning in Scientific Discovery*, ed. by L. Magnani, N. Nersessian, P. Thagard (Plenum, New York 1999) pp. 5–22

1.40 J.C. Maxwell: *The Scientific Papers of James Clerk Maxwell*, Vol. 1 (Cambridge Univ. Press, Cambridge 1890), ed. by W.D. Niven

1.41 R. Frigg: Models and fiction, Synthese **172**(2), 251–268 (2010)

1.42 A.C. Doyle: The Adventure of the Three Garridebs. In: *The Casebook of Sherlock Holmes*, ed. by J. Miller (Dover, 2005)

1.43 A. Meinong: *Untersuchungen zur Gegenstandstheorie und Psychologie* (Barth, Leipzig 1904)

1.44 K. Walton: *Mimesis as Make-Believe: On the Foundations of the Representational Arts* (Harvard Univ. Press, Cambridge 1990)

1.45 A. Toon: *Models as Make-Believe: Imagination, Fiction, and Scientific Representation* (Palgrave-Macmillan, Basingstoke 2012)

1.46 A. Toon: The ontology of theoretical modelling: Models as make-believe, Synthese **172**(2), 301–315 (2010)

1.47 H.G. Wells: *War of the Worlds* (Penguin, London 1897), 1978

1.48 R. Giere: How models are used to represent reality, Proc. Philosophy of Science, Vol. 71 (2004) pp. S742–S752

1.49 M. Morrison, M. Morgan (Eds.): *Models as Mediators* (Cambridge Univ. Press, Cambridge 1999)

1.50 M. Morrison, M. Morgan: Models as mediating instruments. In: *Models as Mediators*, ed. by M. Morrison, M. Morgan (Cambridge Univ. Press, Cambridge 1999) pp. 10–37

1.51 T. Knuuttila: *Models as Epistemic Artefacts: Toward a Non-Representationalist Account of Scientific Representation* (Univ. Helsinki, Helsinki 2005)

1.52 T. Knuuttila: Modelling and representing: An artefactual approach to model-based representation, Stud. Hist. Philos. Sci. **42**(2), 262–271 (2011)

1.53 A. Gelfert: *How to Do Science with Models: A Philosophical Primer* (Springer, Cham 2016)

2. Models and Theories

Demetris Portides

Both the received view (RV) and the semantic view (SV) of scientific theories are explained. The arguments against the RV are outlined in an effort to highlight how focusing on the syntactic character of theories led to the difficulty in characterizing theoretical terms, and thus to the difficulty in explicating how theories relate to experiment. The absence of the representational function of models in the picture drawn by the RV becomes evident; and one does not fail to see that the SV is in part a reaction to – what its adherents consider to be an – excessive focus on syntax by its predecessor and in part a reaction to the complete absence of models from its predecessor's philosophical attempt to explain the theory–experiment relation. The SV is explained in an effort to clarify its main features but also to elucidate the differences between its different versions. Finally, two kinds of criticism are explained that affect all versions of the SV but which do not affect the view that models have a warranted degree of importance in scientific theorizing.

Scientists use the term *model* with reference to iconic or scaled representations, analogies, and mathematical (or abstract) descriptions. Although all kinds of models in science may be philosophically interesting, mathematical models stand out. Representation with iconic or scale models, for instance, mostly applies to a particular state of a system at a particular time, or it requires the mediation of a mathematical (or abstract) model in order to relate to theories. Representation via mathematical models, on the other hand, is of utmost interest because it applies to *types* of target systems and it can be used to draw inferences about the time-evolution of such systems, but more importantly for our purposes because of its obvious link to scientific theories.

In the history of philosophy of science, there have been two systematic attempts to explicate the relation of such models to theory. The first is what had been labeled the *received view* (RV) of scientific theories that grew out of the logical positivist tradition. According to this view, theories are construed as formal axiomatic calculi whose logical consequences extend to observational sentences. Models are thought to have no representational role; their role is understood metamathematically, as interpretative structures of subsets of sentences of the formal calculus. Ultimately it became clear that such a role ascribed to models does not do justice to how science achieves theoretical representations of phenomena. This conclusion was reached largely due to the advent of the second systematic attempt to explore the relation between theory and models, the *semantic view* (SV) or model-theoretic view of scientific theories. The semantic view regards theories as classes of models that are directly defined without resort to a formal calculus. Thus, models in this view are integral parts of theories, but they are also the devices by which representation of phenomena is achieved.

Although, the SV recognized the representational capacity of models and exposed that which was concealed by the logical positivist tradition, namely that one of the primary functions of scientific models is to apply the abstract theoretical principles in ways that actual physical systems can be represented, it also generated a debate concerning the complexities involved in scientific representation. This recent debate has significantly enhanced our understanding of the representational role of scientific models. At the same time it gave rise, among other things, to questions regarding the relation between models and theory. The adherents of the SV claim that a scientific theory is identified with a class of models, hence that models are constitutive parts of theory and thus they represent by means of the conceptual apparatus of theory. The critics of the SV, however, argue that those models that are successful representations of physical systems utilize a much richer conceptual apparatus than that provided by theory and thus claim that they should be understood as partially autonomous from theory.

A distinguishing characteristic of this debate is the notion of representational model, that is, a scientific entity which possesses the necessary features that render it representational of a physical system. In the SV, theoretical models, that is, mathematical models that are constitutive parts of theory structure, are considered to be representational of physical systems. Its critics, however, argue that in order to provide a model with the capacity to represent actual physical systems, the theoretical principles from which the model arises are typically supplemented with ingredients that derive from background knowledge, from semiempirical results and from experiment. In order to better understand the character of successful representational models, according to this latter view, we must move away from a purely theory-driven view of model construction and also place our emphasis on the idea that representational models are entities that consist of assortments of the aforementioned sorts of conceptual ingredients.

In order to attain insight into how models could relate to theory and also be able to use that insight in addressing other issues regarding models, in what follows, I focus on the RV and the SV of scientific theories. Each of the two led to a different conception of the nature of theory structure and subsequently to a different suggestion for what scientific models are, what they are used for, and how they function. In the process of explicating these two conceptions of theory structure, I will also review the main arguments that have been proposed against them. The RV has long been abandoned for reasons that I shall explore in Sect. 2.1, but the SV lives on despite certain inadequacies that I shall also explore in Sect. 2.2. Toward the end of Sect. 2.2, in Sect. 2.2.4, I shall very briefly touch upon the more recent view that the relation between theory and models is far more complex than advocates of the RV or the SV have claimed, and that models in science demonstrate a certain degree of partial autonomy from the theory that prompts their construction and because of this a unitary account of models obscures significant features of scientific modeling practices.

2.1 The Received View of Scientific Theories

What has come to be called the RV of scientific theories is a conception of the structure of scientific theories that is associated with logical positivism, and which was the predominant view for a large part of the twentieth century. It is nowadays by and large overlooked hence it is anything but received. Despite its inappropriate label, clarifying its major features as well as understanding the major philosophical arguments that revealed its inadequacies would not only facilitate acquaintance with the historical background of the debate about the structure of scientific theories and give the reader a flavor of the difficulties involved in characterizing theory structure, but it would also be helpful in understanding some characteristics of contemporary views and how models came to occupy central stage in current debates on how theories represent and explain phenomena. With this intention in mind, I proceed in this section by briefly explaining the major features of the RV and continue with sketching the arguments that exposed its weaknesses in Sects. 2.1.1–2.1.6.

The RV construes scientific theories as Hilbert-style formal axiomatic calculi, that is, axiomatized sets of sentences in first-order predicate calculus with identity. A scientific theory is thus identified with a formal language, L, having the following features. The nonlogical terms of L are divided into two disjoint classes: (1) the theoretical terms that constitute the theoretical vocabulary, V_{T}, of L and (2) the observation terms that constitute the observation vocabulary, V_{O}, of L. Thus, L can be thought of as consisting of an observation language, L_{O}, that is, a language that consists only of observation terms, a theoretical language, L_{T}, that is, a language that consists only of theoretical terms, and a part that consists of mixed sentences that are made up of both observation and theoretical terms. The theoretical postulates or the axioms of the theory, T (i. e., what we, commonly, refer to as the high-level scientific laws), consist only of terms from V_T. This construal of theories is a syntactic system, which naturally requires semantics in order to be useful as a model of scientific theories.

It is further assumed that the terms of V_O refer to directly observable physical objects and directly observable properties and relations of physical objects. Thus the semantic interpretation of such terms, and the sentences belonging to L_O, is provided by direct observation. The terms of V_T, and subsequently all the other sentences of L not belonging to L_O, are partially interpreted via the theoretical postulates, T, and – a finite set of postulates that has come to be known as – the *correspondence rules*, C. The latter are mixed sentences of L, that is, they are constructed with at least one term from each of the two classes V_T and V_O. (The reader could consult *Suppe* [2.1] for a detailed exposition of the RV, but also for a detailed philosophical study of the developments that the RV underwent under the weight of several criticisms until it reached, what Suppe calls, the "final version of the RV").

We could synopsize how scientific theories are conceived according to the RV as follows: The scientific laws, which as noted constitute the axioms of the theory, specify relations holding between the theoretical terms. Via a set of *correspondence rules*, theoretical terms are reduced to, or defined by, observation terms. Observation terms refer to objects and relations of the physical world and thus are interpreted. Hence, a scientific theory, according to the RV, is a formal axiomatic system having as point of departure a set of theoretical postulates, which when augmented with a set of correspondence rules has deductive consequences that stretch all the way to terms, and sentences consisting of such terms, that refer to the directly observable physical objects. Since according to this view, the backbone of a scientific theory is the set of theoretical postulates, T, and a partial interpretation of L is given via the set of correspondence rules, C, let TC (i. e., the union set of T and C) designate the scientific theory.

From this sketch, it can be inferred that the RV implies several philosophically interesting things. For the purposes of this chapter, it suffices to limit the discussion only to those implications of the RV that are relevant to the criticisms that have contributed to its downfall. These implications, which in one way or another relate to the difficulty in characterizing V_T terms, are:

1. It relies on an observational–theoretical distinction of the terms of L.
2. It embodies an analytic–synthetic distinction of the sentences of L.
3. It employs the obscure notion of correspondence rules to account for the interpretation of theoretical terms and to account for theory application.
4. It does not assign a representational function to models.
5. It assigns a deductive status to the relation between empirical theories and experiment.
6. It commits to a theory consistency condition and to a meaning invariance condition.

Part A | 2.1

2.1.1 The Observation–Theory Distinction

The separation of L into V_O and V_T terms implies that the RV requires an observational–theoretical distinction in the terms of the vocabulary of the theory. This idea was criticized in two ways. The first kind of objection to the observation–theory distinction relied on a twofold argument. On the one hand, the critics claim that an observation–theory distinction of scientific terms cannot be drawn; and on the other, that a classification of terms following such a distinction would give rise to a distinction of observational–theoretical statements, which also cannot be drawn for scientific languages. The second kind of objection to the distinction relies on attempts to establish accounts of *observation* that are incompatible with the observation–theory distinction and on showing that observation statements are theory laden.

The Untenability of the Observation–Theory Distinction

The argument of the first kind that focuses on the untenability of the observation–theory distinction is due to *Achinstein* [2.2, 3] and *Putnam* [2.4]. Achinstein explores the sense of observation relevant to science, that is, "the sense in which observing involves visually attending to something," and he claims that this sense exhibits the following characteristics:

1. Observation involves attention to the various aspects or features of an item depending on the observer's concerns and knowledge.
2. Observation does not necessarily involve recognition of the item.
3. Observation does not imply that whatever is observed is in the visual field or in the line of sight of the observer.
4. Observation could be achieved indirectly.
5. The description of what one observes can be done in different ways (The reader could refer to *Achinstein* [2.3, pp. 160–165] for an explication of these characteristics of observation by the use of specific examples).

If now one urges an observation–theory distinction by simply constructing lists of observable and unobservable terms (as proponents of the RV according to Achinstein do), the distinction becomes untenable. For example, according to typical lists of unobservables, *electron* is a theoretical term. But based on points (3) and (4) above, Achinstein claims, this could be rejected. Similarly based on point (5), Achinstein also rejects the tenability of such a distinction at the level of statements, because "what scientists as well as others observe is describable in many different ways, using terms from both vocabularies" [2.3, p. 165].

Furthermore, if, as proponents of the RV have often claimed, (For instance, *Hempel* [2.5], *Carnap* [2.6] and [2.7]), items in the observational list are directly observable whereas those in the theoretical list are not, then *Achinstein* [2.3, pp. 172–177] claims that a close construal of *directly observable* reveals that the desired classification of terms into the two lists fails. He explains that *directly observable* could mean that it can be observed without the use of instruments. If this is what advocates of the RV require, then it does not warrant the distinction. First, it is not precise enough to classify things seen by images and reflections. Second, if something is not observable without instruments means that no aspect of it is observable without instruments then things like temperature and mass would be observables, since some aspects of them are detected without instruments. If however directly observable means that no instruments are required to detect its presence, then it would be insufficient because one cannot talk about the presence of temperature. Finally, if it means that no instruments are required to measure it or its properties, then such terms as volume, weight, etc. would have to be classified as theoretical terms. Hence, Achinstein concludes that the notion of direct observability is unclear and thus fails to draw the desired observation–theory distinction.

Along similar lines, *Putnam* [2.4] argues that the distinction is completely *broken-backed* mainly for three reasons. First, if an observation term is one that only refers to observables then there are no observation terms. For example, the term *red* is in the observable class but it was used by Newton to refer to a theoretical term, namely red corpuscles. Second, many terms that refer primarily to the class of unobservables are not theoretical terms. Third, some theoretical terms, that are of course the outcome of a scientific theory, refer primarily to observables. For example, the theory of evolution, as put forward by Darwin, referred to observables by employing theoretical terms.

What these arguments accomplish is to highlight the fact that scientific languages employ terms that cannot clearly and easily be classified into observational or theoretical. They do not however show the untenability of the observation–theory distinction as employed by the RV. As *Suppe* [2.8] correctly observes, what they show is that the RV needs a sufficiently rich artificial language for science, no matter how complex it may turn out to be. Such a language, in which presumably the observation–theory distinction is tenable, must have a plethora of terms, such that, to use his example, the designated term red_o will refer to the observable occurrences of the predicate red, and the designated term red_t will refer to the unobservable occurrences.

The Theory-Ladenness of Observation

Hanson's argument is a good example of the second kind, in which an attempt is made to show that there is no theory-neutral observation language and that observation is theory-laden and thus establish an account of *observation* that is incompatible with the observation–theory distinction required by the RV (*Hanson* [2.9, pp. 4–30]. *Hanson* [2.10, pp. 59–198]. Also *Suppe* [2.1, pp. 151–166]). He does this by attempting to establish that an observation language that intersubjectively can be given a theory-independent semantic interpretation, as the RV purports, cannot exist.

He begins by asking whether two people see the same things when holding different theories. We could follow his argument by reference to asking whether Kepler and Tycho Brahe see the same thing when looking at the sun rising. Kepler, of course, holds that the earth revolves around the sun, while Tycho holds that the sun revolves around the earth. Hanson addresses this question by considering ambiguous figures, that is, figures that sometimes can be seen as one thing and other times as another. The most familiar example of this kind is the duck–rabbit figure.

When confronted with such figures, viewers see either a duck or a rabbit depending on the perspective they take, but in both cases they see the same distal object (i. e., the object that emits the rays of light that impinge the retina). Hanson uses this fact to develop a sequence of arguments to counter the standard interpretations of his time. There were two standard interpretations at the time. The first was that the perceptual system delivers the same visual representation and then cognition (thought) interprets this either as a duck or as a rabbit. The other was that the perceptual system outputs both representations and then cognition chooses one of the two. Both interpretations are strongly linked with the idea that the perceptual process and the cognitive process function independently of one another, that is, the perceptual system delivers its output independent of any cognitive influences. However, Hanson challenges the assumption that the two observers see the same thing and via thought they interpret it differently. He claims that perception does not deliver either a duck or a rabbit, or an ambiguous figure, and then via some other independent process thought chooses one or the other. On the contrary, the switch from seeing one thing to seeing the other seems to take place spontaneously and moreover a process of back and forth seeing without any thinking seems to be involved. He goes on to ask, what could account for the difference in what is seen? His answer is that what changes is the organization of the ambiguous figure as a result of the conceptual background of each viewer. This entails that what one sees, the percept, depends on the conceptual background that results from one's experience and knowledge, which means that thought affects the formation of the percept; thus perception and cognition become intertwined. When Tycho and Kepler look at the sun, they are confronted with the same distal object but they see different things because their conceptual organizations of their experiences are vastly different. In other words, Hanson's view is that the percept depends on background knowledge, which means that cognition influences perceptual processing. Consequently, observation is theory laden, namely, observation is conditional on background knowledge.

By this argument, Hanson undermines the RV's position, which entails that Kepler and Brahe see the same thing but interpret it differently; and also establishes that conceptual organizations are features of *seeing* that are indispensable to scientific observation and thus that Kepler and Brahe see two different things because perception inherently involves interpretation, since the former is conditional on background knowledge. It is, however, questionable whether Hanson's arguments are conclusive. *Fodor* [2.11–13], *Pylyshyn* [2.14, 15], and *Raftopoulos* [2.16–18], for example, have extensively argued on empirical grounds that perception, or at least a part of it, is theory independent and have proposed explanations of the ambiguous figures that do not invoke cognitive effects in explaining the percept and the switch between the two interpretations of the figure. This debate, therefore, has not yet reached its conclusion; and many today would argue that fifty or so years after Hanson the arguments against the theory ladenness of observation are much more tenable.

2.1.2 The Analytic–Synthetic Distinction

The RV's dependence on the observation–theory distinction is intimately linked to the requirement for an analytic–synthetic distinction. An argument to defend this claim is given by *Suppe* [2.1, pp. 68–80]. Here is a sketch of that argument. The analytic–synthetic distinction is embodied in the RV, because (as suggested by *Carnap* [2.19]) implicit in *TC* are meaning postulates (or semantical rules) that specify the meanings of sentences in L. However, if meaning specification were the only function of *TC* then *TC* would be analytic, and in such case it would not be subject to empirical investigation. *TC* must therefore have a factual component, and the meaning postulates must separate the meaning from the factual component. This would imply an analytic–synthetic separation, if those sentences in L that are logical truths or logical consequences of the meaning postulates are analytic and all nonanalytic sentences are understood to be synthetic. Moreover, any nonanalytic sentence in L taken in conjunction with the class of meaning postulates would have certain empirical consequences. If the conjunction is refuted or confirmed by directly observable evidence, this will reflect only on the truth value of the conjunction and not on the meaning postulates. Hence such conjunctive sentences can only be synthetic. Thus every nonanalytic sentence of L_O and every sentence of L constituted by a mixed vocabulary is synthetic. So the observation–theory distinction supports an analytic–synthetic distinction of the sentences of L.

The main criticism against the analytic–synthetic distinction consists of attempts to show its untenability. *Quine* [2.20] points out that there are two kinds of analytic statements: (a) logical truths, which remain true under all interpretations, and (b) statements that are true by virtue of the meaning of their nonlogical terms, for example, *No bachelor is married*. He then argues that the analyticity of statements of the second kind cannot be established without resort to the notion of synonymy, and that the latter notion is just as problematic as the notion of analyticity. The argument runs roughly as follows. Given that meaning (or intension) is clearly distinguished from its extension, that is, the class of entities to which it refers, a theory of meaning is primarily concerned with cognitive synonymy (i. e., the synonymy of linguistic forms). For example, to say that *bachelor* and *unmarried man* are cognitively synonymous is to say that they are interchangeable in all contexts without change of truth value. If such were the case then the statement *No bachelor is married* would become *No unmarried man is married*, which would be a logical truth. In other words, statements of kind (b) are reduced to statements of kind (a) if only we could interchange synonyms for synonyms. But as Quine argues, the notion of interchangeability salva veritate is an extensional concept and hence does not help with analyticity. In fact, no analysis of the interchangeability salva veritate account of synonymy is possible without recourse to analyticity, thus making such an effort circular, unless interchangeability is "[...] relativized to a language whose extent is specified in relevant respects" [2.20, p. 30]. That is to say,

Part A | 2.1

we first need to know what statements are analytic in order to decide which expressions are synonymous; hence appeal to synonymy does not help with the notion of analyticity.

Similarly *White* [2.21] argues that an artificial language, L_1, can be constructed with appropriate definitional rules, in which the predicates P_1 and Q_1 are synonymous whereas P_1 and Q_2 are not; hence making such sentences as $\forall x\,(P_1(x) \rightarrow Q_1(x))$ logical truths and such sentences as $\forall x\,(P_1(x) \rightarrow Q_2(x))$ synthetic. In a different artificial language L_2, P_1 could be defined to be synonymous to Q_2 and not to Q_1, hence making the sentence $\forall x\,(P_1(x) \rightarrow Q_2(x))$ a logical truth and the sentence $\forall x\,(P_1(x) \rightarrow Q_1(x))$ synthetic. This relies merely upon convention. However, he asks, in a natural language what rules are there that dictate what choice of synonymy can be made such that one formula is a synthetic truth rather than analytic? The key point of the argument is therefore that in a natural language or in a scientific language, which are not artificially constructed and which do not contain definitional rules, the notion of analyticity is unclear.

Nevertheless, it could be argued that such arguments as the above are not entirely conclusive, primarily because the RV is not intended as a description of actual scientific theories. Rather, the RV is offered as a *rational reconstruction* of scientific theories, that is, an explication of the structure of scientific theories. It does not aim to describe how actual theories are formulated, but only to indicate a logical framework (i. e., a canonical linguistic formulation) in which theories can be essentially reformulated. Therefore, all that proponents of the RV, needed to show was that the analytic–synthetic distinction is tenable in some artificial language (with meaning postulates) in which scientific theories could potentially be reformulated. In view of this, in order for the RV to overcome the obscurity of the notion of analyticity, pointed out by Quine and White, it would require the conclusion of a project that Carnap begun: To spell out a clear way by which to characterize meaning postulates for a specified theoretical language (This is clearly *Carnap*'s intention in his [2.19]).

2.1.3 Correspondence Rules

In order to distinguish the character and function of theoretical terms from speculative metaphysical ones (e.g., *unicorn*), logical positivists sought for a connection of theoretical to observational terms by giving an analysis of the empirical nature of theoretical terms contrary to that of metaphysical terms. This connection was formulated in what we can call, following *Achinstein* [2.22], the *Thesis of Partial Interpretation*, which is basically the following: As indicated above, in the brief sketch of the main features of the RV, the RV allows that a complete empirical semantic interpretation in terms of directly observables is given to V_{O} terms and to sentences that belong to L_{O}. However, no such interpretation is intended for V_{T} terms and consequently for sentences of L containing them. It is TC as a whole that supplies the empirical content of V_{T} terms. Such terms receive a partial observational meaning indirectly by being related to sets of observation terms via correspondence rules. To use one of Achinstein's examples [2.22, p. 90]:

> "it is in virtue of [a correspondence-rule] which connects a sentence containing the theoretical term *electron* to a sentence containing the observational term *spectral line* that the former theoretical term gains empirical meaning within the Bohr theory of the atom"

Correspondence rules were initially introduced to serve three functions in the RV:

1. To define theoretical terms.
2. To guarantee the cognitive significance of theoretical terms.
3. To specify the empirical procedures for applying theory to phenomena.

In the initial stages of logical positivism it was assumed that if observational terms were cognitively significant, then theoretical terms were cognitively significant if and only if they were explicitly defined in terms of observational terms. The criteria of explicit definition and cognitive significance were abandoned once proponents of the RV became convinced that dispositional terms, which are cognitively significant, do not admit of explicit definitions (*Carnap* [2.23, 24], also *Hempel* [2.25, pp. 23–29], and *Hempel* [2.5]). Consider, for example, the dispositional term *tearable* (let us assume all the necessary conditions for an object to be torn apart hold), if we try to explicitly define it in terms of observables we end up with something like this:

> "An object x is tearable if and only if, if it is pulled sharply apart at time t then it will tear at t (assuming for simplicity that pulling and tearing occur simultaneously)."

The above definition could be rendered as $\forall x\,(T(x) \leftrightarrow \forall t(P(x,t) \rightarrow Q(x,t)))$, where, T is the theoretical term *tearable*, P is the observational term *pulled apart*, and Q is the observational term *tears*. But this does not correctly define the actual dispositional property *tearable*, because the right-hand side of the biconditional will be true of objects that are never pulled apart. As a result, some objects that are not tearable and have never being pulled apart will by definition have the property *tearable*.

Because of this, *Carnap* [2.23, 24] proposed to replace the construal of correspondence rules as explicit definitions, by *reduction sentences* that partially determine the observational content of theoretical terms. A reduction sentence defined the dispositional property tearable as follows: $\forall x \forall t\ (P(x,t) \rightarrow (Q(x,t) \leftrightarrow T(x)))$. That is, (*Carnap* calls such sentences *bilateral reduction sentences* [2.23, 24]):

> "If an object x is pulled-apart at time t, then it tears at time t if and only if it is tearable."

Unlike the explicit definition case, if a is a non-tearable object that is never pulled apart then it is not implied that $T(a)$ is true. What will be implied, in such case, is that $\forall t\ (P(a,t) \rightarrow (Q(a,t) \leftrightarrow T(a)))$, is true. Thus the above shortcoming of explicit definitions is avoided, because a reduction sentence does not completely define a disposition term. In fact, this is also the reason why correspondence rules supply only partial observational content, since many other reduction sentences can be used to supply other empirical aspects of the term *tearable*, for example, being torn by excessively strong shaking. Consequently, although correspondence rules were initially meant to provide explicit definitions and cognitive significance to V_{T} terms, these functions were abandoned and substituted by *reduction sentences* and *partial interpretation* (A detailed explication of the changes in the use of correspondence rules through the development of the RV can be found in [2.1]).

Therefore, in its most defensible version the RV could be construed to assign the following functions to correspondence rules: First, they specify empirical procedures for the application of theory to phenomena and second, as a constitutive part of TC, they supply V_{T} and L_{T} with partial interpretation. Partial interpretation in the above sense is all the RV needs since, given its goal of distinguishing theoretical from speculative metaphysical terms, it only needs a way to link the V_{T} terms to the V_{O} terms. The version of the RV that employs correspondence rules for these two purposes motivated two sorts of criticisms. The first concerns the idea that correspondence rules provide partial interpretation to V_{T} terms, and the second concerns the function of correspondence rules for providing theory application.

The thesis of partial interpretation came under attack from *Putnam* [2.4] and *Achinstein* [2.3, 22]. The structure of their arguments is similar. They both think that partial interpretation is unclear and they attempt to clarify the concept. They do so by suggesting plausible explications for *partial interpretation*. Then they show that for each plausible explication that each of them suggests partial interpretation is either an incoherent notion or inadequate for the needs of the RV. Thus, they both conclude that any attempt to elucidate the notion of partial interpretation is problematic and that partial interpretation of V_{T} terms cannot be adequately explicated. For example, Putnam gives the following plausible explications for *partial interpretation*:

1. To partially interpret V_{T} terms is to specify a class of intended models.
2. To partially interpret a term is to specify a verification–refutation procedure that applies only to a proper subset of the extension of the term.
3. To partially interpret a formal language L is to interpret only part of the language.

In similar spirit, Achinstein gives three other plausible explications. One of Putnam's counterexamples is that (1) above cannot meet its purpose because the class of intended models, that is, the semantic structures or interpretations that satisfy TC and which are so intended by scientists, is not well defined (A critical assessment of these arguments can be found in [2.1]).

The other function of correspondence rules, that of specifying empirical procedures for theory application to phenomena, also came under criticism. *Suppe* [2.1, pp. 102–109] argued that the account of correspondence rules inherent in the RV is inadequate for understanding actual science on the following three grounds:

1. They are mistakenly viewed as components of the theory rather than as auxiliary hypotheses.
2. The sorts of connections (e.g., explanatory causal chains) that hold between theories and phenomena are inadequately captured.
3. They oversimplify the ways in which theories are applied to phenomena.

The first argument is that the RV considers TC as postulates of the theory. Hence C is assumed to be an integral part of the theory. But, if a new experimental procedure is discovered it would have to be incorporated into C, and the result would be a new set of rules C' that subsequently leads to a new theory TC'. But obviously the theory does not undergo any change. When new experimental procedures are discovered we only improve our knowledge of how to apply theory to phenomena. So we must think of correspondence rules as auxiliary hypotheses distinct from theory.

The second argument is based upon *Schaffner*'s [2.26] observation that there is a way in which theories are applied to phenomena, which is not captured by the RV's account of correspondence rules. This is the case when various auxiliary theories (independent of T) are used to describe a *causal sequence*, which obtains between the states described by T and the observation reports. These causal sequences are descriptions of the mechanisms involved within physical systems to cause the measurement apparatus to behave as it does. Thus, they supplement theoretical explanations of the observed behavior of the apparatus by linking the theory to the observation reports via a causal story. For example, such auxiliary hypotheses are used to establish a causal link between the motion of an electron (V_T term) and the spectral line (V_O term) in a spectrometer photograph. Schaffner's point is that the relation between theory and observation reports is frequently achieved by the use of these auxiliary hypotheses that establish explanations of the behavior of physical systems via causal mechanisms. Without recognizing the use of these auxiliaries the RV may only describe a type of theory application whereby theoretical states are just correlated to observational states. If these kinds of auxiliaries were to be viewed as part of C then it is best that C is dissociated from the core theory and is regarded as a separate set of auxiliary hypotheses required for establishing the relation between theory and experiment, because such auxiliaries are obviously not theory driven, but if they are not to be considered part of C then C does not adequately explain the theory–experiment relation.

Finally, the third argument is based on *Suppes*' [2.27, 28] analysis of the complications involved in relating theoretical predictions to observation reports. Suppes observes that in order to reach the point where the two can meaningfully be compared, several epistemologically important modifications must take place on the side of the observation report. For example, Suppes claims, on the side of theory we typically have predictions derived from continuous functions, and on the side of an observation report we have a set of discrete data. The two can only be compared after the observation report is modified accordingly. Similarly, the theory's predictions may be based on the assumption that certain idealizing conditions hold, for example, no friction. Assuming that in the actual experiment these conditions did not hold, it would mean that to achieve a reasonable comparison between theory and experiment the observational data will have to be converted into a corresponding set that reflects the result of an *ideal* experiment. In other words, the actual observational data must be converted into what they would have been had the idealizing conditions obtained. According to Suppes, these sorts of conversion are obtained by employing appropriate *theories of data*. So, frequently, there will not be a direct comparison between theory and observation, but a comparison between theory and observation-altered-by-theory-of-data.

By further developing Suppes' analysis, *Suppe* [2.8] argues that because of its reliance on the observation–theory distinction, the RV employs correspondence rules in such a way as to blend together unrelated aspects of the scientific enterprise. Such aspects are the design of experiments, the interpretation of theories, the various calibration procedures, the employment of results and procedures of related branches of science, etc. All these unrelated aspects are compounded into the correspondence rules. Contrary to the implications of the RV, Suppe claims, in applying a theory to phenomena we do not have any direct link between theoretical terms and observational terms. In a scientific experiment we collect data about the phenomena, and often enough the process of collecting the data involves rather sophisticated bodies of theory. Experimental design and control, instrumentation, and reliability checks are necessary for the collection of data. Moreover, sometimes generally accepted laws or theories are also employed in collecting these data. All these features of experimentation and data collection are then employed in ways as to structure the data into forms (which Suppe calls, *hard data*) that allow meaningful comparison to theoretical predictions. In fact, theory application according to Suppe involves contrasting theoretical predictions to *hard data*, and not to something directly observed [2.8, p. 11]:

> "Accordingly, the correspondence rules for a theory should not correlate direct-observation statements with theoretical statements, but rather should correlate *hard data* with theoretical statements."

In a nutshell, although both Suppes' and Suppe's arguments do not establish with clarity how the theory–experiment relation is achieved they do make the following point: Actual scientific practice, and in particular theory–application, is far more complex than the description given by the RV's account of correspondence rules.

2.1.4 The Cosmetic Role of Models According to the RV

The objection that the RV obscures several epistemologically important features of scientific theories is implicitly present in all versions of the SV of theories. Suppe, however, brings this out explicitly in the form of a criticism (*Suppe* [2.1, 29, 30]). To clarify the sort of criticism presented by Suppe, we need to make use of some elements of the alternative picture of scientific theories given by the SV, which we shall explore in detail in Sect. 2.2.

The reasoning behind Suppe's argument is the following. Science, he claims, has managed so far to go about its business without involving the observation–theory distinction and all the complexities that it gives rise to. Since, he suggests, the distinction is not required by science, it is important to ask not only whether an analysis of scientific theories that employs the distinction is adequate or not, that is, the issue on which (as we have seen so far) many of the criticisms of the RV have focused, but whether or not the observation–theory distinction which leads to the notion of correspondence rules subsequently steers toward obscuring epistemological aspects of scientific theorizing.

The sciences, he argues, do not deal with all the detailed features of phenomena and not with phenomena in all their complexity. Rather they isolate a certain number of physical parameters by abstraction and idealization and use these parameters to characterize *physical systems* (Suppe's terminology is idiosyncratic, he uses the term *physical system* to refer to the abstract entity that an idealized model of the theory represents and not to the actual target physical system), which are highly abstract and idealized replicas of phenomena. A classical mechanical description of the earth–sun system of our solar system, would not deal with the actual system, but with a physical system in which some relevant parameters are abstracted (e.g., mass, displacement, velocity) from the complex features of the actual system. And in which some other parameters are ignored, for example, the intensity of illumination by the sun, the presence of electromagnetic fields, the presence of organic life. In addition, these abstracted parameters are not used in their full complexity to characterize the physical system. Indeed, the description would idealize the physical system by ignoring certain factors or features of the actual system that may plausibly be causally relevant to the actual system. For instance, it may assume that the planets are point masses, or that their gravitational fields are uniform, or that there are no disturbances to the system by external factors and that the system is in a vacuum. What scientific theories do is attempt to characterize the behavior of such physical systems not the behavior of directly observable phenomena.

Although this is admittedly a rough sketch of Suppe's view, it is not hard to see that the aim of the argument is to lead to the conclusion that the directly observable phenomena are connected to a scientific theory via the physical system. That is to say, (if we put together this idea with the one presented at the end of Sect. 2.1.3 above) the connection between the theory and the phenomena, according to Suppe, requires an analysis of theories and of theory–application that involves a two-stage move. The first move involves the connection between raw phenomena and the *hard data* about the particular target system in question. The second move involves the connection between the *physical system* that represents the *hard data* and the theoretical postulates of the theory. According to Suppe's understanding of the theory–experiment relation, the physical system plays the intermediate role between phenomena and theory and this role, which is operative in theory–application, is what needs to be illuminated. The RV implies that the correspondence rules "[...] amalgamate together the two sorts of moves [...] so as to eliminate the physical system" [2.29, p. 16], thus obscuring this important epistemological feature of scientific theorizing.

So, according to Suppe, correspondence rules must give way to this two-stage move, if we are to identify and elucidate the epistemic features of *physical systems*. Suppe's suggestion is that the only way to accommodate physical systems into our understanding of how theories relate to phenomena is to give models of the theory their representational status. The representational means of the RV are linguistic entities, for example, sentences. Models, within the RV, are denied any representational function. They are conceived exclusively as interpretative devices of the formal calculus, that is, as structures that satisfy subsets of sentences of the theory. This reduces models to metamathematical entities that are employed in order to make intelligible the abstract calculus, which amounts to treating them as more or less *cosmetic* aspects of science. But this understanding of the role of models leads to the incapacity of the RV to elucidate the epistemic features of physical systems, and thus obscures – what Suppe considers to be – epistemologically important features of scientific theorizing.

2.1.5 Hempel's Provisos Argument

In one of his last writings, *Hempel* [2.31] raises a problem that suggests a flaw in interpreting the link between empirical theories and experimental reports as mere deduction. Assuming that a theory is a formal axiomatic system consisting of T and C, as we did so far, consider Hempel's example. If we try to apply the theory of magnetism for a simple case we are faced with the following inferential situation. From the observational sentence *b is a metal bar to which iron filings are clinging* (S_{O1}), by means of a suitable correspondence rule we infer the theoretical sentence *b is a magnet* (S_{T1}). Then by using the theoretical postulates in T, we infer *if b is broken into two bars, then both are magnets and their poles will attract or repel each other* (S_{T2}). Finally using further correspondence rules we derive the observational sentence *if b is broken into two shorter bars and these are suspended, by long thin threads, close to each other at the same distance from the ground, they will orient themselves so as to fall into a straight line* (S_{O2}) ([2.31, p. 20]). If the inferential structure is assumed to be deductive then the above structure can be read as follows: S_{O1} in combination with the theory deductively implies S_{O2}. Hempel concludes that this deductivist construal faces a difficulty, which he calls *the problem of provisos*.

To clarify the problem of provisos, we must look into the third inferential step from S_{T2} to S_{O2}. What is necessary here is for the theory of magnetism to provide correspondence rules that would turn this step into a deductive inference. The theory however, as Hempel points out, clearly does not do this. In fact, the theory allows for the possibility that the magnets orient themselves in a way other than a straight line, for example, if an external magnetic field of suitable strength and direction is present. This leads to recognizing that the third inferential step presupposes the additional assumption that there are no disturbing influences to the system of concern. *Hempel* uses the term *provisos*, "[...] to refer to assumptions [of this kind] [...], which are essential, but generally unstated, presuppositions of theoretical inferences" [2.31, p. 23]. Therefore, provisos are presupposed in the application of a theory to phenomena (The problem we saw in Sect. 2.1.3 which Suppes raises, namely that in science theoretical predictions are not confronted with raw observation reports but with observation-altered-by-theory-of-data reports, neighbors this problem but it is not the same. Hempel's problem of provisos concerns whether it is possible to deductively link theory to observational statements no matter how the latter are constructed).

What is the character of provisos? Hempel suggests we may view provisos as *assumptions of completeness*. For example, in a theoretical inference from a sentence S_1 to another S_2, a proviso is required that asserts that in a given case "[...] no factors other than those specified in S_1 are present that could affect the event described by S_2" [2.31, p. 29]. As, for example, is the case in the application of the Newtonian theory to a two-body system, where it is presupposed that their mutual gravitational attraction are the only forces the system is subjected to. It is clear that [2.31, p. 26]:

> "[...] a proviso as here understood is not a clause that can be attached to a theory as a whole and vouchsafe its deductive potency by asserting that in all particular situations to which the theory is applied, disturbing factors are absent. Rather, a proviso has to be conceived as a clause that pertains to some particular application of a given theory and asserts that in the case at hand, no effective factors are present other than those explicitly taken into account."

Thus, if a theory is conceived as a deductively closed set of statements and its axioms conceived as empirical universal generalizations, as the RV purports, then to apply theory to phenomena, that is, to deductively link theoretical to observational statements, provisos are required. However, in many theory applications there would be an indefinitely large number of provisos, thus trivializing the concept of scientific laws understood as empirical universal generalizations. In other cases, some provisos would not even be expressible in the language of the theory, thus making the deductive step impossible. Hempel's challenge is that theory–applications presuppose provisos and this does not cohere with the view that theory relates to observation sentences deductively (For an interesting discussion of Hempel's problem of provisos, see [2.32–35]).

2.1.6 Theory Consistency and Meaning Invariance

Feyerabend criticized the logical positivist conception of scientific theories on the ground that it imposes on them a *meaning invariance condition* and a *consistency condition*. By the consistency condition he meant that [2.36, p. 164]

> "[...] only such theories are [...] admissible in a given domain which either *contain* the theories already used in this domain, or which are at least *consistent* with them inside the domain."

By the condition of meaning invariance he meant that [2.36, p. 164]:

> "[...] meanings will have to be invariant with respect to scientific progress; that is, all future theories will have to be framed in such a manner that their use in explanations [or reductions] does not affect what is said by the theories, or factual reports to be explained"

Feyerabend's criticisms are not aimed directly at the RV, but rather at two other claims of logical positivism that are intimately connected to the RV, namely the theses of *the development of theories by reduction* and *the covering law model of scientific explanation*.

A brief digression, in order to look into the aforementioned theses, would be helpful. The development of theories by reduction involves the reduction of one theory (secondary) into a second more inclusive theory (primary). In such developments, the former theory may employ [2.37, p. 342]

> "[...] in its formulations [...] a number of distinctive descriptive predicates that are not included in the basic theoretical terms or in the associated rules of correspondence of the primary [theory] [...]."

That is to say, the V_T terms of the secondary theory are not necessarily all included in the theoretical vocabulary of the primary theory. Nagel builds up his case based on the example of the reduction of thermodynamics to statistical mechanics. There are several requirements that have to be satisfied for theory reduction to take place, two of which are: (1) the V_T terms for both theories involved in the reduction must have unambiguously fixed meanings by codified rules of usage or by established procedures appropriate to each discipline, for example, theoretical postulates or correspondence rules. (2) for every V_T term in the secondary theory that is absent from the theoretical vocabulary of the primary theory, assumptions must be introduced that postulate suitable relations between these terms and corresponding theoretical terms in the primary theory. (See *Nagel* [2.37, pp. 345–358]. In fact *Nagel* presents a larger set of conditions that have to hold in order for reduction to take place [2.37, pp. 336–397], but these are the only two relevant to Feyerabend's arguments).

The covering law model of scientific explanation is, in a nutshell, explanation in terms of a deductively valid argument. The sentence to be explained (explanandum) is a logical consequence of a set of law-premises together with a set of premises consisting of initial conditions or other particular facts involved (explanans). For the special case when the explanandum is a scientific theory, T', the covering law model can be formulated as follows: A theory T explains T' if and only if T together with initial conditions constitute a deductively valid inference with consequence T'. In other words, if T' is derivable from T together with statements of particular facts involved then T' is explained by T. It seems that reduction and explanation of theories go hand in hand, that is, if T' is reduced to T, then T explains T' and conversely.

Feyerabend points out that Nagel's two assumptions – (1) and (2) above – for theory reduction respectively impose a condition of meaning invariance and a consistency condition to scientific progress. The thesis of development of theories by reduction condemns science to restrict itself to theories that are mutually consistent. But the consistency condition requires that terms in the admissible theories for a domain must be used with the same meanings. Similarly, it can be shown that the covering law model of explanation also imposes these two conditions. In fact, the consistency condition follows from the requirement that the explanandum must be a logical consequence of the explanans, and since the meanings of the terms and statements in a logically valid argument must remain constant, an obvious demand for explanation – imposed by the covering law model – is that meanings must be invariant. Feyerabend objects to the meaning invariance and the consistency conditions and argues his case inductively by drawing from historical examples of theory change. For example, the concept of mass does not have the same meaning in relativity theory as it does in classical mechanics. Relativistic mass is a relational concept between an object and its velocity, whereas in classical mechanics mass is a monadic property of an object. Similarly, Galileo's law asserts that acceleration due to gravity is constant, but if Newton's law of gravitation is applied to the surface of the earth it yields a variable acceleration due to gravity. Hence, Galileo's law cannot be derived from Newton's law. By such examples, he attempts to undermine Nagel's assumptions (1) and (2) above and establish that neither meaning invariance nor the related notion of theory consistency characterize actual science and scientific progress (see *Feyerabend* [2.36, 38–40]. Numerous authors have criticized Feyerabend's views. For instance, objections to his views have been raised based on his idiosyncratic analysis of *meaning*, on which his arguments rely. His views are hence not presented here as conclusive criticisms of the RV; but only to highlight that they cast doubt on the adequacy of the theses of theory development by reduction and the covering law model of explanation).

2.1.7 General Remark on the Received View

The RV is intended as an explicative and not a descriptive view of scientific theories. We have seen that even as such it is vulnerable to a great deal of criticism. One way or another, all these criticisms rely on one weakness of the RV: Its inability to clearly spell out the nature of theoretical terms (and how they acquire their meaning) and its inability to specify how sentences consisting of such terms relate to experimental reports. This is a weakness that has been understood by the RV's critics to stem from the former's focus on syntax. By shifting attention away from the representational function of models and attempting to characterize theory structure in syntactic terms, the RV makes itself vulnerable to such objections. Despite all of the above criticisms pointing to the difficulty in explicating how theoretical terms relate to observation, I do not think that any one of them is conclusive in the ultimate sense of rebutting the RV. Nevertheless, the subsequent result was that under the weight of all of these criticisms together the RV eventually made room for its successor.

2.2 The Semantic View of Scientific Theories

The SV has for the last few decades been the standard-bearer of the view that theories are families of models. The slogan *theories are families of models* was meant by the philosophers that originally put forward the SV to stand for the claim that it is more suitable – for understanding scientific theorizing – that the structure of theory is identified with, or presented as, classes of models. A logical consequence of identifying theory structure with classes of models is that models and modeling are turned into crucial components of scientific theorizing. Indeed, this has been one of the major contributions of the SV, since it unquestionably assisted in putting models and modeling at the forefront of philosophical attention. However, identifying theory structure with classes of models is not a logical consequence of the thesis that models (and modeling) are important components of scientific theorizing. Some philosophers who came to this conclusion have since defended the view that although models are crucial to scientific theorizing, the relation between theory and models is much more complex than that of set-theoretical inclusion. I shall proceed in this section by articulating the major features of the SV; in the process I shall try to clarify the notion of model inherent in the view and also explain – what I consider to be – the main difference among its proponents, and finally I will briefly discuss the criticisms against it, which, nevertheless, do not undermine the importance of models in science.

Patrick Suppes was the first to attempt a model-theoretic account of theory structure. He was one of the major denouncers of the attempts by the logical positivists to characterize theories as first-order calculi supplemented by a set of correspondence rules. (See [2.27, 28, 41–43]; much of the work developed in these papers is included in [2.44]). His objections to the RV led him on the one hand to suggest that in scientific practice the theory–experiment relation is more sophisticated than what is implicit in the RV and that theories are not confronted with raw experimental data (as we have seen in Sect. 2.1) but with, what has since been dubbed, *models of data*. On the other hand, he proposed that theories be construed as collections of models. The models are possible realizations (in the Tarskian sense) that satisfy sets of statements of theory, and these models, according to Suppes, are entities of the appropriate set-theoretical structure. Both of these insights have been operative in shaping the SV.

Suppes urged against standard formalizations of scientific theories. First, no substantive example of a scientific theory is worked out in a formal calculus, and second the [2.28, p. 57]

> "[...] very sketchiness [of standard formalizations] makes it possible to omit both important properties of theories and significant distinctions that may be introduced between different theories."

He opts for set-theoretical axiomatization as the way by which to overcome the shortcomings of standard formalization. As mentioned by *Gelfert*, Chap. 1, Suppe's example of a set-theoretical axiomatization is classical particle mechanics (CPM). Three axioms of kinematics and four axioms of dynamics (explicitly stated in Chap. 1 of this volume: *The Ontology of Models*) are articulated by the use of predicates that are defined in terms of set theoretical notions. The structure $\wp = \langle P, T, s, m, f, g \rangle$ can then be understood to be a model of CPM if and only if it satisfies those axioms [2.41, p. 294]. Such a structure is what logicians would label a (semantic) model of the theory, or more accurately a class of models. In general, the model–theoretic notion of a structure, S, is that of an entity consisting of a nonempty set of individuals, D, and a set of relations defined upon the former, R, that is, $S = \langle D, R \rangle$. The set D specifies the domain of the structure and the set R specifies the relations that hold between the individuals in D. (Note that as far as the notion of a structure is concerned, it only matters how many individuals are there and not what they are, and it only matters that the relations in R hold between such and such individuals of D and not what the relations are. For more on this point and a detailed analysis of the notion of structure *Frigg* and *Nguyen*, Chap. 3).

Models of data, according to Suppes, are possible realizations of the experimental data. It is to models of data that models of the theory are contrasted. The RV would have it that the theoretical predictions have a *direct analogue* in the observation statements. This view however, is, according to Suppes, a distorting simplification. As we have seen in Sect. 2.1.3, Suppes defends the claim that by the use of theories of experimental design and other auxiliary theories, the raw data are regimented into a structural form that bears a relation to the models of the theory. To structure the data, as we saw earlier, various influencing factors that the theory does not account for, but are known to influence the experimental data, must be accommodated by an appropriate conversion of the data into canonical form. This regimentation results in a finished product that Suppes dubbed *models of data*, which are structures that could reasonably be contrasted to the models of the theory. Suppes' picture of science as an enterprise of theory construction and empirical testing of theories involves establishing a *hierarchy of models*, roughly consisting of the general categories of models of the theory and models of the data. Furthermore, since the theory–experiment relation is construed as no more than a comparison (i. e., a mapping) of mathematical structures, he invokes the mathematical notion of *isomorphism* of structure to account for the link between theory and experiment. (An isomorphism between structures U and V exists, if there is a function that maps each element of U onto each element of V). Hence, Suppes can be read as urging the thesis that defining the models of the theory and checking for isomorphism with models of data, is a rational reconstruction that does more justice to actual science than the RV does.

The backbone of Suppes' account is the sharp distinction between models of theory and models of data. In his view, the traditional syntactic account of the relation between theory and evidence, which could be captured by the schema: $(T\&A) \to E$ (where, T stands for theory, A for auxiliaries, E for empirical evidence), is replaced by theses (1), (2), and (3) below:

1. $M_\text{T} \subseteq TS$, where M_T stands for model of the theory TS for the theory structure, and $\subseteq$ for the relation of inclusion
2. $(A\&E\&D) \mapsto M_\text{D}$, where M_D stands for model of data, A for auxiliary theories, E for theories of experimental design etc., D for raw empirical data, and $\mapsto$ for ... *used in the construction of* ...
3. $M_\text{T} \approx M_\text{D}$, where $\approx$ stands for mapping of the elements and relations of one structure onto the other.

$M_\text{T} \subseteq TS$ expresses Suppes' view that by defining a theory structure a class of models is laid down for the representation of physical systems. $(A\&E\&D) \mapsto M_\text{D}$ is meant to show how Suppes distances himself from past conceptions of the theory–experiment relation, by claiming that theories are not directly confronted with raw experimental data (collected from the target physical systems) but rather that the latter are used, together with much of the rest of the scientific inventory, in the construction of data structures, M_D. These data structures are then contrasted to a theoretical model, and the theory–experiment relation consists in an isomorphism, or more generally in a mapping of a data onto a theoretical structure, that is, $M_\text{T} \approx M_\text{D}$. The proponents of the SV would, I believe, concur to the above three general theses. Furthermore, they would concur with two of the theses' corollaries: that scientific representation of phenomena can be explicated exclusively by mapping of structures, and that all scientific models constructed within the framework of a particular scientific theory are united under a common mathematical or relational structure. We shall look into these two contentions of the SV toward the end of this section. For now, let me turn our attention to some putative differences between the various proponents of the SV.

Despite agreeing about focusing on the mathematical structure of theories for giving a unitary account of models, it is not hard to notice in the relevant literature that different proponents of the SV have spelled out the details of thesis (1) in different ways. This is because different proponents of the SV have chosen different mathematical entities with which to characterize theory structure. As we saw above, Suppes chooses set theoretical predicates a choice that seems to be shared by *da Costa* and *French* [2.45, 46]. *Van Fraassen* [2.47] on the other hand prefers state-spaces, and *Suppe* [2.30] uses relational systems.

Let us, by way of example, briefly look into van Fraassen's state-space approach. The objects of concern of scientific theories are physical systems. Typically, mathematical models represent physical systems that can generally be conceived as admitting of a certain set of states. *State-spaces* are the mathematical spaces the elements of which can be used to represent the states of physical systems. It is a generic notion that refers to what, for example, physicists would label as phase space in classical mechanics or Hilbert space in quantum mechanics. A simple example of a state-space would be that of an n-particle system. In CPM, the state of each particle at a given time is specified by its position $\boldsymbol{q} = (q_\text{x}, q_\text{y}, q_\text{z})$ and momentum $\boldsymbol{p} = (p_\text{x}, p_\text{y}, p_\text{z})$ vectors. Hence the state-space of an n-particle system would be a Euclidean $6n$-dimensional space, whose points are the $6n$-tuples of real numbers

$$\langle q_{1\text{x}}, q_{1\text{y}}, q_{1\text{z}}, \ldots, q_{\text{nx}}, q_{\text{ny}}, q_{\text{nz}},\\ p_{1\text{x}}, p_{1\text{y}}, p_{1\text{z}}, \ldots, p_{\text{nx}}, p_{\text{ny}}, p_{\text{nz}}\rangle\ .$$

More generally, a state-space is the collection of mathematical entities such as, vectors, functions, or numbers, which is used to specify the set of possible states for a particular physical system. A model, in van Fraassen's characterization of theory structure, is a particular sequence of states of the state-space over time, that is, the state of the modeled physical system evolves over time according to the particular sequence of states admitted by the model. State-spaces unite clusters of models of a theory, and they can be used to single out the class of intended models just as set-theoretical predicates would in Suppes' approach. The presentation of a scientific theory, according to van Fraassen, consists of a description of *a class of state-space types*. As *van Fraassen* explains [2.47, p. 44]:

> "[w]henever certain parameters are left unspecified in the description of a structure, it would be more

> accurate to say [...] that we described a structure type."

The Bohr model of the atom, for example, does not refer to a single structure, but to a structure type. Once the necessary characteristics are specified, it gives rise to a structure for the hydrogen atom, a structure for the helium atom, and so forth.

The different choices of different authors on how theory structure is characterized, however, belong to the realm of personal preference and do not introduce any significant differences on the substance of thesis (1) of the SV, which is that all models of the theory are united under an all-inclusive theory structure. So, irrespective of the particular means used to characterize theory structure, the SV construes models as structures (or structure types) and theories as collections of such structures. Neither have disagreements been voiced regarding thesis (2). On the contrary, there seems to be a consensus among adherents of the SV that models of theory are confronted with models of data and not the direct result of an experimental setup (Not much work has been done to convincingly analyze particular scientific examples and to show the details of the use of *models of data* in science; rather, adherents of the SV repeatedly use the notion with reference to something very general with unclear applications in actual scientific contexts).

2.2.1 On the Notion of Model in the SV

An obvious objection to thesis (1) would be that a standard formalization could be used to express the theory and subsequently define the class of semantic models metamathematically, as the class of structures that satisfy the sentences of the theory, despite Suppes suggestion that such a procedure would be unnecessarily complex and tedious.

In fact, proponents of the SV have often encouraged this objection. *Van Fraassen* and Suppe are notable examples as the following quotations suggest [2.48, p. 326]:

> "There are natural interrelations between the two approaches [i. e., the RV and the SV]: An axiomatic theory may be characterized by the class of interpretations which satisfy it, and an interpretation may be characterized by the set of sentences which it satisfies; though in neither case is the characterization unique. These interrelations [...] would make implausible any claim of philosophical superiority for either approach. But the questions asked and methods used are different, and with respect to fruitfulness and insight they may not be on a par with specific contexts or for special purposes."

> *Suppe* [2.30, p. 82]:
>
> "This suggests that theories be construed as propounded abstract structures serving as models for sets of interpreted sentences that constitute the linguistic formulations. These structures are metamathematical models of their linguistic formulations, where the same structure may be the model for a number of different, and possibly nonequivalent, sets of sentences or linguistic formulations of the theory."

From such remarks, one is justifiably led to believe that propounding a theory as a class of models directly defined, without recourse to its syntax, only aims at convenience in avoiding the hustle of constructing a standard formalization, and at easier adaptability of our reconstruction with common scientific practices. Epigrammatically, the difference – between the SV and the RV – would then be methodological and heuristic. Reasons such as this have led some authors to question the *logical* difference between defining the class of models directly as opposed to metamathematically.

Examples are *Friedman* and *Worrall* who in their separate reviews of *van Fraassen* [2.47] ask whether the class of models that constitutes the theory, according to the proponents of the SV, is to be identified with an elementary class, that is, a class that contains all the models (structures) that satisfy a first-order theory. They both notice that not only does van Fraassen and other proponents of the SV offer no reason to oppose such a supposition, but also they even encourage it (as in the above quotations). But if that is the case [2.49, p. 276]:

> "[t]hen the completeness theorem immediately yields the equivalence of van Fraassen's account and the traditional syntactic account [i. e., that of the RV]."
>
> In other words [2.50, p. 71]:
>
> "So far as logic is concerned, syntax and semantics go hand-in-hand – to every consistent set of first-order sentences there corresponds a nonempty set of models, and to every normal (elementary) set of models there corresponds a consistent set of first-order sentences."

If we assume (following Friedman and Worrall) that the proponents of the SV are referring to the elementary class of models then the preceding argument is sound. The SV, in agreement with the logical positivists, retains formal methods as the primary tool for philosophical analysis of science. The only new elements of its own would be the suggestions that first it is more convenient that rather than developing these methods using proof–theory we should instead use formal semantics (model-theory), and second we should assign to models (i. e., the semantic interpretations of sets of sentences) a representational capacity.

Van Fraassen, however, resists the construal of the class of models of the SV with an elementary class (See *van Fraassen* [2.51, pp. 301–303] and his [2.52]). Let me rehearse his argument. The SV claims that to present a theory is to define a class M of models. This is the class of structures the theory makes available for modeling its domain. For most scientific theories, the real number continuum would be included in this class. Now his argument goes, if we are able to formalize what is meant to be conveyed by M in some appropriate language, then we will be left with a class N of models of the language, that is, the class of models in which the axioms and theorems of the language are satisfied. Our hope is that every structure in M occurs in N. However, the real number continuum is infinite and [2.52, p. 120]:

> "[t]here is no elementary class of models of a denumerable first-order language each of which includes the real numbers. As soon as we go from mathematics to metamathematics, we reach a level of formalization where many mathematical distinctions cannot be captured."

Furthermore, "[t]he *Löwenheim–Skolem* theorems [...] tell us [...] that N contains many structures not isomorphic to any member of M" [2.51, p. 302]. Van Fraassen relies, here, on the following reasoning: The Löwenheim–Skolem theorem tells us that all satisfiable first-order theories that admit infinite models will have models of all different infinite cardinalities. Now models of different cardinality are nonisomorphic. Consequently, every theory that makes use of the real number continuum will have models that are not isomorphic to the intended models (i. e., nonstandard interpretations) but which satisfy the axioms of the theory. So van Fraassen is suggesting that M is the intended class of models, and since the limitative meta-theorems tell us that it cannot be uniquely determined by any set of first-order sentences we can only define it directly. Here is his concluding remark [2.51, p. 302]:

> "The set N contains [...] [an] image M^* of M, namely, the set of those members of N which consist of structures in M accompanied by interpretations therein of the syntax. But, moreover, [...] M^* is not an elementary class."

Evidently, van Fraassen's argument aims to establish that the directly defined class of models is not an elementary class. It is hard, however, to see that defining the models of the theory directly without resort to formal syntax yields only the intended models of theory (i. e., excludes all nonstandard models), despite the possibility that one could see the prospect of the SV being heuristically superior to the RV. (Of course, we must not forget that this superiority would not necessarily be the result of thesis (1) of the SV, but it could be the result of its consequence of putting particular emphasis on the significance of scientific models that, as noted earlier, does not logically entail thesis (1)).

Let us, for the sake of argument, ignore the Friedman–Worrall argument. Now, according to the SV, models of theory have a dual role. On the one hand, they are devices by which phenomena are represented, and on the other, they are structures that would satisfy a formal calculus were the theory formalized. The SV requires this dual role. First because the representational role of models is the way by which the SV accounts for scientific representation without the use of language; and second because the role of interpreting a set of axioms ensures that a unitary account of models is given. Now, *Thompson-Jones* [2.53] notices that the notion of model implicit in the SV is either that of an interpretation of a set of sentences or a mathematical structure (the disjunction is of course inclusive). He analyzes the two possible notions and argues that the SV becomes more tenable if the notion of model is only understood as that of a mathematical structure that functions as a representation device. If that were the case then the adherents of the SV could possibly claim that defining the class of structures directly indeed results in something distinct from the metamathematical models of a formal syntax. Thompson-Jones' suggestion, however, would give rise to new objections. Here is one. It would give rise to the following question: How could a theory be identified with a class of models (i. e., mathematical structures united under an all-inclusive theory structure) if the members of such a class do not attain membership in the class because they are interpretations of the same set of theory axioms? In other words, the proponents of the SV would have to explain what it is that *unites* the mathematical models other than the satisfaction relation they have to the theoretical axioms. To my knowledge, proponents of the SV have not offered an answer to this question. If Thompson-Jones' suggestion did indeed offer a plausible way to overcome the Friedman–Worrall argument then the SV would have to abandon the quest of giving a unitary account of models. Given the dual aim of the SV, namely to give a unitary account of models and to account for scientific representation by means of structural relations, it seems that the legitimate notion of model integral to this view must have these two-hard to reconcile-roles; namely, to function both as an interpretation of sets of sentences and as a representation of phenomena. (Notice that this dual function of models is an aspect of all versions of the SV, independent of how one chooses to characterize theory structure and of how one chooses to interpret that structure).

2.2.2 The Difference Between Various Versions of the SV

The main difference among the various versions of the SV relates to two intertwined issues that relate to thesis (3), namely how the theory structure is construed and how the theory–experiment mapping relation is construed. To a first approximation we could divide the different versions of the SV, from the perspective of these two issues, into two sorts. Those in which particular emphasis is given to the presence of abstraction and idealization in scientific theorizing for explicating the theory–experiment (or model–experiment) relation, and those in which the significance of this nature of scientific theorizing is underrated.

Idealization and Abstraction Underrated

Van Fraassen (Suppes most probably could be placed in this group too), for example, seems to be a clear case of this sort. Here is how he encapsulates his conception of scientific theories and of how theory relates to experiment [2.47, p. 64]:

> "To present a theory is to specify a family of structures, its *models*; and secondly, to specify certain parts of those *models* (*the empirical substructures*) as candidates for the direct representation of observable phenomena. The structures which can be described in experimental and measurement reports we can call *appearances*: The theory is empirically adequate if it has some model such that all appearances are isomorphic to empirical substructures of that model."

Appearances (which is van Fraassen's term for *models of data*) are relational structures of measurements of observable aspects of the target physical system, for example, relative distances and velocities. For example, in the Newtonian description of the solar system, as *van Fraassen* points out, the relative motions of the planets "[...] form relational structures defined by measuring relative distances, time intervals, and angles of separation" [2.47, p. 45]. Within the theoretical model for this physical system, "[...] we can define structures that are meant to be exact reflections of those appearances [...]" [2.47, p. 45]. Van Fraassen calls these *empirical substructures*. When a theory structure is defined each of its models, which are candidates for the representation of phenomena, includes empirical substructures. So within representational models we could specify a division between observable/nonobservable features (albeit this division is not drawn in linguistic terms), and the empirical substructures of such models are assumed to be isomorphic to the observable aspects of the physical system. In other words, the theory structure is interpreted as having distinctly divided observable and nonobservable features, and the theory–experiment relation is interpreted as being an isomorphic relation between the data model and the observable parts of the theoretical model. Now, the state-space is a class of models, it thus includes – for CPM – many models in which the world is a Newtonian mechanical system. In fact, it seems that the state-space includes (unites) all logically possible models, as the following dictum suggests ([2.52, p. 111], [2.54, p. 226]):

> "In one such model, nothing except the solar system exists at all; in another the fixed stars also exist, and in a third, the solar system exists and dolphins are its only rational inhabitants."

According to van Fraassen, the theory is empirically adequate if we can find a model of the theory in which we can specify empirical substructures that are isomorphic to the data model. The particular view of scientific representation that resides within this idea is this: *A model represents its target if and only if it is isomorphic to a data model constructed from measurements of the target*. Not much else seems to matter for a representation relation to hold but the *isomorphism condition*. Many would argue, however, that such a condition for the representation relation is too strong to explicate how actual scientific models relate to experimental results and would object to this view on the ground that for isomorphism to occur it would require that target physical systems occur under highly idealized conditions or in isolated circumstances. (Admittedly, it would not be such a strong requirement for models that would only describe observable aspects of the world. In such cases isomorphism could be achieved, but at the expense of the model's epistemic significance. I do not think, for instance, that such models would be of much value to a science like Physics as, more often than not, they would be useless in predicting the future behavior of their targets).

Idealization and Abstraction Highlighted

In the second camp of the SV, we encounter several varieties. One of these is *Suppe* [2.30], who interprets theory structure and the theory–experiment relation as follows. Theories characterize particular classes of target systems. However, target systems are not characterized in their full complexity, as already mentioned in Sect. 2.1.4. Instead, Suppe's understanding is that certain parameters are abstracted and employed in this characterization. In the case of CPM, these are the position and momentum vectors. These two parameters are abstracted from all other characteristics that target systems may possess. Furthermore, once the factors, which are assumed to influence the class of target systems in the theory's intended scope, have been abstracted the characterization of physical systems (as mentioned in Sect.2.1.4, physical systems in Suppe's terminology refer to the abstract entities that models of the theory represent and not to the actual target systems) still does not *fully* account for target systems. Physical systems are not concerned with the actual values of the parameters the particulars possess, for example, actual velocities, but with the values of these parameters under certain conditions that obtain only within the physical system itself. Thus in CPM, where the behavior of dimensionless point-masses are studied in isolation from outside interactions, physical systems characterize this behavior only by reference to the positions and momenta of the point-masses at given times.

An example can serve to demonstrate Suppe's idea in bit more detail. The linear harmonic oscillator, that is, a *mathematical instrument*, is expressed by the following equation of motion $\ddot{x} + (k/m)x = 0$, which is the result of applying Newton's second law to a linear restoring force. The mathematical model is interpreted (and thus characterizes a *physical system*) as follows: Periodic oscillations are assumed to take place with respect to time, x is the displacement of an oscillating mass-point, and k and m are constant coefficients that may be replaced by others. When the mathematical parameters in the above equation are linked to features of a specific object, the equation can be used to model for instance the torsion pendulum, that is, an elastic rod connected to a disk that oscillates about an equilibrium position. This sort of linking of mathematical terms to features of objects could be understood to be a manifestation of what Giere calls identification. Giere introduces a useful distinction between *interpretation* and *identification* [2.55, p. 75]:

> "[...] [Interpretation] is the linking of the mathematical symbols with *general terms*, or concepts, such as *position*[...] [Identification] is the linking of a mathematical symbol with some feature of a *specific object*, such as *the position of the moon*."

In the torsion pendulum model, x is identified with the angle of twist, k with the torsion constant, and m with the moment of inertia. By linking the mathematical symbols of a model to features of a target system we can reasonably assume, according to Suppe, that the model could be associated with an actual system of the world; the model characterizes, as Suppe would say in his own jargon, "a causally possible physical system."

However, even when a certain mathematical product of theory is identified with a causally possible physical system, we still know that typically the situation described by the physical system does not obtain. The actual torsion pendulum apparatus is subject to a number of different factors (or may have a number of different characteristics) that may or may not influence the process of oscillation. Some influencing factors are the amplitude of the angle of oscillation, the mass distribution of the rod and disc, the nonuniformity of the gravitational field of the earth, the buoyancy of the rod and disc, the resistance of the air and the stirring up of the air due to the oscillations. In modeling the torsion pendulum by means of the linear harmonic oscillator the physical system is abstracted from factors assumed to influence the oscillations in the same manner as from those assumed not to. Therefore, the replicating relation between the physical system, P, and the target system, S, which Suppe urges cannot be understood as one of identity or isomorphism. *Suppe* is explicit about this [2.30, p. 94]:

> "The attributes in P determine a sequence of states over time and thus indicate a possible behavior of S [...] Accordingly, P is a kind of *replica* of S; however, it need not replicate S in any straight-forward manner. For the state of P at t does not indicate what attributes the particulars in S possess at t; rather, it indicates what attributes they *would have* at t *were* the abstracted parameters the only ones influencing the behavior of S and were certain idealized conditions met. In order to see how P replicates S we need to investigate these abstractive and idealizing conditions holding between them."

In summary, the replicating relation is counterfactual: If the conditions assumed to hold for the description of the physical system were to hold for the target system, then the target system would behave in the way described by the physical system. The behavior of actual target systems, however, may be subject to other unselected parameters or other conditions, for which the theory does not account.

The divergence of Suppe's view from that of van Fraassen is one based primarily on the representation relation of theory to phenomena. Suppe understands the theory structure as being a highly abstract and idealized representation of the complexities of the real world. Van Fraassen disregards this because he is concerned with the observable aspects of theories and assumes that these can, to a high degree of accuracy, be captured by experiments. Thus van Fraassen regards theories as containing empirical substructures that stand in isomorphic relations to the observable aspects of the world. Suppe's understanding of theory structure, however, points to a significant drawback present in van Fraassen's view: How can isomorphism obtain between a data model and an empirical substructure of the model, given that the model is abstract and idealized? Suppe's difference with van Fraassen's view of the representation relation and of the epistemic inferences that can be drawn from it is this, if indeed it is the case that isomorphism obtains between a data model and an empirical substructure, then it is so for either of two reasons: (1) the experiment is highly idealized, or (2) the data model is converted to what the measurements *would have been* if the influences that are not accounted by the theory did not have any effect on the experimental setup. This is a significantly different claim from what van Fraassen would urge, to wit that the world or some part of it is isomorphic to the model. According to Suppe's understanding of theory structure, no part of the world is or can be isomorphic to a model of the theory, because abstraction and idealization are involved in scientific theorizing.

Geire [2.55] is another example of a version of the SV that places the emphasis on abstraction and idealization. Following Suppes and van Fraassen, Giere understands theories as classes of models. He does not have any special preference about the mathematical entities by which theory structure is characterized, but he is interested in looking at the characteristics of actual science and how these could be captured by the SV. This leads him to a similar claim as Suppe. He claims that although he does not see any logical reason why a real target system could not be isomorphic to a model, nevertheless for the examples of models found in mechanics texts, typically, no claim of isomorphism is made, indeed "[...] the texts often explicitly note respects in which the model fails to be isomorphic to the real system" [2.55, p. 80]. He attributes this to the abstract and idealized nature of models of the theory. His solution is to substitute the strict criterion of isomorphism, as a way by which to explicate the theory–experiment relation, with that of similarity in relevant respects and degrees between the model and its target.

Finally, there is another example of a version of the SV that also gives attention to idealization and abstraction, namely the version advocated by *da Costa* and *French* in [2.45, 46, 56]. They do this indirectly by interpreting theories as partial structures, that is, structures consisting of a domain of individuals and a set of partial relations defined on the domain, where a partial relation is one that is not defined for all the n-tuples of individuals of the domain for which it presumably holds. If models of theory are interpreted in this manner and if it is assumed that models of data are also partial structures, then the theory–experiment relation is explicated by *da Costa* and *French* [2.46] as a partial isomorphism. A partial isomorphism between two partial structures U and V exists when a partial substructure of U is isomorphic to a partial substructure of V. In other words, partial isomorphism exists when some elements of the set of relations in U are mapped onto elements of the set of relations in V. If a model of theory is partially isomorphic to a data model then, da Costa and French claim, the model is partially true. The notion of partial truth is meant to convey a pragmatic notion of truth, which plausibly could avoid the problems of correspondence or complete truth, and capture the commonplace idea that theories (or models) are incomplete or imperfect or abstract or idealized descriptions of target systems.

In conclusion, if we could speak of *different* versions of the SV and not just different formulations of the same idea, if, in other words, the proposed versions of the semantic conception of theories can be differentiated in any significant way amongst them, it is on the basis of how thesis (3) is conceived: There are those that understand the representation relation, $M_{\mathrm{T}} \approx M_{\mathrm{D}}$, as a strict isomorphic relation, and those that construe it more liberally, for example, as a similarity relation. In particular, van Fraassen prefers an isomorphic relation between theory and experiment, whereas Suppe and others understand theories as being abstract and idealized representations of phenomena. It would seem therefore that particular criticisms would not necessarily target both versions. This has not been the case however, as we shall examine in the next two subsections. Critics of the SV have either targeted theses (1) and (2) and the unitary account of models implicit in the SV, or thesis (3) and the representation relation however the latter is conceived. The arguments against the unitary account of scientific models, which obviously aim indiscriminately at all versions of the SV, will be explored in Sect.2.2.4. The arguments against the nature of the representation relation implied by the SV, which shall be explored in Sect.2.2.3, if properly adapted affect both versions of the SV.

2.2.3 Scientific Representation Does not Reduce to a Mapping of Structures

Suarez [2.57] presents five arguments against the idea that scientific representation can be explicated by appealing to a structural relation (like isomorphism or similarity) that may hold between the representational device and the represented target. (*Suarez* [2.57] also develops his arguments for other suggested interpretations of theses (3), such as partial isomorphism). These arguments, which are summarized below, imply that the representational capacity of scientific models cannot derive from having a structural relation with its target. Suarez's first argument is that in science many disparate things act as representational devices, for example, a mathematical equation, or a Feynman diagram, or an architect's model of a building, or the double helix macro-model of the DNA molecule. Neither isomorphism nor similarity can be applied to such disparate representational devices in order to explicate their representational function. A similar point is also made by *Downes* [2.58], who by also exploring some examples of scientific models, argues that models in science relate to their target systems in various ways, and that attempts to explicate this relation by appeal to isomorphism or similarity does little to serve the purpose of understanding the theory–experiment relation.

The second argument concerns the logical properties of representation vis-a-vis those of isomorphism and similarity. Suarez explains that representation is nonsymmetric, nonreflexive and nontransitive. If scientific representation is a type of representation then any attempt to explicate scientific representation cannot imply different logical features from representation. But appeal to a structural relation does not accomplish this, because "[...] similarity is reflexive and symmetric, and isomorphism is reflexive, symmetric and transitive" [2.55, p. 233].

His third argument is that any explication of representation must allow for misrepresentation or inaccurate representation. Misrepresentation, he explains, occurs either when the target of a representation is mistaken or when a representation is inaccurate because it is either incomplete or idealized. Neither isomorphism nor similarity allows for the first kind of misrepresentation and isomorphism does not allow for the second kind. Although, similarity does account for the second kind of representation, Suarez argues, it does so in a restrictive sense. That is, if we assume that an incomplete representation is given according to theory X then similarity does account for misrepresentation. However, if a complete representation were given according to theory X (i. e., if we have similarity in all relevant respects that X dictates) but the predictions of this representation still diverge from measurements of the values of the target's attributes then similarity does not account for this kind of misrepresentation.

The fourth argument is that neither isomorphism nor similarity is necessary for representation. Our intuitions about the notion of representation allow us to accept the representational device derived from theory X as a *representation* of its target, even though we may know that isomorphism or similarity does not obtain because, for example, an alternative theory Y not only gives us better predictions about the target but also tells us why X fails to produce representational devices that are isomorphic or similar to their targets. A different argument but with the same conclusion is given by *Portides* [2.59], who argues that isomorphism, or other forms of structural mapping, is not necessary for representation because it is possible to explicate the representational function of some successful quantum mechanical models, which are not isomorphic to their targets. Suarez's final argument is that neither isomorphism nor similarity is sufficient for representation. In other words, even though there may not be a representation relation between A and B, A and B may, however, be isomorphic or similar.

Aiming at the same feature of the SV as Suarez, *Frigg* [2.60] reiterates some of the arguments above and gives further reasons to fortify them, but he also presents two more arguments that undermine the notion of representation as dictated by thesis (3) of the SV. Employed in his first argument is a particular notion of abstractness of concepts advocated by *Cartwright* [2.61]. A concept is considered abstract in relation to a set of more concrete concepts if for the former to apply it is necessary that one of its concrete instances apply. One of Frigg's intuitive examples is that the concept of traveling is more abstract than the concept of sitting in a moving train. So according to this sense of *abstractness* the concept of traveling applies whenever one is sitting in a moving train and that the abstract concept does not apply if one is not performing some action that belongs to the set of concrete instances of traveling. *Frigg* then claims, "[...] that possessing a structure is abstract in exactly this sense and it therefore does not apply without some more concrete concepts applying as well" [2.60, p. 55]. He defends this claim with the following argument. Since to have a structure means to consist of a set of individuals which enter into some relations, then it follows that whenever the concept of possessing a structure applies to S the concept of being an individual applies to members of a set of S and the concept of being in a relation applies to some parts of that set. The concepts of being an individual and being in a relation are abstract in the above sense. For example, given the proper context, for *being an individual* to apply, *occupying a certain space-time region* has to apply. Similarly, given the proper context, for *being in a relation* to apply it must be the case that *being greater than* applies. Therefore, both being an individual and being in a relation are abstract. Thus Frigg concludes, *possessing a structure* is abstract; hence for it to apply, it must be the case that a concrete description of the target applies. Because, the claim that the representation relation can be construed as an isomorphism (or similarity) of structures presupposes that the target possesses a structure, Frigg concludes that such a claim "[...] presupposes that there is a more concrete description that is true of the [target] system" [2.60, p. 56]. This argument shows that to reduce the representation relation to a mapping of structures the proponents of the SV need to invoke nonstructural elements into their account of representation, so pure and simple reduction fails.

Frigg's second argument, as he states, is inductive. He examines several examples of systems from different contexts in order to support the claim that a target system does not have a unique structure. For a system to have a structure it must be made of individuals and relations, but slicing up the physical systems of the world into individuals and relations is dependent on how we conceptualize the world. The world itself does not provide us with a unique slicing. "Because different conceptualizations may result in different structures there is no such thing as the one and only structure of a system" [2.60, p. 57]. One way that Frigg's argument could be read is this: Thesis (2) of the SV implies that the measurements of an experiment are structured to form a data model. But, according to Frigg, this structuring is not unique. So the claim of thesis (3), that there is, for example, an isomorphism between a theoretical model and a data model is not epistemically informative since there may be numerous other structures that could be constructed from the data that are not isomorphic to the theoretical model.

2.2.4 A Unitary Account of Models Does not Illuminate Scientific Modeling Practices

The second group of criticisms against the SV consists of several heterogeneous arguments stemming from different directions and treating a variety of features and functions of models. Despite this heterogeneity, they can be grouped together because they all indirectly undermine the idea that the unitary account of scientific models given by employing a set theoretical (or other mathematical) characterization of theory structure is adequate for understanding the notion of representational model and the model–experiment relation. This challenge to the SV is indirect because the main purpose of these arguments is to illuminate particular features of actual scientific models. In highlighting these features, these arguments illustrate that actual representational models in science are constructed in ways that are incompatible with the SV, they function in ways that the SV does not adequately account for and they represent in ways that is incompatible with the SV's account of representation; furthermore, they indicate that models in science are complex entities that cannot be thoroughly understood by unitary accounts such as set-theoretical inclusion. In other words, a consequence of most of these arguments is that the unitary account of models that the SV provides through thesis (1) that all models are constitutive parts of theory structure, obscures the particular features that representational scientific models demonstrate.

One such example is *Morrison* [2.62], who argues that models are partially autonomous from the theories that may be responsible for instigating their construction. This partial autonomy is something that may derive from the way they function but also from the way they are constructed. She discusses Prandtl's hydrodynamic model of the boundary layer in order to mark out that the inability of theory to provide an explanation of the phenomenon of fluid flow did not hinder scientific modeling. Prandtl constructed the model with little reliance on high-level theory and with a conceptual apparatus that was partially independent from the conceptual resources of theory. This partial independence in construction, according to Morrison, gives rise to functional independence and renders the model partially autonomous from theory. Furthermore, *Morrison* raises another issue (see [2.62], as well as [2.63]); that theories, and hence theoretical models as direct conceptual descendants of theory, are highly abstract and idealized descriptions of phenomena, and therefore they represent only the general features of phenomena and do not explain the specific mechanisms at work in physical systems. In contrast, actual representational scientific models – that she construes as partially autonomous mediators between theories and phenomena – are constructed in ways that allow them to function as explanations of the specific mechanisms and thus function as sources of knowledge about corresponding target systems and their constitutive parts. (As she makes clear in *Morrison* [2.64], to regard a model as partially independent from theory does not mean that theory plays an unimportant role in its construction). This argument, in which representational capacity is correlated to the explanatory power of models, is meant to achieve two goals. Firstly, to offer a way by which to go beyond the narrow understanding of scientific representation as a mapping relation of structure, and second, to offer a general way to understand the representational function of both kinds of models that physicists call theory-driven and phenomenological (In *Portides* [2.65] a more detailed contrast between Morrison's view of the representation relation and that of the SV is offered). *Cartwright* et al. [2.66] and *Portides* [2.67] have also argued that by focusing exclusively on theory-driven models and the mapping relation criterion, the SV obscures the representational function of phenomenological models and also many aspects of scientific theorizing that are the result of phenomenological methods.

It is noteworthy that the unitary account that the SV offers may be applicable to theory-driven models. Whether that is helpful or not is debatable. However, more often than not representation in science is achieved by the use of phenomenological models or phenomenological elements incorporated into theory-driven models. One aspect of Morrison's argument is that if we are not to dismiss the representational capacity of such models we should give up unitary accounts of models. Cartwright makes a similar point but her approach to the same problem is from another angle.

Cartwright [2.61, 68] claims that theories are highly abstract and thus do not and cannot represent what happens in actual situations. Cartwright's observation seems similar to versions of the SV such as Suppe's, however her approach is much more robust. To claim that theories represent what happens in actual situations, she argues, is to overlook that the concepts used in them – such as, *force functions* and *Hamiltonians* – are abstract. Such abstract concepts could only apply to the phenomena whenever more concrete descriptions (as those present in models) can stand-in for them and for this to happen the bridge principles of theory must mediate. Hence the abstract terms of theory apply to actual situations via bridge principles, and this makes bridge principles an operative aspect of theory-application to phenomena. It is only when bridge principles sanction the use of theoretical models that we are led to the construction of a model – with a relatively close relation to theory – that represents the target system. But Cartwright observes that there are only a small number of such theoretical models that can be used successfully to construct representations of physical systems and this is because there are only a handful of theory bridge principles. In most other cases, where no bridge principles exist that enable the use of a theoretical model, concrete descriptions of phenomena are achieved by constructing phenomenological models. Phenomenological models are constructed with minimal aid from theory, and surely there is no deductive (or structural) relation between them and theory. The relation between the two should be sought in the nature of the abstract–concrete distinction between scientific concepts, according to Cartwright. Models in science, whether constructed phenomenologically or by the use of available bridge principles, encompass descriptions that are in some way independent from theory because they are made up of more concrete conceptual ingredients. A weak reading of this argument is that the SV could be a plausible suggestion for understanding the structure of scientific theories for use in foundational work. But in the context of utilizing the theory to construct representations of phenomena, focusing on the structure of theory does not illuminate much because it is not sufficient as to account for the abstract–concrete distinction that exists between theory and models. A stronger reading of the argument is that the structure of theories is completely irrelevant to how theories represent the world, because they just do not represent it at all. Only models represent pieces of the world and they are partially independent from theory because they are constituted by concrete concepts that apply only to particular physical systems.

Other essays in the volume by *Morgan* and *Morrison* [2.69] discuss different aspects of partial independence of models from theory. Here are two brief examples that aim to show the partial independence of model construction from theory. *Suarez* [2.70] explains how simplifications and approximations that are introduced into representational models (such as the London brothers model of superconductivity) are decided independently of theory and of theoretical requirements. This process gives rise to a model that mediates in the sense that the model itself is the means by which corrections are established that may be incorporated into theory in order to facilitate its applications. But even in cases of models that are strongly linked to theory such as the MIT-bag model of quark confinement, *Hartmann* [2.71] argues, many parts of the model are not motivated by theory but by an accompanying *story* about quarks. From the empirical fact that quarks were not observed physicists were eventually led to the hypothesis that quarks are confined. But confinement is not something that follows from theory. Nevertheless, via the proper amalgam of theory and *story* about quarks the MIT-bag model was constructed to account for quark confinement.

I mentioned earlier in Sect. 2.2.2 that *Giere* [2.55] is also an advocate of the SV. However, his later writings [2.72, 73] suggest that he makes a gradual shift from his earlier conception of representational models in science to a view that neighbors that of Morrison and Cartwright. Even in *Giere* [2.55] the reader notices that he, unlike most other advocates of the SV, is less concerned with the attempt to give a unitary account of models and more concerned with the importance of models in actual scientific practices. But in [2.72] and [2.73] this becomes more explicit. *Giere* [2.55] espouses the idea that the laws of a theory are definitional devices of theoretical models. This view is compatible with the use of scientific laws in the SV. However, in *Giere* [2.72, p. 94] he suggests that scientific laws "[...] should be understood as rules devised by humans to be used in building models to represent specific aspects of the natural world." It is patent that operating as rules for building models is quite a different thing from understanding laws to be the means by which models are defined. The latter view is in line with the three theses of the SV; the former however is only in line with the view that models are important in scientific theorizing. Moreover, in *Giere* [2.73] he makes a more radical step in distinguishing between the abstract models (which he calls *abstract objects*) defined by the laws and those models used by scientists to represent physical systems (which he calls *representational models*). The latter [2.73, p. 63]

> "[. . .] are designed for use in representing aspects of the world. The abstract objects defined by scientific principles [i. e., scientific laws] are, on my view, not intended directly to represent the world."

Giere points to the important difference between the SV and its critics. The SV considers the models that the theory directly delivers representations of target systems of the world. Its critics do not think that; they argue that many successful representational models are constructed by a variety of conceptual ingredients and thus have a degree of autonomy from theory. But if each representational model is partially autonomous from the theory that prompted its construction then a unitary account of representational models does not seem to be much enlightening in enhancing our understanding of why models are so important in scientific theorizing.

2.2.5 General Remark on the Semantic View

Just like its predecessor the SV employs formal methods for the philosophical analysis of scientific theories. In the SV, models of the theory are directly defined by the laws of the theory, and are thus united under a common mathematical structure. Of course, mathematical equations satisfy a structure, no one disputes that mathematically formulated theories can be presented in terms of mathematical structures. Nonetheless, keen to overcome the philosophical problems associated with the RV and its focus on the syntactic elements of theories, the proponents of the SV take the idea of presenting theories structurally one step further. They claim that the SV not only offers a canonical structural formulation for theories, into which any theory can be given an equivalent reformulation (an idea that, no doubt, is useful for the philosophy of mathematics), but they also contend that a scientific theory represents phenomena *because* this structure can be linked to empirical data. To defend this assertion, the proponents of the SV assume that in science there is a sharp distinction between models of theory and models of data and argue that scientific representation is no more than a mapping relation between these two kinds of structures. As we have seen, serious arguments against the idea that representation can be reduced to structural mapping have surfaced; and these arguments counter the SV independently of how the details of the mapping relation is construed.

Furthermore, the SV implies that by defining a theory structure an indefinite number of models that are thought to be antecedently available for modeling the theory's domain are laid down. Neither this position has gone unnoticed. Critics of the SV claim that this idea does not do justice to actual science because it undervalues the complexities involved in actual scientific model construction and the variety of functions that models have in science, but more importantly because it obscures the features of representational models that distinguish them from the models that are direct descendants of theory.

I claimed that the SV employs a notion of model that has two functions – interpretation and representation. In addition, it requires models that have this dual role to be united under a common structure. It is hard to reconcile these two ideas and do justice to actual science. The devices by which the theoretical models are defined, according to the SV, are the laws of the theory. Hence the laws of the theory provide the constraints that determine the structure of these models. Now, it is not hard to see that models viewed as interpretations are indeed united under a common structure determined by the laws of the theory. What is problematic, however, is that the SV assumes that models that are interpretations also function as representations and this means that models functioning as representations can be united under a common structure. The truth value of the conjunction *models are interpretations and representations* is certainly not a trivial issue. When scientists construct representational models, they continuously impose constraints that alter their initial structure. The departure of the resulting constructs from the initial structure is such that it is no longer easily justified to think of them all as united under a common theory structure. Indeed, in many scientific cases this departure of individual representational models is such that they end up having features that may be incompatible with other models that are also instigated by the same theory. These observations lead to the thought that the model-theory and the model–experiment relations may in the end be too complex for our formal tools to capture.

References

2.1 F. Suppe: The search for philosophic understanding of scientific theories. In: *The Structure of Scientific Theories*, ed. by F. Suppe (Univ. Illinois Press, Urbana 1974) pp. 1–241

2.2 P. Achinstein: The problem of theoretical terms, Am. Philos. Q. **2**(3), 193–203 (1965)

2.3 P. Achinstein: *Concepts of Science: A Philosophical Analysis* (Johns Hopkins, Baltimore 1968)

2.4 H. Putnam: What theories are not. In: *Logic, Methodology and Philosophy of Science*, ed. by E. Nagel, P. Suppes, A. Tarski (Stanford Univ. Press, Stanford 1962) pp. 240–251

2.5 Theoretician's dilemma: A study in the logic of theory construction. In: *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, ed. by C. Hempel, C. Hempel (Free Press, New York 1958) pp. 173–226

2.6 R. Carnap: The methodological character of theoretical concepts. In: *Minnesota Studies in the Philosophy of Science: The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, Vol. 1, ed. by H. Feigl, M. Scriven (Univ. Minnesota Press, Minneapolis 1956) pp. 38–76

2.7 R. Carnap: *Philosophical Foundations of Physics* (Basic Books, New York 1966)

2.8 F. Suppe: Theories, their formulations, and the operational imperative, Synthese **25**, 129–164 (1972)

2.9 N.R. Hanson: *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science* (Cambridge Univ. Press, Cambridge 1958)

2.10 N.R. Hanson: *Perception and Discovery: An Introduction to Scientific Inquiry* (Freeman, San Francisco 1969)

2.11 J. Fodor: *The Modularity of Mind* (MIT, Cambridge 1983)

2.12 J. Fodor: Observation reconsidered, Philos. Sci. **51**, 23–43 (1984)

2.13 J. Fodor: The modularity of mind. In: *Meaning and Cognitive Structure*, ed. by Z. Pylyshyn, W. Demopoulos (Ablex, Norwood 1986)

2.14 Z. Pylyshyn: Is vision continuous with cognition?, Behav. Brain Sci. **22**, 341–365 (1999)

2.15 Z. Pylyshyn: *Seeing and Visualizing: It's Not What You Think* (MIT, Cambridge 2003)

2.16 A. Raftopoulos: Is perception informationally encapsulated?, The issue of the theory-ladenness of perception, Cogn. Sci. **25**, 423–451 (2001)

2.17 A. Raftopoulos: Reentrant pathways and the theory-ladenness of observation, Phil. Sci. **68**, 187–200 (2001)

2.18 A. Raftopoulos: *Cognition and Perception* (MIT, Cambridge 2009)

2.19 R. Carnap: Meaning postulates, Philos. Stud. **3**(5), 65–73 (1952)

2.20 W.V. Quine: Two dogmas of empiricism. In: *From a Logical Point of View*, (Harvard Univ. Press, Massachusetts 1980) pp. 20–46

2.21 M.G. White: The analytic and the synthetic: An untenable dualism. In: *Semantics and the Philosophy of Language*, ed. by L. Linsky (Univ. Illinois Press, Urbana 1952) pp. 272–286

2.22 P. Achinstein: Theoretical terms and partial interpretation, Br. J. Philos. Sci. **14**, 89–105 (1963)

2.23 R. Carnap: Testability and meaning, Philos. Sci. **3**, 420–468 (1936)

2.24 R. Carnap: Testability and meaning, Philos. Sci. **4**, 1–40 (1937)

2.25 C. Hempel: *Fundamentals of Concept Formation in Empirical Science* (Univ. Chicago Press, Chicago 1952)

2.26 K.F. Schaffner: Correspondence rules, Philos. Sci. **36**, 280–290 (1969)

2.27 P. Suppes: Models of data. In: *Logic, Methodology and Philosophy of Science*, ed. by E. Nagel, P. Suppes, A. Tarski (Stanford Univ. Press, Stanford 1962) pp. 252–261

2.28 P. Suppes: What is a scientific theory? In: *Philosophy of Science Today*, ed. by S. Morgenbesser (Basic Books, New York 1967) pp. 55–67

2.29 F. Suppe: What's wrong with the received view on the structure of scientific theories?, Philos. Sci. **39**, 1–19 (1972)

2.30 F. Suppe: *The Semantic Conception of Theories and Scientific Realism* (Univ. Illinois Press, Urbana 1989)

2.31 C. Hempel: Provisos: A problem concerning the inferential function of scientific theories. In: *The Limitations of Deductivism*, ed. by A. Grünbaum, W.C. Salmon (Univ. California Press, Berkeley 1988) pp. 19–36

2.32 M. Lange: Natural laws and the problem of provisos, Erkenntnis **38**, 233–248 (1993)

2.33 M. Lange: Who's afraid of ceteris paribus laws?, or: How I learned to stop worrying and love them, Erkenntnis **57**, 407–423 (2002)

2.34 J. Earman, J. Roberts: Ceteris paribus, there is no problem of provisos, Synthese **118**, 439–478 (1999)

2.35 J. Earman, J. Roberts, S. Smith: Ceteris paribus lost, Erkenntnis **57**, 281–301 (2002)

2.36 P.K. Feyerabend: Problems of empiricism. In: *Beyond the Edge of Certainty*, ed. by R.G. Colodny (Prentice-Hall, New Jersey 1965) pp. 145–260

2.37 E. Nagel: *The Structure of Science* (Hackett Publishing, Indianapolis 1979)

2.38 P.K. Feyerabend: Explanation, reduction and empiricism. In: *Minnesota Studies in the Philosophy of Science: Scientific Explanation, Space and Time*, Vol. 3, ed. by H. Feigl, G. Maxwell (Univ. Minnesota Press, Minneapolis 1962) pp. 28–97

2.39 P.K. Feyerabend: How to be a good empiricist – A plea for tolerance in matters epistemological. In: *Philosophy of Science: The Delaware Seminar*, Vol. 2, ed. by B. Baumrin (Interscience, New York 1963) pp. 3–39

2.40 P.K. Feyerabend: Problems of empiricism, Part II. In: *The Nature and Function of Scientific Theories*, ed. by R.G. Colodny (Univ. Pittsburgh Press, Pittsburgh 1970) pp. 275–353

2.41 P. Suppes: *Introduction to Logic* (Van Nostrand, New York 1957)

2.42 P. Suppes: A Comparison of the meaning and uses of models in mathematics and the empirical sciences. In: *The Concept and the Role of the Model in Mathematics and the Natural and Social Sciences*, ed. by H. Freudenthal (Reidel, Dordrecht 1961) pp. 163–177

2.43 P. Suppes: *Set-Theoretical Structures in Science* (Stanford Univ., Stanford 1967), mimeographed lecture notes

2.44 P. Suppes: *Representation and Invariance of Scientific Structures* (CSLI Publications, Stanford 2002)

2.45 N.C.A. Da Costa, S. French: The model-theoretic approach in the philosophy of science, Philos. Sci. **57**, 248–265 (1990)

2.46 N.C.A. Da Costa, S. French: *Science and Partial Truth, a Unitary Approach to Models and Scientific Reasoning* (Oxford Univ. Press, Oxford 2003)

2.47 B.C. Van Fraassen: *The Scientific Image* (Oxford Univ. Press, Oxford 1980)

2.48 B.C. Van Fraassen: On the extension of beth's semantics of physical theories, Philos. Sci. **37**, 325–339 (1970)

2.49 M. Friedman: Review of Bas C. van Fraassen: The scientific image, J. Philos. **79**, 274–283 (1982)

2.50 J. Worrall: Review article: An unreal image, Br. J. Philos. Sci. **35**, 65–80 (1984)

2.51 B.C. Van Fraassen: *An Introduction to the Philosophy of Time and Space*, 2nd edn. (Columbia Univ. Press, New York 1985)

2.52 B.C. Van Fraassen: The semantic approach to scientific theories. In: *The Process of Science*, ed. by N.J. Nersessian (Martinus Nijhoff, Dordrecht 1987) pp. 105–124

2.53 M. Thompson-Jones: Models and the semantic view, Philos. Sci. **73**, 524–535 (2006)

2.54 B.C. Van Fraassen: *Laws and Symmetry* (Oxford Univ. Press, Oxford 1989)

2.55 R.N. Giere: *Explaining Science: A Cognitive Approach* (The Univ. Chicago Press, Chicago 1988)

2.56 S. French: The structure of theories. In: *The Routledge Companion to the Philosophy of Science*, ed. by S. Psillos, M. Curd (Routledge, London 2008) pp. 269–280

2.57 M. Suarez: Scientific representation: Against similarity and isomorphism, Int. Stud. Philos. Sci. **17**(3), 225–244 (2003)

2.58 S.M. Downes: The importance of models in theorising: A deflationary semantic view, PSA 1992, Vol. 1, ed. by D. Hull, M. Forbes, K. Okruhlik (Philosophy of Science Associaion, Chicago 1992) pp. 142–153

2.59 D. Portides: Scientific models and the semantic view of scientific theories, Philos. Sci. **72**(5), 1287–1298 (2005)

2.60 R. Frigg: Scientific representation and the semantic view of theories, Theoria **55**, 49–65 (2006)

2.61 N.D. Cartwright: *The Dappled World: A Study of the Boundaries of Science* (Cambridge Univ. Press, Cambridge 1999)

2.62 M.C. Morrison: Models as autonomous agents. In: *Models as Mediators*, ed. by M.S. Morgan, M. Morrison (Cambridge Univ. Press, Cambridge 1999) pp. 38–65

2.63 M.C. Morrison: Modelling nature: Between physics and the physical world, Philos. Naturalis **35**, 65–85 (1998)

2.64 M.C. Morrison: Where have all the theories gone?, Philos. Sci. **74**, 195–228 (2007)

2.65 D. Portides: Models. In: *The Routledge Companion to the Philosophy of Science*, ed. by S. Psillos, M. Curd (Routledge, London 2008) pp. 385–395

2.66 N.D. Cartwright, T. Shomar, M. Suarez: The tool-box of science. In: *Theories and Models In Scientific Processes*, Poznan Studies, Vol. 44, ed. by E. Herfel, W. Krajewski, I. Niiniluoto, R. Wojcicki (Rodopi, Amsterdam 1995) pp. 137–149

2.67 D. Portides: Seeking representations of phenomena: Phenomenological models, Stud. Hist. Philos. Sci. **42**, 334–341 (2011)

2.68 N.D. Cartwright: Models and the limits of theory: Quantum hamiltonians and the BCS models of superconductivity. In: *Models as Mediators*, ed. by M.S. Morgan, M. Morrison (Cambridge Univ. Press, Cambridge 1999) pp. 241–281

2.69 M.S. Morgan, M. Morrison (Eds.): *Models as Mediators: Perspectives on Natural and Social Science* (Cambridge Univ. Press, Cambridge 1999)

2.70 M. Suarez: The role of models in the application of scientific theories: Epistemological implications. In: *Models as Mediators: Perspectives on Natural and Social Science*, ed. by M.S. Morgan, M. Morrison (Cambridge Univ. Press, Cambridge 1999) pp. 168–196

2.71 S. Hartman: Models and stories in hadron physics. In: *Models as Mediators: Perspectives on Natural and Social Science*, ed. by M.S. Morgan, M. Morrison (Cambridge Univ. Press, Cambridge 1999) pp. 326–346

2.72 R. Giere: *Science Without Laws* (Univ. Chicago Press, Chicago 1999)

2.73 R. Giere: *Scientific Perspectivism* (Univ. Chicago Press, Chicago 2006)

3. Models and Representation

Roman Frigg, James Nguyen

Models are of central importance in many scientific contexts. We study models and thereby discover features of the phenomena they stand for. For this to be possible models must be representations: they can instruct us about the nature of reality only if they represent the selected parts or aspects of the world we investigate. This raises an important question: In virtue of what do scientific models represent their target systems? In this chapter we first disentangle five separate questions associated with scientific representation and offer five conditions of adequacy that any successful answer to these questions must meet. We then review the main contemporary accounts of scientific representation – similarity, isomorphism, inferentialist, and fictionalist accounts – through the lens of these questions. We discuss each of their attributes and highlight the problems they face. We finally outline our own preferred account, and suggest that it provides the most promising way of addressing the questions raised at the beginning of the chapter.

Models play a central role in contemporary science. Scientists construct models of atoms, elementary particles, polymers, populations, genetic trees, economies, rational decisions, airplanes, earthquakes, forest fires, irrigation systems, and the world's climate – there is hardly a domain of inquiry without models. Models are essential for the acquisition and organization of scientific knowledge. We often study a model to discover features of the thing it stands for. How does this work? The answer is that a model can instruct us about the nature of its subject matter if it represents the selected part or aspect of the world that we investigate. So if we want to understand how models allow us to learn about the world, we have to come to understand how they represent.

The problem of representation has generated a sizable literature, which has been growing fast in particular over the last decade. The aim of this chapter is to review this body of work and assess the strengths and weaknesses of the different proposals. This enterprise faces an immediate difficulty: Even a cursory look at the literature on scientific representation quickly reveals that there is no such thing as *the* problem of scientific representation. In fact, we find a cluster of interrelated problems. In Sect. 3.1 we try to untangle this web and get clear on what the problems are and on how they relate to one another (for a historical introduction to the issue, see [3.1]). The result of this effort is a list with five problems and five conditions of adequacy, which provides the analytical lens through which we look at the different accounts. In Sect. 3.2 we discuss Griceanism and *stipulative fiat*. In Sect. 3.3 we look at the time-honored similarity approach, and in Sect. 3.4 we examine its modern-day cousin, the structuralist approach. In Sect. 3.5 we turn to inferentialism, a more recent family of conceptions. In Sect. 3.6 we discuss the fiction view of models, and in Sect. 3.7 we consider the conception of representation-as.

Before delving into the discussion, a number of caveats are in order. The first is that our discussion in no way presupposes that models are the sole unit of scientific representation, or that all scientific representation is model-based. Various types of images have their place in science, and so do graphs, diagrams, and drawings (*Perini* [3.2–4] and *Elkins* [3.5] provide discussions of visual representation in the sciences). In some contexts scientists use what *Warmbrōd* [3.6] calls *natural forms of representation* and what *Peirce* [3.7] would have classified as indices: tree rings, fingerprints, disease symptoms. These are related to thermometer readings and litmus paper indications, which are commonly classified as measurements. Measurements also provide representations of processes in nature, sometimes together with the subsequent condensation of measurement results in the form of charts, curves, tables and the like (*Tal* [3.8] provides a discussion of measurement). And, last but not least, many would hold that theories represent too. At this point the vexing problem of the nature of theories and the relation between theories and models rears is head again. We refer the reader to *Portides*' contribution to this volume, Chap. 2, for a discussion of this issue. Whether these other forms of scientific representation have features in common with how models represent is an interesting question, but this is a problem for another day. Our aim here is a more modest one: to understand how models represent. To make the scope of our investigation explicit we call the kind of representation we are interested in *model-representation*.

The second point to emphasize is that our discussion is not premised on the claim that *all* models are representational; nor does it assume that representation is the only (or even primary) function of models. It has been emphasized variously that models perform a number of functions other than representation. To mention but few: *Knuuttila* [3.9, 10] 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 that allow us to gather knowledge in diverse ways; *Morgan* and *Morrison* [3.11] emphasize the role models play in the mediation between theories and the world; *Hartmann* [3.12] discusses models as tools for theory construction; *Peschard* [3.13] investigates the way in which models may be used to construct other models and generate new target systems; and *Bokulich* [3.14] and *Kennedy* [3.15] present nonrepresentational accounts of model explanation (*Woody* [3.16] and *Reiss* [3.17] provide general discussions of the relation between representation and explanation). Not only do we not see projects like these as being in conflict with a view that sees some models as representational; we think that the approaches are in fact complementary.

Finally, there is a popular myth according to which a representation is a mirror image, a copy, or an imitation of the thing it represents. In this view representation is ipso facto realistic representation. This is a mistake. Representations can be realistic, but they need not. And representations certainly need not be copies of the real thing. This, we take it, is the moral of the satire about the cartographers who produce maps as large as the country itself only to see them abandoned. The story has been told by Lewis Carroll in *Sylvie and Bruno* and Jorge Luis Borges in *On Exactitude in Science*. Throughout this review we encounter positions that make room for nonrealistic representation and hence testify to the fact that representation is a much broader notion than mirroring.

There is, however, a sense in which we presuppose a minimal form of realism. Throughout the discussion we assume that target systems exist independently of human observers, and that they are how they are irrespective of what anybody thinks about them. That is, we assume that the targets of representation exist independently of the representation. This is a presupposition not everybody would share. Constructivists (and other kinds of metaphysical antirealists) assume that there is no phenomenon independent of its representation: representations constitute the phenomena they represent (this view is expounded for instance by *Lynch* and *Wooglar* [3.18]; *Giere* [3.19] offers a critical discussion). It goes without saying that an assessment of the constructivist program is beyond the scope of this review. It is worth observing, though, that many of the discussions to follow are by no means pointless from a constructivist perspective. What in the realist idiom is conceptualized as the representation of an object in the world by a model would, from the constructivist perspective, turn into the study of the relation between a model and another representation, or an object constituted by another representation. This is because even from a constructivist perspective, models and their targets are not identical, and the fact that targets are representationally constituted would not obliterate the differences between a target representation and scientific model.

3.1 Problems Concerning Model-Representation

In this section we say what questions a philosophical account of model-representation has to answer and reflect on what conditions such an answer has to satisfy. As one would expect, different authors have framed the problem in different ways. Nevertheless, recent discussions about model-representation have tended to cluster around a relatively well-circumscribed set of issues. The aim of this section is to make these issues explicit and formulate five problems that an account of model-representation has to answer. These problems will help us in structuring the discussion in later sections and put views and positions into perspective. In the course of doing so we also articulate five conditions of adequacy that every account of model-representation has to satisfy.

Models are representations of a selected part or aspect of the world. This is the model's *target system*. The first and most fundamental question about a model therefore is: In virtue of what is a model a representation of something else? Attention has been drawn to this issue by *Frigg* ([3.20, p. 17], [3.21, p. 50]), *Morrison* [3.22, p. 70], and *Suárez* [3.23, p. 230]. To appreciate the thrust of this question it is instructive to briefly ponder the same problem in the context of pictorial representation. When seeing, say, Soutine's *The Groom or the Bellboy* we immediately realize that it depicts a man in a red dress. Why is this? Per se the painting is a plane surface covered with pigments. How does an arrangement of pigments on a surface represent something outside the picture frame? Likewise, models, before being representations of atoms, populations, or economies, are equations, structures, fictional scenarios, or mannerly physical objects. The problem is: what turns equations and structures, or fictional scenarios and physical objects into representations of something beyond themselves? It has become customary to phrase this problem in terms of necessary and sufficient conditions and throughout this review we shall follow suit (some may balk at this, but it's worth flagging that the standard arguments against such an analysis, e.g., those surveyed in *Laurence* and *Margolis* [3.24], lose much of their bite when attention is restricted to core cases as we do here). The question then is: What fills the blank in *M is a model-representation of T iff* ____, where *M* stands for model and *T* for target system?

To spare ourselves difficulties further down the line, this formulation needs to be adjusted in light of a crucial condition of adequacy that any account of model-representation has to meet. The condition is that models represent in a way that allows us to form hypotheses about their target systems. We can generate claims about a target system by investigating a model that represents it. Many investigations are carried out on models rather than on reality itself, and this is done with the aim of discovering features of the things models stands for. Every acceptable theory of scientific representation has to account for how reasoning conducted on models can yield claims about their target systems. Let us call this the *surrogative reasoning condition*.

The term *surrogative reasoning* was introduced by *Swoyer* [3.25, p. 449], and there seems to be widespread agreement on this point (although *Callender* and *Cohen* [3.26], whose views are discussed in Sect. 3.3, provide a noteworthy exception). To mention just a few writers on the subject: *Bailer-Jones* [3.27, p. 59] emphasizes that models "*tell us* something about certain features of the world" (original emphasis). *Boliskna* [3.28] and *Contessa* [3.29] both call models *epistemic representations*; *Frigg* ([3.21, p. 51], [3.30, p. 104]) sees the potential for learning as an essential explanandum for any theory of representation; *Liu* [3.31, p. 93] emphasizes that the main role for models in science and technology is epistemic; *Morgan* and *Morrison* [3.11, p. 11] regard models as *investigative tools*; *Suárez* ([3.23, p. 229], [3.32, p. 772]) submits that models license specific inferences about their targets; and *Weisberg* [3.33, p. 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". This distinguishes models from lexicographical representations such as words. Studying the internal constitution of a model can provide information about the target. Not so with words. The properties of a word (consisting of so and so many letters and syllables, occupying this or that position in a dictionary, etc.) do not matter to its functioning as a word; and neither do the physical properties of the ink used to print words on a piece of paper. We can replace one word by another at will (which is what happens in translations from one language to another), and we can print words with other methods than ink on paper. This is possible because the properties of a word as an object do not matter to its semantic function.

This gives rise to a problem for the schema *M is a model-representation of T iff* ____. The problem is that any account of representation that fills the blank in a way that satisfies the *surrogative reasoning condition* will almost invariably end up covering other kinds of representations too. Geographical maps, graphs, diagrams, charts, drawings, pictures, and photographs often provide epistemic access to features of the items they represent, and hence are likely to fall under an account of representation that explains this sort of reasoning. This is a problem for an analysis of model-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 of an account of model-representation, then one either has to conclude that the account fails because it does not provide necessary conditions, or that first impressions are wrong and other representations (such as maps or photographs) are in fact model-representations.

Neither of these options is appealing. To avoid this problem we follow a suggestion of *Contessa*'s [3.29] and broaden the scope of the investigation. Rather than analyzing the relatively narrow category of model-representation, we analyze the broader category of *epistemic representation*. This category comprises model-representations, but it also includes other representations that allow for surrogative reasoning. The task then becomes to fill the blank in *M is an epistemic representation of T iff* ____. For brevity we use $R(M,T)$ as a stand in for *M is an epistemic representation of T*, and so the biconditional becomes $R(M,T)$ *iff* ____. We call the general problem of figuring out in virtue of what something is an epistemic representation of something else the *epistemic representation problem* (*ER-problem*, for short), and the above biconditional the *ER-scheme*. So one can say that the ER is to fill the blank in the ER-scheme. *Frigg* [3.21, p. 50] calls this the "enigma of representation" and in *Suárez*'s [3.23, p. 230] terminology this amounts to identifying the *constituents* of a representation (although he questions whether both necessary *and* sufficient conditions can be given; see Sect. 3.5 for further discussion on how his views fit into the ER-framework).

Analyzing the larger category of epistemic representation and placing model-representations in that category can be seen as giving rise to a demarcation problem for scientific representations: How do scientific model-representations differ from other kinds of epistemic representations? We refer to this question as the *representational demarcation problem*. *Callender* and *Cohen* [3.26, p. 69] formulate this problem, but then voice skepticism about our ability to solve it [3.26, p. 83]. The representational demarcation problem has received little, if any, attention in the recent literature on scientific representation, which would suggest that other authors either share Callender and Cohen's skepticism, or regard it as a nonissue to begin with. The latter seems to be implicit in approaches that discuss scientific representation alongside pictorial representation such as *Elgin* [3.34], *French* [3.35], *Frigg* [3.21], *Suárez* [3.32], and *van Fraassen* [3.36]. But a dismissal of the problem is in no way a neutral stance. It amounts to no less than the admission that model-representations are not fundamentally different from other epistemic representations, or that we are unable to pin down what the distinguishing features are. Such a stance should be made explicit and, ideally, justified.

Two qualifications concerning the ER-scheme need to be added. The first concerns its flexibility. Some might worry that posing the problem in this way prejudges what answers can be given. The worry comes in a number of variants. A first variant is that the scheme presupposes that representation is an intrinsic relation between M and T (i. e., a relation that only depends on intrinsic properties of M and T and on how they relate to one another rather than on how they relate to other objects) or even that it is naturalisable (a notion further discussed in Sect. 3.3). This is not so. In fact, R might depend on any number of factors other than M and T themselves, and on ones that do not qualify as natural ones. To make this explicit we write the ER-scheme in the form $R(M,T)$ *iff* $C(M,T,x_1,\dots,x_{\mathrm{n}})$, where n is a natural number and C is an $(n+2)$-ary relation that grounds representation. The x_i can be anything that is deemed relevant to epistemic representation, for instance a user's intentions, standards of accuracy, and specific purposes. We call C the *grounding relation* of an epistemic representation.

Before adding a second qualification, let us introduce the next problem in connection with model-representation. Even if we restrict our attention to scientific epistemic representations (if they are found to be relevantly different to nonscientific epistemic representations as per the demarcation problem above), not all representations are of the same kind. In the case of visual representations this is so obvious that it hardly needs mention: An Egyptian mural, a two-point perspective ink drawing, a pointillist oil painting, an architectural plan, and a road map represent their respective targets in different ways. This pluralism is not limited to visual representations. Model-representations do not all seem to be of the same kind either. *Woody* [3.37] argues that chemistry as a discipline has its own ways to represent molecules. But differences in style can also appear in models from the same discipline. Weizsäcker's liquid drop model represents the nucleus of an atom in a manner that seems to be different from the one of the shell model. A scale model of the wing of a plane represents the wing in a way that is different from how a mathematical model of its cross section does. Or Phillips and Newlyn's famous hydraulic machine and Hicks' mathematical models both represent a Keynesian economy but they seem to do so in different ways. This gives rise to the question: What styles are there and how can they be characterized? This is the *problem of style* [3.21, p. 50]. There is no expectation that a *complete* list of styles be provided in response. Indeed, it is unlikely that such a list can ever be drawn up, and new styles will be invented as science progresses. For this reason a response to the problem of style will always be open-ended, providing a taxonomy of what is currently available while leaving room for later additions.

With this in mind we can now turn to the second qualification concerning the ER-scheme. The worry is this: The scheme seems to assume that representation is a monolithic concept and thereby make it impossible to distinguish between different kinds of representation. The impression is engendered by the fact the scheme asks us to fill a blank, and blank is filled only once. But if there are different kinds of representations, we should be able to fill the blank in different ways on different occasions because a theory of representation should not force upon us the view that the different styles are all variations of one overarching concept of representation.

The ER-scheme is more flexible than it appears at first sight. There are at least three ways in which different styles of representations can be accommodated. For the sake of illustration, and to add some palpability to an abstract discussion, let us assume that we have identified two styles: analogue representation and idealized representation. The result of an analysis of these relations is the identification of their respective grounding relations. Let $C_A(M, T, \ldots)$ and $C_I(M, T, \ldots)$ be these relations. The first way of accommodating them in the ER-scheme is to fill the blank with the disjunction of the two: $R(M, T)$ *iff* $C_A(M, T, \ldots)$ *or* $C_I(M, T, \ldots)$. In plain English: M represents T if and only if M is an analogue representation of T or M is an idealized representation of T. This move is possible because, first appearances notwithstanding, nothing hangs on the grounding relation being homogeneous. The relation can be as complicated as we like and there is no prohibition against disjunctions. In the above case we have $C = [C_A \text{ or } C_I]$. Furthermore, the grounding relation could even be an open disjunction. This would help accommodating the above observation that a list of styles is potentially open-ended. In that case there would be a grounding relation for each style and the scheme could be written as $R(M, T)$ *iff* $C_1(M, T \ldots)$ *or* $C_2(M, T \ldots)$ *or* $C_3(M, T \ldots)$ *or* $\ldots$, where the C_i are the grounding relations for different styles. This is not a new scheme; it's the old scheme where $C = [C_1 \text{ or } C_2 \text{ or } C_3 \text{ or } \ldots]$ is spelled out.

Alternatively one could formulate a different scheme for every kind of representation. This would amount to changing the scheme slightly in that one does not analyze epistemic representation per se. Instead one would analyze different kinds of epistemic representations. Consider the above example again. Let $R_1(M, T)$ stand for *M is an analogue epistemic representation of T* and $R_2(M, T)$ for *M is an idealized epistemic representation of T*. The response to the ER-problem then consists in presenting the two biconditionals $R_1(M, T)$ *iff* C_A and $R_2(M, T)$ *iff* C_I. This generalizes straightforwardly to the case of any number of styles, and the open-endedness of the list of styles can be reflected in the fact that an open-ended list of conditionals of the form $R_i(M, T)$ *iff* C_i can be given, where the index ranges over styles.

In contrast with the second option, which pulls in the direction of more diversity, the third aims for more unity. The crucial observation here is that the grounding relation can in principle be an abstract relation that can be concretized in different ways, or a determinable that can have different determinates. On the third view, then, the concept of representation is like the concept of force (which is abstract in that in a concrete situation force is gravity or electromagnetic attraction or some other specific force), or like color (where a colored object must be blue or green or ____). This view would leave $R(M, T)$ *iff* $C(M, T, x_1, \ldots, x_n)$ unchanged and take it as understood that C is an abstract relation.

At this point we do not adjudicate between these options. Each has its own pros and cons, and which one is the most convenient to work with depends on one's other philosophical commitments. What matters is that the ER-scheme does have the flexibility to accommodate different representational styles, and that it can in fact accommodate them in at least three different ways.

The next problem in line for the theory of model-representation is to specify standards of accuracy. Some representations are accurate; others aren't. The Schrödinger model is an accurate representation of the hydrogen atom; the Thomson model isn't. On what grounds do we make such judgments? In *Morrison*'s words: "how do we identify what constitutes a accurate representation?" [3.22, p. 70]. We call this the problem of *standards of accuracy*. Answering this question might make reference to the purposes of the model and model user, and thus it is important to note that by *accuracy* we mean something that can come in degrees and may be context dependent. Providing a response to the problem of accuracy is a crucial aspect of an account of epistemic representation.

This problem goes hand in hand with a second condition of adequacy: the *possibility of misrepresentation*. Asking what makes a representation an accurate representation already presupposes that inaccurate representations are representations too. And this is how it should be. If M does not accurately portray T, then it is a misrepresentation but not a nonrepresentation. It is therefore a general constraint on a theory of epistemic representation that it has to make misrepresentation possible. This can be motivated by a brief glance at the history of science, but is plausibly also part of the concept of representation, and as such is found in discussions of other kinds of representation (*Stitch* and *Warfield* [3.38, pp. 6–7], for instance, suggest that a theory of mental representation should be able to account for misrepresentation, as do *Sterelny* and *Griffiths* [3.39, p. 104] in their discussion of genetic representation). A corollary of this requirement is that representation is a wider concept than accurate representation and that representation cannot be analyzed in terms of accurate representation.

A related condition concerns models that misrepresent in the sense that they lack target systems. Models of ether, phlogiston, four-sex populations, and so on, are all deemed scientific models, but ether, phlogiston, and four-sex populations don't exist. Such models lack (actual) target systems, and one hopes that an account of epistemic representation would allow us to understand how these models work. We call this the problem of targetless models (or models without targets).

The fourth condition of adequacy for an account of model-representation is that it must account for the directionality of representation. Models are about their targets, but (at least in general) targets are not about their models. So there is an essential directionality to representations, and an account of model-representation has to identify the root of this directionality. We call this the *requirement of directionality*.

Many scientific models are highly mathematized, and their mathematical aspects are crucial to their cognitive as well as their representational function. This forces us to reconsider a time-honored philosophical puzzle: the applicability of mathematics in the empirical sciences. Even though the problem can be traced back at least to Plato's *Timaeus*, its canonical modern expression is due to *Wigner*, who famously remarked that "the enormous usefulness of mathematics in the natural sciences is something bordering on the mysterious and that there is no explanation for it" [3.40, p. 2]. One need not go as far as seeing the applicability of mathematics as an inexplicable miracle, but the question remains: How does mathematics hook onto the world?

The recent discussion of this problem has taken place in a body of literature that grew out of the philosophy of mathematics (see *Shapiro* [3.41, Chap. 8] for a review). But, with the exception of *Bueno* and *Colyvan* [3.42], there has been little contact with the literature on scientific modeling. This is a regrettable state of affairs. The question of how a mathematized model represents its target implies the question of how mathematics applies to a physical system. So rather than separating the question of model-representation from the problem of the applicability of mathematics and dealing with them in separate discussions, they should be seen as the two sides of the same coin and be dealt with in tandem. For this reason, our fifth and final condition of adequacy is that an account of representation has to explain how mathematics is applied to the physical world. We call this the *applicability of mathematics condition*.

In answering the above questions one invariably runs up against a further problem, the *problem of ontology*: What kinds of objects are models? Are they structures in the sense of set theory, fictional entities, descriptions, equations or yet something else? Or are there no models at all? While some authors develop an ontology of models, others reject an understanding of models as *things* and push a program that can be summed up in the slogan *modeling without models* [3.43]. There is also no presupposition that all models be of the same kind. Some models are material objects, some are things that one holds in one's head rather than one's hands (to use *Hacking*'s phrase [3.44, p. 216]). For the most part, the focus in debates about representation has been on nonmaterial models, and we will follow this convention. It is worth emphasizing, however, that also the seemingly straightforward material models raise interesting philosophical questions: *Rosenblueth* and *Wiener* [3.45] discuss the criteria for choosing an object as a model; *Ankeny* and *Leonelli* [3.46] discuss issues that arise when using organisms as models; and the contributors to [3.47] discuss representation in the laboratory.

A theory of representation can recognize different kinds of models, or indeed no models at all. The requirement only asks us to be clear on our commitments and provide a list with things, if any, that we recognize as models and give an account of what they are in case these entities raise questions (what exactly do we mean by something that one holds in one's head rather than one's hands?).

In sum, an account of model-representation has to do the following:

1. Provide an answer to the *epistemic representation problem* (filling the blank in ER-scheme: *M is an epistemic representation of T iff* ...).
2. Take a stand on the *representational demarcation problem* (the question of how scientific epistemic

representations differ from other kinds of epistemic representations).
3. Respond to the *problem of style* (what styles are there and how can they be characterized?).
4. Formulate *standards of accuracy* (how do we identify what constitutes an accurate representation?).
5. Address the *problem of ontology* (what kinds of objects are models?).

Any satisfactory answer to these five issues will have to meet the following five conditions of adequacy:

1. *Surrogative reasoning condition* (models represent their targets in a way that allows us to generate hypotheses about them).
2. *Possibility of misrepresentation* (if M does not accurately represent T, then it is a misrepresentation but not a nonrepresentation).
3. *Targetless models* (what are we to make of scientific representations that lack targets?).
4. *Requirement of directionality* (models are about their targets, but targets are not about their models).
5. *Applicability of mathematics condition* (how the mathematical apparatus used in M latches onto the physical world).

To frame the problem in this way is not to say that these are separate and unrelated issues, which can be dealt with one after the other in roughly the same way in which we first buy a ticket, walk to the platform and then take a train. This division is analytical, not factual. It serves to structure the discussion and to assess proposals; it does not imply that an answer to one of these questions can be dissociated from what stance we take on the other issues.

3.2 General Griceanism and Stipulative Fiat

Callender and *Cohen* [3.26] submit that the entire debate over scientific representation has started on the wrong foot. They claim that scientific representation is not different from "artistic, linguistic, and culinary representation" and in fact "there is no special problem about scientific representation" [3.26, p. 67]. Underlying this claim is a position Callender and Cohen call *General Griceanism* (GG). The core of GG is the reductive claim that most representations we encounter are "derivative from the representational status of a privileged core of representations" [3.26, p. 70]. GG then comes with a practical prescription about how to proceed with the analysis of a representation [3.26, p. 73]:

> "The General Gricean view consists of two stages. First, it explains the representational powers of derivative representations in terms of those of fundamental representations; second, it offers some other story to explain representation for the fundamental bearers of content."

Of these stages only the second requires serious philosophical work, and this work is done in the philosophy of mind because the fundamental form of representation is mental representation.

Scientific representation is a derivative kind of representation [3.26, pp. 71,75] and hence falls under the first stage of the above recipe. It is reduced to mental representation by an act of stipulation [3.26, pp. 73–74]:

> "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."

So explaining any form of representation other than mental representation is a triviality – all it takes is an act of "stipulative fiat" [3.26, p. 75]. This supplies their answer to the ER-problem:

Definition 3.1 Stipulative fiat
A scientific model M represents a target system T iff a model user stipulates that M represents T.

On this view, scientific representations are cheap to come by. The question therefore arises why scientists spend a lot of time constructing and studying complex models if they might just as well take a salt shaker and turn it into a representation of, say, a Bose–Einstein condensate by an act of fiat. *Callender* and *Cohen* admit that there are useful and not so useful representations, and that salt shakers belong the latter group. However, they insist that this has nothing to do with representation [3.26, p. 75]:

> "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."

So, in sum, scientific representation [3.26, p. 78]

> "is constituted in terms of a stipulation, together with an underlying theory of representation for mental states, isomorphism, similarity, and inference generation are all idle wheels."

The first question we are faced with when assessing this account is the relation between GG and stipulative fiat (Definition 3.1). Callender and Cohen do not comment on this issue, but that they mention both in the same breath would suggest that they regard them as one and the same doctrine, or at least as the two sides of the same coin. This is not so. Stipulative fiat (Definition 3.1) is just one way of fleshing out GG, which only requires that there be *some* explanation of how derivative representations relate to fundamental representations; GG does not require that this explanation be of a particular kind, much less that it consists of nothing but an act of stipulation ([3.48, pp. 77–78], [3.49, p. 244]). Even if GG is correct, it doesn't follow that stipulative fiat is a satisfactory answer to the ER-problem. Model-representation can, in principle, be *reduced* to fundamental representation in many different ways (some of which we will encounter later in this chapter). Conversely, the failure of stipulate fiat does not entail that we must reject GG: one can uphold the idea that an appeal to the intentions of model users is a crucial element in an account of scientific representation even if one dismisses stipulative fiat (Definition 3.1).

Let us now examine stipulative fiat (Definition 3.1). Callender and Cohen emphasize that anything can be a representation of anything else [3.26, p. 73]. This is correct. Things that function as models don't belong to a distinctive ontological category, 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* [3.50, p. 99], *Giere* [3.51, p. 269], *Suárez* [3.32, p. 773], *Swoyer* [3.25, p. 452], and *Teller* [3.52, p. 397]) and, as we shall see, it is a cornerstone of several alternative accounts of representation.

But just because anything can, in principle, be a representation of anything else, it doesn't follow that a mere act of stipulation suffices to turn M into a representation of T. Furthermore, it doesn't follow that an object elevated to the status of a representation by an act of fiat represents its target in a way that can appropriately be characterized as an instance of epistemic representation. We discuss both concerns in reverse order.

Stipulative fiat (Definition 3.1) fails to meet the surrogative reasoning condition: it fails to provide an account of how claims about Madagascar could be extracted from reasoning about the salt shaker. Even if we admit that stipulative fiat (Definition 3.1) establishes that models denote their targets (and as we will see soon, there is a question about this), denotation is not sufficient for epistemic representation. Both the word *Napoleon* and Jacques-Louis David's portrait of Napoleon serve to denote the French general. But this does not imply that they represent him in the same way, as noted by *Toon* [3.48, pp. 78–79]. *Bueno* and *French* [3.53, pp. 871–874] 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.

Supporters of stipulative fiat (Definition 3.1) 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 Madagascar 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. Second, the last quotation from Callender and Cohen suggests that they might want to relegate surrogative reasoning into the realm of pragmatics and deny that it is part of the relation properly called epistemic representation. This, however, in effect amounts to a removal of the surrogative reasoning condition from the desiderata of an account of scientific representation, and we have argued in Sect. 3.1 that surrogative reasoning is one of the hallmarks of scientific representation. And even if it were *pragmatics*, we still would want an account of how it works.

Let us now turn to our first point, that a mere act of stipulation is insufficient to turn M into a representation of T. We take our cue from a parallel discussion in the philosophy of language, where it has been pointed out that it is not clear that stipulation is sufficient to establish a denotational relationship (which is weaker than epistemic representation). A position similar to stipulative fiat (Definition 3.1) faces what is known as the *Humpty Dumpty problem*, named in reference to Lewis Carroll's discussion of Humpty using the word *glory* to mean *a nice knockdown argument* [3.54, 55] (it's worth noting that this debate concerns meaning, rather than denotation, but it's plausible that it can be reconstructed in terms of the latter). 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 rather than proper names, and work in that field focuses on propping up existing accounts so as to ensure that a speaker's intentions successfully establish the denotation of demonstratives uttered by the speaker [3.56]. Whatever the success of these endeavors, their mere existence shows that successfully establishing denotation requires moving beyond a bare appeal to stipulation, or brute intention. But if a brute appeal to intentions fails in the case of demonstratives – the sorts of terms that such an account would most readily be applicable to – then we find it difficult to see how stipulative fiat (Definition 3.1) will establish a representational relationship between models and their targets. Moreover, this whole discussion supposed that an intention-based account of denotation is the correct one. This is controversial – see *Reimer* and *Michaelson* [3.57] for an overview of discussions of denotation in the philosophy of language. If this is not the correct way to think about denotation, then stipulative fiat (Definition 3.1) will fail to get off the ground at all.

It now pays that we have separated GG from stipulative fiat (Definition 3.1). Even though stipulative fiat (Definition 3.1) does not provide an adequate answer to the ER-problem, one can still uphold GG. As *Callender* and *Cohen* note, all that it requires is that there is a privileged class of representations (they take them to be mental states but are open to the suggestion that they might be something else [3.26, p. 82]), and that other types of representations owe their representational capacities to their relationship with the primitive ones. So philosophers need an account of how members of this privileged class of representations represent, and how derivative representations, which includes scientific models, relate to this class.

This is a plausible position, and when stated like this, many recent contributors to the debate on scientific representation can be seen as falling under the umbrella of GG. As we will see below, the more developed versions of the similarity (Sect. 3.3) and isomorphism (Sect. 3.4) accounts of scientific representation make explicit reference to the intentions and purposes of model users, even if their earlier iterations did not. And so do the accounts discussed in the latter sections, where the intentions of model users (in a more complicated manner than that suggested by stipulative fiat (Definition 3.1)) are invoked to establish epistemic representation.

3.3 The Similarity Conception

Moving on from the Gricean account we now turn to the similarity conception of scientific representation (in aesthetics the term *resemblance* is used more commonly than *similarity*, but there does not seem to be a substantive difference between the notions, and we use the terms as synonyms throughout). Similarity and representation initially appear to be two closely related concepts, and invoking the former to ground the latter has a philosophical lineage stretching back at least as far as Plato's *The Republic*.

In its most basic guise the similarity conception of scientific representation asserts that scientific models represent their targets in virtue of being similar to them. This conception has universal aspirations in that it is taken to account for epistemic representation across a broad range of different domains. Paintings, statues, and drawings are said to represent by being similar to their subjects, (see *Abell* [3.58] and *Lopes* [3.59] for relatively current discussions of similarity in the context of visual representation). And recently *Giere*, one of the view's leading contemporary proponents, proclaimed that it covers scientific models alongside "words, equations, diagrams, graphs, photographs, and, increasingly, computer-generated images" [3.60, p. 243] (see also *Giere* [3.61, p. 272], and for further discussion *Toon* [3.49, pp. 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. (Sometimes the similarity view is introduced by categorizing models as icons in Peirce's sense, and, as *Kralemann* and *Lattmann* point out, icons represent "on the basis of a similarity relation between themselves and their objects" [3.62, p. 3398].)

The view also offers an elegant account of surrogative reasoning. Similarities between model and 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 (however, it is worth noting that this kind of knowledge transfer can cause difficulties in some contexts, *Frigg* et al. [3.63] discuss these difficulties in the context of nonlinear dynamic modeling).

However, appeal to similarity in the context of representation leaves open whether similarity is offered as an answer to the ER-problem, the problem of style, or whether it is meant to set standards of accuracy. Proponents of the similarity account typically have offered little guidance on this issue. So we examine each option in turn and ask whether similarity offers a viable answer. We then turn to the question of how the similarity view deals with the problem of ontology.

3.3.1 Similarity and ER-Problem

Understood as response to the ER-problem, a similarity view of representation amounts to the following:

Definition 3.2 Similarity 1
A scientific model M represents a target T iff M and T are similar.

A well-known objection to this account is that similarity has the wrong logical properties. *Goodman* [3.64, pp. 4–5] submits that similarity is symmetric and reflexive yet representation isn't. If object A is similar to object B, then B is similar to A. But if A represents B, then B need not (and in fact in most cases does not) represent A: the Newtonian model represents the solar system, but the solar system does not represent the Newtonian model. And everything is similar to itself, but most things do not represent themselves. So this account does not meet our third condition of adequacy for an account of scientific representation insofar as it does not provide a direction to representation. (Similar problems also arise in connection with other logical properties, e.g., transitivity; see *Frigg* [3.30, p. 31] and *Suárez* [3.23, pp. 232–233].)

Yaghmaie [3.65] argues that this conclusion – along with the third condition itself – is wrong: epistemic representation is symmetric and reflexive (he discusses this in the context of the isomorphism view of representation, which we turn to in the next section, but the point applies here as well). His examples are drawn from mathematical physics, and he presents a detailed case study of a symmetric representation relation between quantum field theory and statistical mechanics. His case raises interesting questions, but even if one grants that Yaghmaie has identified a case where representation is reflexive and symmetrical it does not follow that representation *in general* is. The photograph in Jane's passport represents Jane; but Jane does not represent her passport photograph; and the same holds true for myriads of other representations. Goodman is correct in pointing out that typically representation is not symmetrical and reflexive: a target T does not represent model M just because M represents T.

A reply diametrically opposed to Yaghmaie's emerges from the writings of Tversky and Weisberg. They accept that representation is not symmetric, but dispute that similarity fails on this count. Using a gradual notion of similarity (i. e., one that allows for statements like *A is similar to B to degree d*), *Tversky* found that subjects in empirical studies judged that North Korea was more similar to China than China was to North Korea [3.66]; similarly *Poznic* [3.67, Sect. 4.2] points out with reference to the characters in a Polanski movie that the similarity relation between a baby and the father need not be symmetric.

So allowing degrees into ones notion of similarity makes room for an asymmetry (although degrees by themselves are not sufficient for asymmetry; metric-based notions are still symmetric). This raises the question of how to analyze similarity. We discuss this thorny issue in some detail in the next subsection. For now we concede the point and grant that similarity need not always be symmetrical. However, this does not solve Goodman's problem with reflexivity (as we will see on Weisberg's notion of similarity everything is maximally similar to itself); nor does it, as will see now, solve other problems of the similarity account.

However the issue of logical properties is resolved, there is another serious problem: similarity is too inclusive a concept to account for representation. In many cases neither one of a pair of similar objects represents the other. Two copies of the same book are similar but neither represents the other. Similarity between two items is not enough to establish the requisite relationship of representation; there are many cases of similarity where no representation is involved. 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.

This point has been brought home in a now-classical thought experiment due to *Putnam* [3.68, pp. 1–3] (but see also *Black* [3.69, p. 104]). An ant is crawling on a patch of sand and leaves a trace that happens to resemble Winston Churchill. Has the ant produced a picture of Churchill? Putnam's answer is that it didn't because the ant has never seen Churchill and had no intention to produce 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, shows that "[s]imilarity [...] to the features of Winston Churchill is not sufficient to make something represent or refer to Churchill" [3.68, p. 1]. And what is true of the trace and Churchill is true of every other pair of similar items: similarity on its own does not establish representation.

There is also a more general issue concerning similarity: it is too easy to come by. Without constraints on what counts as similar, any two things can be considered similar to any degree [3.70, p. 21]. This, however, has the unfortunate consequence that anything represents anything else because any two objects are similar in some respect. Similarity is just too inclusive 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. This suggestion has been made explicitly by *Giere* [3.71, p. 81] who suggests that models come equipped with what he calls *theoretical hypotheses*, statements asserting that model and target are similar in relevant respects and to certain degrees. This idea can be molded into the following definition:

Definition 3.3 Similarity 2

A scientific model M represents a target T iff M and T are similar in relevant respects and to the relevant degrees.

On this definition one is free to choose one's respects and degrees so that unwanted similarities drop out of the picture. While this solves the last problem, it leaves the others untouched: similarity in relevant respects and to the relevant degrees is reflexive (and symmetrical, depending on one's notion of similarity); and presumably the ant's trace in the sand is still similar to Churchill in the relevant respects and degrees but without representing Churchill. Moreover, similarity 2 (Definition 3.3) introduces three new problems.

First, a misrepresentation is one that portrays its target as having properties that are not similar in the relevant respects and to the relevant degrees to the true properties of the target. But then, on similarity 2 (Definition 3.3), M is not a representation at all. *Ducheyne* [3.72] embraces this conclusion when he offers a variant of a similarity account that explicitly takes the *success* of the hypothesized similarity between a model and its target to be a necessary condition on the model representing the target. In Sect. 3.2 we argued that the possibility of misrepresentation is a condition of adequacy for any acceptable account of representation and so we submit that misrepresentation should not be conflated with nonrepresentation ([3.20, p. 16], [3.23, p. 235]).

Second, similarity in relevant respects and to the relevant degrees does not guarantee that M represents the right target. As *Suárez* points out [3.23, pp. 233–234], even a regimented similarity can obtain with no corresponding representation. If John dresses up as Pope Innocent X (and he does so perfectly), then he resembles Velázquez's portrait of the pope (at least in as far as the pope himself resembled the portrait). 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. Like in the case of Putnam's ant, the root cause of the problem is that the similarity is accidental. In the case of the ant, the accident occurs at the representation end of the relation, whereas in the case of John's dressing up the accidental similarity occurs at the target end. Both cases demonstrate that similarity 2 (Definition 3.3) cannot rule out accidental representation.

Third, there may simply be nothing to be similar to because some representations represent no actual object [3.64, p. 26]. Some paintings represent elves and dragons, and some models represent phlogiston and the ether. None of these exist. As *Toon* points out, this is a problem in particular for the similarity view [3.49, pp. 246–247]: models without objects cannot represent what they seem to represent because in order for two things to be similar to each other both have to exist. If there is no ether, then an ether model cannot be similar to the ether.

It would seem that at least the second problem could be solved by adding the requirement that M denote T (as considered, but not endorsed, by *Goodman* [3.64, pp. 5–6]). Amending the previous definition accordingly yields:

Definition 3.4 Similarity 3

A scientific model M represents a target T iff M and T are similar in relevant respects and to the relevant degrees and M denotes T.

This account would also solve the problem with reflexivity (and symmetry), because denotation is directional in a way similarity is not. Unfortunately similarity 3 (Definition 3.4) still suffers from the first and the third problems. It would still lead to the conflation of misrepresentatios with nonrepresentations because the first conjunct (similar in the relevant respects) would still be false. And a nonexistent system cannot be denoted and so we have to conclude that models of, say, the ether and phlogiston represent nothing. This seems an unfortunate consequence because there is a clear sense in which models without targets are about something. Maxwell's writings on the ether provide a detailed and intelligible account of a number of properties of the ether, and these properties are highlighted in the model. If ether existed then similarity 3 (Definition 3.4) could explain why these were important by appealing to them as being relevant for the similarity between an ether model and its target. But since ether does not, no such explanation is offered.

A different version of the similarity view sets aside the moves made in similarity 3 (Definition 3.4) and tries to improve on similarity 2 (Definition 3.3). The crucial move is to take the very act of *asserting* a specific similarity between a model and a target as constitutive of the scientific representation.

Definition 3.5 Similarity 4

A scientific model M represents a target system T if and only if a theoretical hypotheses H asserts that M and T are similar in certain respects and to certain degrees.

This comes close to the view *Giere* advocated in *Explaining Science* [3.71, p. 81] (something like this is also found in *Cartwright* ([3.73, pp. 192–193], [3.74, pp. 261–262]) who appeals to a "loose notion of resemblance"; her account of modeling is discussed in more detail in Sect. 3.6.3). This version of the similarity view avoids problems with misrepresentation because, being hypotheses, there is no expectation that the assertions made in H are true. If they are, then the representation is accurate (or the representation is accurate to the extent that they hold). If they are not, then the representation is a misrepresentation. It resolves the problem of mistargeting because hypotheses identify targets before asserting similarities with M (that is, the task of picking the right target is now placed in the court of the hypothesis and is no longer expected to be determined by the similarity relation). Finally it also resolves the issue with directionality because H can be understood as introducing a directionality that is not present in the similarity relation. However, it fails to resolve the problem with representation without a target. If there is no ether, no hypotheses can be asserted about it.

Let us set the issue of nonexistent targets aside for the moment and have a closer look at the notion of representation proposed in similarity 4 (Definition 3.5). A crucial point remains understated in similarity 4 (Definition 3.5). Hypotheses don't assert themselves; hypotheses are put forward by those who work with representations, in the case of models, scientists. So the crucial ingredient – users – is left implicit in similarity 4 (Definition 3.5).

In a string of recent publications *Giere* made explicit the fact that "scientists are intentional agents with goals and purposes" [3.60, p. 743] and proposed to build this insight explicitly into an account of epistemic representation. This involves adopting an agent-based notion of representation that focuses on "the activity of representing" [3.60, p. 743]. Analyzing epistemic representation in these terms amounts to analyzing schemes like "S uses X to represent W for purposes P" [3.60, p. 743], or in more detail [3.51, p. 274]:

> "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."

This conception of representation had already been proposed half a century earlier by *Apostel* when he urged the following analysis of model-representation [3.75, p. 4]:

> "Let then $R(S,P,M,T)$ indicate the main variables of the modeling relationship. The subject S takes, in view of the purpose P, the entity M as a model for the prototype T."

Including the intentions of model agents in the definition of scientific representation is now widely accepted, as we discuss in more detail in Sect. 3.4 (although *Rusanen* and *Lappi* disagree with this, and claim that "the semantics of models as scientific representations should be based on the mind-independent model-world relation" [3.76, p. 317]).

Giere's proposal, in our own terminology comes down to:

Definition 3.6 Similarity 5

A scientific model M represents a target system T iff there is an agent A who uses M to represent a target system T by proposing a theoretical hypothesis H specifying a similarity (in certain respects and to certain degrees) between M and T for purpose P.

This definition inherits from similarity 4 (Definition 3.5) the resolutions of the problems of directionality, misrepresentation, and mistargeting; and for the sake of argument we assume that the problem with nonexistent targets can be resolved in one way or other.

A crucial thing to note about similarity 5 (Definition 3.6) is that, by invoking an active role for the purposes and actions of scientists in constituting epistemic representation, it marks a significant change in emphasis for similarity-based accounts. *Suárez* [3.23, pp. 226–227], drawing on *van Fraassen* [3.77] and *Putnam* [3.78], 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 personal purposes, views or interests of enquirers". By building the purposes of model users directly into an answer to the ER-problem, similarity 5 (Definition 3.6) is explicitly not a naturalistic account (in contrast, for example, to similarity 1 (Definition 3.2)). As noted in Sect. 3.2 we do not demand a naturalistic account of model-representation (and as we will see later, many of the more developed answers to the ER-problem are also not naturalistic accounts).

Does this suggest that similarity 5 (Definition 3.6) is a successful similarity-based solution to the ER-problem? Unfortunately not. A closer look at similarity 5 (Definition 3.6) 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 5 (Definition 3.6) is done by the appeal to agents and similarity has in fact become an idle wheel. *Giere* implicitly admits this when he writes [3.60, p. 747]:

> "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."

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. Being similar in the relevant respects to the relevant degrees now plays the role either of a representational style, or of a normative criterion for accurate representation, rather than of a grounding of representation. We assess in the next section whether similarity offers a cogent reply to the issues of style and accuracy.

A further problem is that there seems to be a hidden circularity in the analysis. As *Toon* [3.49, pp. 251–252] points out, having a scientist form a theoretical hypothesis about the similarity relation between two objects A and B and exploit this similarity for a certain purpose P is not sufficient for representation. A and B could be two cars in a showroom and an engineer inspects car A and then use her knowledge about similarities to make assertions about B (for instance if both cars are of the same brand she can infer something about B's quality of manufacturing). This, Toon submits, is not a case of representation: neither car is representational. Yet, if we delete the expression *to represent* on the right hand side of the biconditional in similarity 5 (Definition 3.6), the resulting condition provides an accurate description of what happens in the showroom. So the only difference between the nonrepresentational activity of comparing cars and representing B by A is that in one case A is *used to represent* and in the other it's only *used*. So representation is explained in terms of *to represent*, which is circular. So similarity 5 (Definition 3.6) does not provide nontrivial conditions for something to be used *as a representation*.

One way around the problem would be to replace *to represent* by *to denote*. This, however, would bring the account close to similarity 3 (Definition 3.4), and it would suffer from the same problems.

Mäki [3.79] suggested an extension of similarity 5 (Definition 3.6), which he explicitly brands as "a (more or less explicit) version" of Giere's. *Mäki* adds two conditions to Giere's: the agent uses the model to address an audience E and adds a commentary C [3.79, p. 57]. The role of the commentary is to specify the nature of the similarity. This is needed because [3.79, p. 57]:

> "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."

What these relevant respects and degrees of resemblance are depends on the purposes of the scientific representation in question. These are not determined *in the model* as it were, but are pragmatic elements. 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.

The addition of an audience, however, is problematic. While models are often shared publicly, this does not seem to be a necessary condition for the representational use of a model. There is nothing that precludes a lone scientist from coining a model M and using it representationally. That some models are easier to grasp, and therefore serve as more effective tools to drive home a point in certain public settings, is an indisputable fact, but one that has no bearing on a model's status as a representation. The pragmatics of communication and the semantics of modeling are separate issues.

The conclusion we draw from this discussion is that similarity does not offer a viable answer to the ER-problem.

3.3.2 Accuracy and Style

Accounting for the possibility of misrepresentation resulted in a shift of the division of labor for the more developed similarity-based accounts. Rather than being the relation that grounds representation, similarity should be considered as setting a standard of accuracy or as providing an answer to the question of style (or both). The former is motivated by the observation that a proposed similarity between M and T could be wrong, and hence if the model user's proposal does in fact hold (and M and T are in fact similar in the specified way) then M is an accurate representation of T. The latter transpires from the simple observation that a judgment of accuracy in fact presupposes a choice of respects in which M and T are claimed to be similar. Simply proposing that they are similar in some unspecified respect is vacuous. But delineating relevant properties could potentially provide an answer to the problem of style. For example, if M and T are proposed to be similar with respect to their causal structure, then we might have a style of causal modeling; if M and T are proposed to be similar with respect to structural properties, then we might have a style of structural modeling; and so on and so forth. So the idea is that if M representing T involves the claim that M and T are similar in a certain respect, the respect chosen specifies the style of the representation; and if M and T are in fact similar in that respect (and to the specified degree), then M accurately represents T within that style.

In this section we investigate both options. But before delving into the details, let us briefly step back and reflect on possible constraints on viable answers. Taking his cue from *Lopes*' [3.59] discussion of pictures, *Downes* [3.80, pp. 421–422] proposes two constraints on allowable notions of similarity. The first, which he calls the *independence challenge*, requires that a user must be able to specify the relevant representation-grounding similarity *before* engaging in a comparison between M and T. Similarities that are recognizable only with hindsight are an unsound foundation of a representation. We agree with this requirement, which in fact is also a consequence of the surrogative reasoning condition: a model can generate novel hypotheses only if (at least some of the) similarity claims are not known only ex post facto.

Downes' second constraint, the *diversity constraint*, is the requirement that the relevant notion of similarity has to be identical in all kinds of representation and across all representational styles. So all models must bear the same similarity relations to their targets. Whatever its merits in the case of pictorial representation, this observation does not hold water in the case of scientific representation. Both Giere and Teller have insisted – rightly, in our view – that there need not be a substantive sense of similarity uniting all representations (see also *Callender* and *Cohen* [3.26, p. 77] for a discussion). A proponent of the similarity view is free to propose different kinds of similarity for different representations and is under no obligation to also show that they are special cases of some overarching conception of similarity.

We now turn to the issue of style. A first step in the direction of an understanding of styles is the explicit analysis of the notion of similarity. 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 instead rather than on similarity in respect and degrees; for a critical discussion see [3.81]. 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* [3.82, pp. 117–118] and *Goodman* [3.83, p. 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 formal accounts to capture it directly in more fully worked out accounts. The two most prominent suggestions are the *geometric* and *contrast* accounts (see [3.84] for an up-to-date discussion). The former, associated with *Shepard* [3.85], assigns objects a place in a multidimensional space based on values assigned to their properties. This space is then equipped with a metric and the degree of similarity between two objects is a function of the distance between the points representing the two objects in that space.

This account is based on the strong assumptions that values can be assigned to all features relevant to similarity judgments, which is deemed unrealistic. This problem is supposed to be overcome in *Tversky*'s *contrast account* [3.86]. This account defines a gradated notion of similarity based on a weighted comparison of properties. *Weisberg* ([3.33, Chap. 8], [3.87]) has recently introduced this 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*. This account is our primary interest here.

The account introduces a set Δ of relevant properties. Let then $\Delta_M \subseteq \Delta$ be the set of properties from Δ that are instantiated by the model M; likewise Δ_T is the set of properties from Δ instantiated by the target system. Furthermore let f be a ranking function assigning a real number to every subset of Δ. 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 [3.87, p. 788] (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 value. This equation provides "a similarity score that can be used in comparative judgments of similarity" [3.87, p. 788]. The score is determined by weighing the properties the model and target have in common against those they do not. (Thus we note that this account could be seen as a quantitative version of *Hesse*'s [3.88] 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*.) 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). One can then use standard mathematical techniques to renormalize S so that it takes values in the unit interval [0, 1] (these technical moves need not occupy us here and we refer the reader to *Weisberg* for details [3.33, Chap. 8]).

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" [3.33, p. 149]. Likewise, the ranking function as well as the values of weighting parameters depend on the goals of the investigation, the context, and the theoretical framework in which the scientists operate. Weisberg 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" [3.33, p. 145]. This distinction is helpful in the application to concrete cases, but for the purpose of our conceptual discussion it can be set aside.

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 in Sect. 3.3.1. Second, S has a property called *maximality*: everything is maximally similar to itself and every other nonidentical object is equally or less similar. Formally: $S(A,A) \geq S(A,B)$ for all objects A and B as long as $A \neq B$ [3.33, p. 154].

What does this account contribute to a response to the question of style? The answer, we think, is that it has heuristic value but does not provide substantive account. In fact, stylistic questions stand outside the proposed framework. The framework can be useful in bringing questions into focus, but eventually the substantive stylistic questions concern inclusion criteria for Δ (what properties do we focus on?), the weight given by f to properties (what is the relative importance of properties?) and the value of the parameters (how significant are disagreements between the properties of M and T?). These questions have to be answered outside the account. The account is a framework in which questions can be asked but which does not itself provide answers, and hence no classification of representational styles emerges from it.

Some will say that this is old news. *Goodman* denounced similarity as "a pretender, an impostor, a quack" [3.83, p. 437] not least because he thought that it merely put a label to something unknown without analyzing it. And even some proponents of the similarity view have insisted that no general characterization of similarity was possible. Thus *Teller* submits that [3.52, p. 402]:

> "[t]here can be no general account of similarity, but there is also no need for a general account because the details of any case will provide the information which will establish just what should count as relevant similarity in that case."

This amounts to nothing less than the admission that no analysis of similarity (or even different kinds of similarity) is possible and that we have to deal with each case in its own right.

Assume now, for the sake of argument, that the stylistic issues have been resolved and full specifications of relevant properties and their relative weights are available. It would then seem plausible to say that $S(M,T)$ provides a degree of accuracy. This reading is supported by the fact that *Weisberg* paraphrases the role of $S(M,T)$ as providing "standards of fidelity" [3.33, p. 147]. Indeed, in response to *Parker* [3.89], Weisberg claims that his weighted feature matching account is supposed to answer the ER-problem *and* provide standards of accuracy.

As we have seen above, $S(M,T)$ is maximal if M is a perfect replica of T (with respect to the properties in Δ), and the fewer properties M and T share, the less accurate the representation becomes. This lack of accuracy is then reflected in a lower similarity score. This is plausible and Weisberg's account is indeed a step forward in the direction of quantifying accuracy.

Weisberg's account is an elaborate version of the co-instantiation account of similarity. It improves significantly on simple versions, but it cannot overcome that account's basic limitations. *Niiniluoto* distinguishes between two different kinds of similarities [3.90, pp. 272–274]: partial identity and likeness (which also feature in *Hesse*'s discussion of analogies, see, for instance [3.88, pp. 66–67]). Assume M instantiates the relevant properties $P_1, \ldots, P_n$ and T instantiates the relevant properties $Q_1, \ldots, Q_n$. If these properties are identical, i. e., if $P_i = Q_i$ for all $i = 1, \ldots, n$, then M and T are similar in the sense of being *partially identical*. Partial identity contrasts with what Niiniluoto calls *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 $i = 1, \ldots, n$. 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 color, even if they do not instantiate the exact same shade of red. As *Parker* [3.89, p. 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 which encompass the P_i and the Q_i. *Parker* [3.89] suggests that this can be done by introducing intervals in the feature set, for instance of the form "the value of feature X lies in the interval $[x-\varepsilon, x+\varepsilon]$" where ε is a parameter specifying the precision of overlap. To illustrate she uses Weisberg's example of the 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 has to claim something like [3.89, p. 273]:

> "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."

In his response *Weisberg* 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 [3.91, p. 302]. But such interval-valued properties have to be part of 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 has to delineate relevant properties, but, at least for the quantitative ones, also has 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 [3.33, p. 136].

3.3.3 Problems of Ontology

Another problem facing similarity-based approaches concerns their treatment of the ontology of models. 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 homely physical objects. The Army Corps of Engineers' model of the San Francisco bay is a water basin equipped with pumps to simulate the action of tidal flows [3.33]; ball and stick models of molecules are made of metal or wood [3.92]; the Phillips–Newlyn model of an economy is system of pipes and reservoirs [3.93]; and model organisms in biology are animals like worms and mice [3.46]. 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 that 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* aptly describes as "something you hold in your head rather than your hands" [3.44, p. 216]. Following *Thomson-Jones* [3.94] we call such models *nonconcrete models*. The question then is what kind of objects nonconcrete models are. *Giere* submits that they are abstract objects ([3.60, p. 747], cf. [3.51, p. 270], [3.71, p. 81]):

> "Models in advanced sciences such as physics and biology should be abstract objects constructed in conformity with appropriate general principles and specific conditions."

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 [3.95], and *Hale* [3.96, pp. 86–87] lists no less than 12 different possible characterizations 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. *Giere* could counter that he limits attention to those abstract objects that possess "all and only the characteristics specified in the principles" [3.60, p. 745], where principles are general rules like Newton's laws of motion. He 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" [3.60, p. 747]. While this narrows down the choices somehow, 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 skepticism about the view that philosophers of science "need a deeper understanding of imaginative processes and of the objects produced by these process" [3.97, p. 250] or that "we need say much more [...] to get on with the job of investigating the functions of models in science" [3.97].

We remain unconvinced about this skepticism, 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 models are *abstract* in the sense that they have no spatiotemporal location. *Teller* [3.52, p. 399] and *Thomson-Jones* [3.98] supply arguments suggesting that 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 idealized pendulum model makes this clear. If the idealized 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* [3.52] is that we should instead say that whilst "concrete objects HAVE properties [...] properties are PARTS of models" [3.52, p. 399] (original capitalization). It is not entirely clear what Teller means by this, but our 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 ones account of instantiation, of properties and other abstract objects, and of the way properties enter into models" [3.52].

But as *Thompson-Jones* [3.98, pp. 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 [3.99]. One might speculate that addressing these issues would push Teller either towards the kind of more robust account of abstract objects that he endeavored to avoid, or towards a fictionalist understanding of models.

The latter option has been discussed by Giere, who points out that a natural response to Teller's and 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 an imaginary detective, and smoking is a physical property. At times, Giere seems sympathetic to this view. He says [3.97, p. 249]:

> "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."

And he observes that [3.51, p. 278]:

> "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."

However, these seem to be occasional slips and he recently positioned himself as an outspoken opponent of any approach to models that likens them to literary fiction. We discuss these approaches as well as Giere's criticisms of them in Sect. 3.6.

In sum, the similarity view is yet to be equipped with a satisfactory account of the ontology of models.

3.4 The Structuralist Conception

The structuralist conception of model-representation originated in the so-called semantic view of theories that came to prominence in the second half of the 20th century (*Suppes* [3.100], *van Fraassen* [3.101], and *Da Costa* and *French* [3.102] provide classical statements of the view; *Byerly* [3.103], *Chakravartty* [3.104], *Klein* [3.105] and *Portides* [3.106, 107] provide critical discussions). The semantic view was originally proposed as an account of theory structure rather than model-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? Defenders of the semantic view of theories take models to be structures, which represent their target systems in virtue of there being some kind of *mapping* (isomorphism, partial isomorphism, homomorphism, ...) between the two. (It is worth noting that Giere, whose account of scientific representation we discussed in the previous section, is also associated with the semantic view, despite not subscribing to either of these positions.)

This conception has two prima facie advantages. The first advantage is that it offers a straightforward answer to the ER-problem, and one that accounts 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. There is time-honored position in the philosophy of mathematics that sees mathematics as the study of structures; see, for instance, *Resnik* [3.108] and *Shapiro* [3.109]. It is a natural move for the scientific structuralist to adopt this point of view, which, without further ado, provides a neat explanation of how mathematics is used in scientific modeling.

3.4.1 Structures and the Problem of Ontology

Almost anything from a concert hall to a kinship system 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* [3.110]), but by far the most common and widely used is the notion of structure one finds in set theory and mathematical logic. A structure S in that sense (sometimes *mathematical structure* or *set-theoretic structure*) is a composite entity consisting of the following: a nonempty set U of objects called the domain (or universe) of the structure and a nonempty indexed set R of relations on U. With the exception of the caveat below regarding interpretation functions, this definition of structure is widely used in mathematics and logic; see for instance *Machover* [3.111, p. 149], *Hodges* [3.112, p. 2], and *Rickart* [3.113, p. 17]. It is convenient to write these as $S = \langle U, R \rangle$, where $\langle\, , \rangle$ denotes an ordered tuple. Sometimes operations are also included in the definition of a structure. While convenient in some applications, operations are redundant because operations reduce to relations (see *Boolos* and *Jeffrey* [3.114, pp. 98–99]).

It is important to be clear on what we mean by *object* and *relation* in this context. As *Russell* [3.115, p. 60] points out, in defining the domain of a structure it is irrelevant what the objects are. All that matters from a structuralist point of view is that there are so and so many of them. Whether the object is a desk or a planet is irrelevant. All we need are dummies or placeholders whose *only* property is *objecthood*. Similarly, when defining relations one disregards completely what the relation is *in itself*. Whether we talk about *being the mother of* or *standing to the left of* is of no concern in the context of a structure; all that matters is between which objects it holds. For this reason, a relation is specified purely extensionally: as a class of ordered n-tuples. The relation literally is nothing over and above this class. So a structure consists of dummy objects between which purely extensionally defined relations hold.

Let us illustrate this with an example. Consider the structure with the domain $U = \{a, b, c\}$ and the following two relations: $r_1 = \{a\}$ and $r_2 = \{\langle a, b \rangle, \langle b, c \rangle, \langle a, c \rangle\}$. Hence R consists of r_1 and r_2, and the structure itself is $S = \langle U, R \rangle$. This is a structure with a three-object domain endowed with a monadic property and a transitive relation. Whether the objects are books or iron rods is of no relevance to the structure; they could be literally anything one can think of. Likewise r_1 could be literally any monadic property (being green, being waterproof, etc.) and r_2 could be any (irreflexive) transitive relation (larger than, hotter than, more expensive than, etc.).

It is worth pointing out that this use of *structure* differs from the use one sometimes finds in logic, where linguistic elements are considered part of the model as well. Specifically, over and above $S = \langle U, R\rangle$, a structure is also taken to include a language (sometimes called a *signature*) L, and an interpretation function ([3.112, Chap. 1] and [3.116, pp. 80–81]). But in the context of the accounts discussed in this section, a structure is the ordered pair $S = \langle U, R\rangle$ as introduced above and so we disregard this alternative use of *structure*.

The first basic posit of the structuralist theory of representation is that models are structures in the above sense (the second is that models represent their targets by being suitably morphic to them; we discuss morphisms in the next subsection). *Suppes* has articulated this stance clearly when he declared that "the meaning of the concept of model is the same in mathematics and the empirical sciences" [3.117, p. 12]. Likewise, *van Fraassen* posits that a "scientific theory gives us a family of models to represent the phenomena", that "[t]hese models are mathematical entities, so all they have is structure [. . .]" [3.118, pp. 528–529] and that therefore [3.118, p. 516]

> "[s]cience is [. . .] interpreted as saying that the entities stand in relations which are transitive, reflexive, etc. but as giving no further clue as to what those relations are."

Redhead submits that "it is this abstract structure associated with physical reality that science aims, and to some extent succeeds, to uncover [. . .]" [3.119, p. 75]. Finally, *French* and *Ladyman* affirm that "the specific material of the models is irrelevant; rather it is the structural representation [. . .] which is important" [3.120, p. 109]. Further explicit statements of this view are offered by: *Da Costa* and *French* [3.121, p. 249], *Suppes* ([3.122, p. 24], [3.123, Chap. 2]) and *van Fraassen* ([3.101, pp. 43, 64], [3.118, pp. 516, 522], [3.124, p. 483], [3.125, p. 6]).

These structuralist accounts have typically been proposed in the framework of the so-called semantic view of theories. There are differences between them, and formulations vary from author to author. However, as *Da Costa* and *French* [3.126] point out, all these accounts share a commitment to analyzing models as structures. So we are presented with a clear answer to the problem of ontology: models are structures. The remaining issue is what structures themselves are. Are they platonic entities, equivalence classes, modal constructs, or yet something else? This is a hotly debated issue in the philosophy of logic and mathematics; for different positions see for instance *Dummett* [3.127, 295ff.], *Hellman* [3.128, 129], *Redhead* [3.119], *Resnik* [3.108], and *Shapiro* [3.109]. But philosophers of science need not resolve this issue and can pass off the burden of explanation to philosophers of mathematics. This is what usually happens, and hence we don't pursue this matter further.

An extension of the standard conception of structure is the so-called partial structures approach (for instance, *Da Costa* and *French* [3.102] and *Bueno* et al. [3.130]). Above we defined relations by specifying between which tuples it holds. This naturally allows a sorting of all 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 tuples it is indeterminate whether or not they belong to the relation. Such a relation is a *partial relation*. A structure with a set R containing partial relations is a *partial structure* (formal definitions can be found in references given above). Partial structures make room for a process of scientific investigation where one begins not knowing whether a tuple falls under the relation and then learns whether or not it does.

Proponents of that approach are more guarded as regards the ontology of models. *Bueno* and *French* emphasize that "advocates of the semantic account need not be committed to the ontological claim that models *are* structures" [3.53, p. 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* [3.131] 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 models is just one way of representing a theory; we can also use an intrinsic characterization and represent the same theory as a set of sentences in order to account for how they can be objects of our epistemic attitudes [3.132].

He therefore adopts a quietist position with respect to what a theory or a model *is*, declining to answer the question [3.131, 133]. There are thus two important notions of representation at play: representation of targets by models, which is the job of scientists, and representation of theories and models by structures, which is the job of philosophers of science. The question for this approach then becomes whether or not the structuralist representation of models and epistemic representation – as partial structures and morphisms that hold between them – is an accurate or useful one. And the concerns raised below remain when translated into this context as well.

There is an additional question regarding the correct formal framework for thinking about models in the structuralist position. *Landry* [3.134] argues that in certain contexts group, rather than set, theory should be used when talking about structures and morphisms between them, and *Halvorson* [3.135, 136] argues that theories should be identified with categories rather than classes or sets. Although these discussions highlight important questions regarding the nature of scientific theories, the question of how individual models represent remains unchanged. Halvorson still takes individual models to be set-theoretic structures. And Landry's paper is not an attempt to reframe the representational relationship between models and their targets (see [3.137] for her skepticism regarding how structuralism deals with this question). Thus, for reasons of simplicity we will focus on the structuralist view that identifies models with set-theoretic structures throughout the rest of this section.

3.4.2 Structuralism and the ER-Problem

The most basic structuralist conception of scientific representation asserts that scientific models, understood as structures, represent their target systems in virtue of being isomorphic to them. Two structures $S_a = \langle U_a, R_a \rangle$ and $S_b = \langle U_b, R_b \rangle$ are isomorphic iff there is a mapping $f : U_a \rightarrow U_b$ such that (i) f is one-to-one (bijective) and (ii) f preserves the system of relations in the following sense: The members $a_1, \ldots, a_n$ of U_a satisfy the relation r_a of R_a iff the corresponding members $b_1 = f(a_1), \ldots, b_n = f(a_n)$ of U_b satisfy the relation r_b of R_b, where r_b is the relation corresponding to r_a (for difficulties in how to cash out this notion of correspondence without reference to an interpretation function see *Halvorson* [3.135] and *Glymour* [3.138]).

Assume now that the target system T exhibits the structure $S_T = \langle U_T, R_T \rangle$ and the model is the structure $S_M = \langle U_M, R_M \rangle$. Then the model represents the target iff it is isomorphic to the target:

Definition 3.7 Structuralism 1

A scientific model M represents its target T iff S_M is isomorphic to S_T.

This view is articulated explicitly by *Ubbink*, who posits that [3.139, p. 302]

> "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."

Views similar to Ubbink's seem operable in many versions of the semantic view. In fairness to proponents of the semantic view it ought to be pointed out, though, that for a long time representation was not the focus of attention in the view and the attribution of (something like) structuralism 1 (Definition 3.7) to the semantic view is an extrapolation. 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 (Definition 3.7), or pointed out that they never held such a view. We turn to more advanced positions shortly, but to understand what motivates such positions it is helpful to understand why structuralism 1 (Definition 3.7) fails.

An immediate question concerns the target end structure S_T. At least prima facie target systems aren't structures; they are physical objects like planets, molecules, bacteria, tectonic plates, and populations of organisms. An early recognition that the relation between targets and structures is not straightforward can be found in *Byerly*, who emphasizes that structures are abstracted from objects [3.103, pp. 135–138]. The relation between structures and physical targets is indeed a serious question and we will return to it in Sect. 3.4.4. In this subsection we grant the structuralist the assumption that target systems are (or at least have) structures.

The first and most obvious problem is the same as with the similarity view: isomorphism is symmetrical, reflexive, and transitive, but epistemic representation isn't. This problem could be addressed by replacing isomorphism with an alternative mapping. *Bartels* [3.140], *Lloyd* [3.141], and *Mundy* [3.142] suggest homomorphism; *van Fraassen* [3.36, 101, 118] and *Redhead* isomorphic embeddings [3.119]; advocates of the partial structures approach prefer partial isomophisms [3.102, 120, 121, 143–145]; and *Swoyer* [3.25] introduces what he calls $\Delta/\Psi-$ morphisms. We refer to these collectively as *morphisms*.

This solves some, but not all problems. While many of these mappings are asymmetrical, they are all still reflexive, and at least some of them are also transitive. But even if these formal issues could be resolved in one way or another, a view based on structural mappings would still face other serious problems. For ease of presentation we discuss these problems in the context of the isomorphism view; mutatis mutandis other formal mappings suffer from the same difficulties (For detailed discussions of homomorphism and partial isomorphism see *Suárez* [3.23, pp. 239-241] and *Pero* and *Suárez* [3.146]; *Mundy* [3.142] discusses general constraints one may want to impose on morphisms.)

Like similarity, isomorphism is too inclusive: not all things that are isomorphic 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. Many mathematical structures have been discovered and discussed long before they have been used in science. Non-Euclidean geometries were studied by mathematicians long 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 was nothing over and above isomorphism, then we would have to conclude that Riemann discovered general relativity or that that Hilbert invented quantum mechanics. This is obviously wrong. Isomorphism on its own does not establish representation [3.20, p. 10].

Isomorphism is more restrictive than similarity: not everything is isomorphic to everything else. But isomorphism is still too abundant to correctly identify the extension of a representation (i. e., the class of systems it represents), which gives rise to a version of the mistargeting problem. The root of the difficulties is that the same structures can be instantiated in different target systems. The $1/r^2$ law of Newtonian gravity is also the *mathematical skeleton* of Coulomb's law of electrostatic attraction and the weakening of sound or light as a function of the distance to the source. The mathematical structure of the pendulum is also the structure of an electric circuit with condenser and solenoid (a detailed discussion of this case is provided by *Kroes* [3.147]). Linear equations are ubiquitous in physics, economics and psychology. Certain geometrical structures are instantiated by many different systems; just think about how many spherical things we find in the world. This shows that the same structure can be exhibited by more than one target system. Borrowing a term from the philosophy of mind, one can say that structures are *multiply realizable*. If representation is explicated solely in terms of isomorphism, then we have to conclude that, say, a model of a pendulum also represents an electric circuit. But this seems wrong. Hence isomorphism is too inclusive to correctly identify a representation's extension.

One might try to dismiss this point as an artifact of a misidentification of the target. *Van Fraassen* [3.101, p. 66], mentions a similar problem under the heading of "unintended realizations" and then expresses confidence that it will "disappear when we look at larger observable parts of the world". Even if there are multiply realizable structures to begin with, they vanish as science progresses and considers more complex systems because these systems are unlikely to have the same structure. Once we focus on a sufficiently large part of the world, no two phenomena will have the same structure. There is a problem with this counter, however. To appeal to *future* science to explain how models work *today* seems unconvincing. It is a matter of fact that we *currently* have models that represent electric circuits and sound waves, and we do not have to await future science providing us with more detailed accounts of a phenomenon to make our models represent what they actually already do represent.

As we have seen in the last section, a misrepresentation is one that portrays its target as having features it doesn't have. In the case of an isomorphism account of representation this presumably means that the model portrays 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. A misrepresentation won't be isomorphic to the target. By the lights of structuralism 1 (Definition 3.7) it is therefore is not a representation at all. Like simple similarity accounts, structuralism 1 (Definition 3.7) conflates misrepresentation with nonrepresentation.

Muller [3.148, p. 112] suggests that this problem can be overcome in a two-stage process: one first identifies a submodel of the model, which in fact is isomorphic to at least a part of the target. This *reduced* isomorphism establishes representation. One then constructs "a tailor-made morphism on a case by case basis" [3.148, p. 112] to account for accurate representation. *Muller* is explicit that this suggestion presupposes that there is "at least one resemblance" [3.148, p. 112] between model and target because "otherwise one would never be called a representation of the other" [3.148, p. 112]. While this may work in some cases, it is not a general solution. It is not clear whether all misrepresentations have isomorphic submodels. Models that are gross distortions of their targets (such as the liquid drop model of the nucleus or the logistic model of a population) may well not have such submodels. More generally, as *Muller* admits, his solution "precludes total misrepresentation" [3.148, p. 112]. So in effect it just limits the view that representation coincides with correct representation to a submodel. However, this is too restrictive a view of representation. Total misrepresentations may be useless, but they are representations nevertheless.

Another response refers to the partial structures approach and emphasizes that partial structures are in fact constructed to accommodate a mismatch between model and target and are therefore not open to this objection [3.53, p. 888]. It is true that the partial structures framework has a degree of flexibility that the standard view does not. However, we doubt that this flexibility stretches far enough. While the partial structure approach deals successfully with incomplete representations, it does not seem to deal well with distortive representations (we come back to this point in the next subsection). So the partial structures approach, while enjoying an advantage over the standard approach, is nevertheless not yet home and dry.

Like the similarity account, structuralism 1 (Definition 3.7) has a problem with nonexistent 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 target cannot represent what they seem to represent.

Most of these problems can be resolved by making moves similar to the ones that lead to similarity 5 (Definition 3.6): introduce agents and hypothetical reasoning into the account of representation. Going through the motions one finds:

Definition 3.8 Structuralism 2

A scientific model M represents a target system T iff there is an agent A who uses M to represent a target system T by proposing a theoretical hypothesis H specifying an isomorphism between S_M and S_T.

Something similar to this was suggested by *Adams* [3.149, p. 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 the *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" [3.36, p. 23]. Likewise, *Bueno* submits that "representation is an *intentional* act relating two objects" [3.150, p. 94] (original emphasis), and *Bueno* and *French* point 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" [3.53, p. 885]. This, of course, gives up on the idea of an account that 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 5 (Definition 3.6), structuralism 2 (Definition 3.8) accounts for the directionality of representation and has no problem with misrepresentation. But, again as in the case of similarity 5 (Definition 3.6), this is a Pyrrhic victory as the role of isomorphism has shifted. The crucial ingredient is the agent's intention and isomorphism has in fact become either a representational style or normative criterion for accurate representation. Let us now assess how well isomorphism fares as a response to these problems.

3.4.3 Accuracy, Style and Demarcation

The problem of style is to identify representational styles and characterize them. Isomorphism offers an obvious response to this challenge: one can represent a system by coming up with a model that is structurally isomorphic to it. We call this the isomorphism-style. This style also offers a clear-cut condition of accuracy: the representation is accurate if the hypothesized 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? The former is uncontentious. However, the emphasis many structuralists place on isomorphism suggests that they do not regard isomorphism as merely one way among others to represent something. What they seem to have in mind is the stronger claim that a representation *must* be of that sort, or that the isomorphism-style is the only acceptable style.

This claim seems to conflict with scientific practice. Many representations are inaccurate in some way. As we have seen above, partial structures are well equipped to deal with incomplete representations. However, not all inaccuracies are due to something being left out. Some models distort, deform and twist properties of the target in ways that seem to undercut isomorphism. Some models in statistical mechanics have an infinite number of particles and the Newtonian model of the solar system represents the sun as perfect sphere where it in reality is fiery ball with no well-defined surface at all. It is at best unclear how isomorphism, partial or otherwise, can account for these kinds of idealizations. From an isomorphism perspective all one can say about such idealizations is that they are failed isomorphism representations (or isomorphism misrepresentations). This is rather uninformative. One might try to characterize these idealizations by looking at *how* they fail to be isomorphic to their targets, but we doubt that this is going very far. Understanding how distortive idealizations work requires a positive characterization of them, and we cannot see how such a characterization could be given within the isomorphism framework. So one has to recognize styles of representation other than isomorphism.

This raises that question of whether other mappings such as homomorphisms or embeddings would fit the bill. They would, we think, make valuable additions to the list of styles, but they would not fill all gaps. Like isomorophism, these mappings are not designed to accommodate distortive idealizations, and hence a list of styles that includes them still remains incomplete.

Structuralism's stand on the demarcation problem is by and large an open question. Unlike similarity, which has been widely discussed across different domains, isomorphism is tied closely to the formal framework of set theory, and it has been discussed only sparingly outside the context of the mathematized sciences. An exception is *French*, who discusses isomorphism accounts in the context of pictorial representation [3.35]. He discusses in detail *Budd*'s [3.151] 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 [3.35, pp. 1475–1476] (this point is reaffirmed by *Bueno* and *French* [3.53, pp. 864–865]; see *Downes* [3.80, pp. 423–425] for a critical discussion). 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" [3.150, p. 94] – can be accommodated. This would suggest that 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 we will see 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 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. Nonmathematized sciences work with models that aren't structures. *Godfrey-Smith* [3.152], 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 modeling, 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 modeling (this is the thrust of many contributions to [3.11]; see also [3.73]). Although some models used by scientists may be best thought of as set theoretic structures, 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 San Francisco bay model [3.33], but also systems such as the idealized 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* [3.98] dubs this *face value practice*, and there is a question whether structuralism can account for that practice.

3.4.4 The Structure of Target Systems

Target systems are physical objects: atoms, planets, populations of rabbits, economic agents, etc. Isomorphism is a relation that holds between two structures and claiming that a set theoretic structure is isomorphic to a piece of the physical world is prima facie a category mistake. By definition, all of the mappings suggested – isomorphism, partial isomorphism, homomorphism, or isomorphic embedding – only hold between two structures. If we are to make sense of the claim that the model is isomorphic to its target we have to assume that the target somehow exhibits a certain structure $S_{\mathrm{T}} = \langle U_{\mathrm{T}}, R_{\mathrm{T}} \rangle$. But what does it mean for a target system – a part of the physical world – to possess a structure, and where in the target system is the structure located?

The two prominent suggestions in the literature are that data models are the target end structures represented by models, and that structures are, in some sense, instantiated in target systems. The latter option comes in three versions. The first version is that a structure is ascribed to a system; the second version is that systems instantiate structural universals; and the third version claims that target systems simply are structures. We consider all suggestions in turn.

What are data models? Data are what we gather in experiments. When observing the motion of the moon, for instance, we choose a coordinate system and observe the position of the moon in this coordinate system at consecutive instants of time. We then write down these observations. The data thus gathered are called the *raw* data. The raw data then undergo a process of cleansing, rectification and regimentation: we throw away data points that are obviously faulty, take into consideration what the measurement errors are, take averages, and usually idealize the data, for instance by replacing discrete data points by a continuous function. Often, although not always, the result is a smooth curve through the data points that satisfies certain theoretical desiderata (*Harris* [3.153] and *van Fraassen* [3.36, pp. 166–168] elaborate on this process). 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 structure in the requisite sense.

Suppes [3.122] was the first to suggested 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" [3.154, p. 667]. He has defended this position numerous times over the years ([3.77, p. 164], [3.101, p. 64], [3.118, p. 524], [3.155, p. 229] and [3.156, p. 271]) including in his most recent book on representation [3.36, pp. 246, 252]. 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 [3.148, p. 98]:

> "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."

Muller's point is that science aims (or at least has to aim) to represent real systems in the world and not data structures. *Van Fraassen* calls this the "loss of reality objection" [3.36, p. 258] and accepts that the structuralist must ensure that models represent target systems, rather than finishing the story at the level of data. In his [3.36] he addresses this issue in detail and offers a solution. We discuss his solution below, but before doing so we want to articulate the objection in more detail. To this end we briefly revisit the discussion about phenomena and data which took place in the 1980s and 1990s.

Bogen and *Woodward* [3.157], *Woodward* [3.158], and more recently (and in a somewhat different guise) *Teller* [3.159], introduced the distinction between phenomena and data and argue that models represent phenomena, not data. The difference is best introduced with an example: the discovery of weak neutral currents [3.157, pp. 315–318]. What the model at stake consists of is particles: neutrinos, nucleons, and the Z^0 particle, along with the reactions that take place between them. (The model we are talking about here is not the so-called standard model of elementary particles as a whole. Rather, what we have in mind is one specific model about the interaction of certain particles of the kind one would find in a theoretical paper on this experiment.) Nothing of that, however, shows in the relevant data. CERN (Conseil Européen pour la Recherche Nucléaire) in Geneva produced 290 000 bubble chamber photographs of which roughly 100 were considered to provide evidence for the existence of neutral currents. The notable point in this story is that there is no part of the model (provided by quantum field theory) that could be claimed to be isomorphic to these photographs. Weak neutral currents are the phenomenon under investigation; the photographs taken at CERN are the raw data, and any summary one might construct of the content of these photographs would be a data model. But it's weak neutral currents that occur in the model; not any sort of data we gather in an experiment.

This is not to say that these data have nothing to do with the model. The model posits a certain number of particles and informs us about the way in which they interact both with each other and with their environment. Using this knowledge we can place them in a certain experimental context. The data we then gather in an experiment are the product of the elements of the model and of the way in which they operate in that context. Characteristically this context is one that we are able to control and about which we have reliable knowledge (knowledge about detectors, accelerators, photographic plates and so on). Using this and the model we can derive predictions about what the outcomes of an experiment will be. But, and this is the salient point, these predictions involve the entire experimental setup and not only the model and there is nothing in the model itself with which one could compare the data. Hence, data are highly contextual and there is a big gap between observable outcomes of experiments and anything one might call a substructure of a model of neutral currents.

To underwrite this claim Bogen and Woodward notice that parallel to the research at CERN, the National Accelerator Laboratory (NAL) in Chicago also performed an experiment to detect weak neutral currents, but the data obtained in that experiment were quite different. They consisted of records of patterns of discharge in electronic particle detectors. Though the experiments at CERN and at NAL were totally different and as a consequence the data gathered had nothing in common, they were meant to provide evidence for the same theoretical model. But the model, to reiterate the point, does not contain any of these contextual factors. It posits certain particles and their interaction with other particles, not how detectors work or what readings they show. That is, the model is not idiosyncratic to a special experimental context in the way the data are and therefore it is not surprising that they do not contain a substructure that is isomorphic to the data. For this reason, models represent phenomena, not data.

It is difficult to give a general characterization of phenomena because they do not belong to one of the traditional ontological categories [3.157, p. 321]. In fact, phenomena fall into many different established categories, including particular objects, features, events, processes, states, states of affairs, or they defy classification in these terms altogether. This, however, does not detract from the usefulness of the concept of a phenomenon because specifying one particular ontological category to which all phenomena belong is inessential to the purpose of this section. What matters to the problem at hand is the distinctive role they play in connection with representation.

What then is the significance of data, if they are not the kind of things that models represent? The answer to this question is that data perform an evidential function. That is, data play the role of evidence for the presence of certain phenomena. The fact that we find a certain pattern in a bubble chamber photograph is evidence for the existence of neutral currents. Thus construed, we do not denigrate the importance of data in science, but we do not have to require that data have to be embeddable into the model at stake.

Those who want to establish data models as targets can reply to this in three ways. The first reply is an appeal to radical empiricism. By postulating phenomena over and above data we leave the firm ground of observable things and started engaging in transempirical speculation. But science has to restrict its claims to observables and remain silent (or at least agnostic) about the rest. Therefore, so the objection goes, phenomena are chimeras that cannot be part of any serious account of science. It is, however, doubtful that this helps the data model theorist. Firstly, note that it even rules out representing *observable phenomena*. To borrow *van Fraassen*'s example on this story, a population model of deer reproduction would represent data, rather than deer [3.36, pp. 254–260]. 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, irrespective of whether one understands phenomena realistically [3.157] or antirealistically [3.160], it is phenomena that models portray and not data. To deny the reality of phenomena just won't make a theoretical model *represent* data. Whether we regard neutral currents as real or not, it is neutral currents that are portrayed in a field-theoretical model, not bubble chamber photographs. Of course, one can suspend belief about the reality of these currents, but that is a different matter.

The second reply is to invoke a chain of representational relationships. *Brading* and *Landry* [3.137] 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 [3.137, p. 575]. 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, would that 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 [3.137, p. 575]. 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. With respect to the second question, although representation is not intransitive, it is not transitive [3.20, pp. 11–12]. 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* [3.36]. His *Wittgensteinian* solution is to diffuse the loss of reality objection. Once we pay sufficient attention to the pragmatic features of the contexts in which scientific and data models are used, *van Fraassen* claims, there actually is no difference between representing data and representing a target (or a phenomenon in Bogen and Woodward's sense) [3.36, p. 259]:

> "in a context in which a given [data] model is *someone*'s representation of a phenomenon, there is *for that person* no difference between the question *whether a theory* [theoretical model] *fits that representation* and the question *whether that theory fits the phenomenon*."

Van Frasseen's argument for this claim is long and difficult and we cannot fully investigate it here; we restrict attention to one crucial ingredient and refer the reader to *Nguyen* [3.161] for a detailed discussion of the argument.

Moore's paradox is that we cannot assert sentences of the form *p and I don't believe that p*, where *p* is an arbitrary proposition. For instance, someone cannot assert that Napoleon was defeated in the battle of Waterloo and assert, at the same time, that she doesn't believe that Napoleon was defeated in the battle of Waterloo. Van Fraassen's treatment of Moore's paradox is that speakers cannot assert such sentences because the pragmatic commitments incurred by asserting the first conjunct include that the speaker believe that *p*. This commitment is then contradicted by the assertion of the second conjunct. So instances of Moore's paradox are pragmatic contradictions. Van Fraassen then draws an analogy between this paradox and the scientific representation. He submits that a user simply cannot, on pain of pragmatic contradiction, assert that a data model of a target system be embeddable within a theoretical model without thereby accepting that the theoretical model represents the target.

However, *Nguyen* [3.161] argues that in the case of using a data model as a representation of a phenomenon, no such pragmatic commitment is incurred, and therefore no such contradiction follows when accompanied by doubt that the theoretical model also represents the phenomenon. To see why this is the case, consider a more mundane example of representation: a caricaturist can represent Margaret Thatcher as draconian without thereby committing himself to the belief that Margaret Thatcher really is draconian. Pragmatically speaking, acts of representation are weaker than acts of assertion: they do not incur the doxastic commitments required for van Fraassen's analogy to go through. So it seems van Fraassen doesn't succeed in dispelling the loss of reality objection. How target systems enter the picture in the structuralist account of scientific representation remains therefore a question that structuralists who invoke data models as providing the target-end structures must address. Without such an account the structuralist account of representation remains at the level of data, a position that seems implausible, and contrary to actual scientific practice.

We now turn to the second response: that a structure is instantiated in the system. As mentioned above, this response comes in three versions. The first is metaphysically more parsimonious and builds on the systems' constituents. Although target systems are not structures, they are composed of parts that instantiate physical properties and relations. The parts can 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*' discussion of the solar system [3.100, p. 22]). This supplies the required notion of structure. We might then say that physical systems instantiate a certain structure, and it is this structure that models are isomorphic to.

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 regard the atoms as objects and the bonds are relations. Denoting the carbon atom by a, and the four hydrogen atoms by b, c, d, and e, we obtain a structure S with the domain $U = \{a, b, c, d, e\}$ and the relation $r = \{\langle a, b\rangle, \langle b, a\rangle, \langle a, c\rangle, \langle c, a\rangle, \langle a, d\rangle, \langle d, a\rangle, \langle a, e\rangle, \langle e, a\rangle\}$, which can be interpreted as *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 identify 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 [3.21]. One could just as well choose bonds as objects and consider the relation *sharing a node with another bond*. Denoting the bonds by a', b', c' and d', we obtain a structure S' with the domain $U' = \{a', b', c', d'\}$ and the relation $r = \{\langle a', b'\rangle, \langle b', a'\rangle, \langle a', c'\rangle, \langle c', a'\rangle, \langle a', d'\rangle, \langle d', a'\rangle, \langle b', c'\rangle, \langle c', b'\rangle, \langle b', d'\rangle, \langle d', b'\rangle, \langle c', d'\rangle, \langle d', c'\rangle\}$. Obviously S and S' are not isomorphic. So which structure is picked out depends on how the system is described. Depending on which parts one regards as individuals and what relation one chooses, very different structures can emerge. And it takes little ingenuity to come up with further descriptions of the methane molecule, which lead to yet other structures.

There is nothing special about the methane molecule, and any target system can be presented under alternative descriptions, which ground different structures. So the lesson learned generalizes: there is no such thing as *the* structure of a target system. Systems only have a structure under a particular description, and there are many nonequivalent descriptions. This renders talk about a model being isomorphic to target system *simpliciter* meaningless. Structural claims do not *stand on their own* in that their truth rests on the truth of a more concrete description of the target system. As a consequence, descriptions are an integral part of an analysis of scientific representation.

In passing we note that *Frigg* [3.21, pp. 55–56] also provides another argument that pulls in the same direction: structural claims are abstract and are true only relative to a more concrete nonstructural description. For a critical discussion of this argument see *Frisch* [3.162, pp. 289–294] and *Portides*, Chap. 2.

How much of a problem this is depends on how austere one's conception of models is. The semantic view of theories was in many ways the result of an antilinguistic turn in the philosophy of science. Many proponents of the view aimed to exorcise language from an analysis of theories, and they emphasized that the model-world relationship ought to be understood as a *purely* structural relation. *Van Fraassen*, for instance, submits that "no concept which is essentially language dependent has any philosophical importance at all" [3.101, p. 56] and observes that "[t]he semantic view of theories makes language largely irrelevant" [3.155, p. 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.

For someone of that provenance the above argument is bad news. However, a more attenuated position could integrate descriptions in the package of modeling, 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" [3.53, p. 887]. Likewise, Munich structuralists explicitly acknowledge the need for a concrete description of the target system [3.163, pp. 37–38], and they consider these *informal descriptions* to be *internal* to the theory. This is a plausible move, but those endorsing this solution have to concede that there is more to epistemic 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 (a practically identical conclusion is reached in Putnam's so called model-theoretic argument; see *Demopoulos* [3.164] for a discussion). Hence, *any* structure of cardinality C is isomorphic to a target of cardinality C because the target instantiates any structure of cardinality C (see *Ketland* [3.165] and *Frigg* and *Votsis* [3.166] for discussions). This problem is not unsolvable, but all solutions 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. How to do this in the structuralist tradition remains unclear (*Ainsworth* [3.167] provides as useful summary of the different solutions).

Newman's theorem is both stronger and weaker than the argument from multiple descriptions. It's stronger in that it provides more alternative structures than multiple descriptions. It's weaker in that many of the structures it provides are *unphysical* because they are purely set theoretical combinations of elements. By contrast, descriptions pick out structures that a system can reasonably be seen as possessing.

The second version of the second response emerges from the literature on the applicability of mathematics. Structural platonists like *Resnik* [3.108] and *Shapiro* [3.41, 109, 168] take structures to be *ante rem* universals. In this view, structures exist independently of physical systems, yet they can be instantiated in physical systems. In this view systems instantiate structures and models are isomorphic to these instantiated structures.

This view raises all kind of metaphysical issues about the ontology of structures and the instantiation relation. Let us set aside these issues and assume that they can be resolved in one way or another. This would still leave us with serious epistemic and semantic questions. How do we know a certain structure is instantiated in a system and how do we refer to it? Objects do not come with labels on their sleeves specifying which structures they instantiate, and proponents of structural universals face a serious problem in providing an account of *how we access* the structures instantiated by target systems. Even if – as a brute metaphysical fact – target systems only instantiate a small number of structures, and therefore there is a substantial question regarding whether or not scientific models represent them, this does not help us understand how we could ever come to know whether or not the isomorphism holds. It seems that individuating a domain of objects and identifying relations between them is the only way for us to access a structure. But then we are back to the first version of the response, and we are again faced with all the problems that it raises.

The third version of the second response is more radical. One might take target systems themselves to be structures. If this is the case then there is no problem with the idea that they can be isomorphic 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 isomorphism between a model and the world. Surprisingly, some ontic structuralists have been hesitant to adopt such a view (see *French* and *Ladyman* [3.120, p. 113] and *French* [3.169, p. 195]). Others, however, seem to endorse it. *Tegmark* [3.170], for instance, offers an explicit defense 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" [3.170, p. 102] and argues that this entails that the world is a mathematical structure (his "mathematical universe hypothesis") [3.170, p. 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 [3.170, pp. 103–104]. Discussing this in any detail goes beyond our 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 of our current scientific models, whose targets aren't at this level. When modeling an airplane 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 nonfundamental target systems.

3.5 The Inferential Conception

In this section we discuss accounts of scientific representation that analyze 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.

3.5.1 Deflationary Inferentialism

Suárez argues that we should adopt a "deflationary or minimalist attitude and strategy" [3.32, p. 770] when addressing the problem of epistemic representation. We will discuss deflationism in some detail below, but in order to formulate and discuss Suárez's theory of representation we need at least a preliminary idea of what is meant by a deflationary attitude. In fact two different notions of deflationism are in operation in his account. The first is [3.32, p. 771]:

> "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."

We call the view that a theory of representation 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" [3.32, p. 771] or "platitudes" [3.171, p. 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" [3.172, p. 39]. We call this position s-deflationism (s for *surface feature*). As far as we can tell, Suárez intends his account of representation to be deflationary in both senses.

Suárez dubs the account that satisfies these criteria *inferentialism* [3.32, p. 773]:

Definition 3.9 Inferentialism 1

A scientific model M represents a target 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.

Notice that this condition is not an instantiation of the ER-scheme: in keeping with n-deflationism it features a material conditional rather than a biconditional and hence provides necessary (but not sufficient) conditions for M to represent T. We now discuss each condition in turn, trying to explicate in what way they 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" [3.32, p. 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" [3.32, p. 771]. This might look like denotation as in Sect. 3.2. But Suárez stresses 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 [3.172, p. 41]. Secondly, M can denote T only if T exists. Thus including denotation as a necessary condition on scientific representation "would rule out fictional representation, that is, representation of nonexisting entities" [3.32, p. 772], and "any adequate account of scientific representation must accommodate representations with fictional or imaginary targets" [3.172, p. 44].

The second issue is one that besets other accounts of representation too, in particular similarity and isomorphism accounts. 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.

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* we introduced in Sect. 3.1, now taken as a necessary condition on epistemic 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" [3.32, p. 773]. (The insistence on inference makes Suárez's account an instance of what *Chakravartty* [3.173] calls a *functional conception* of representation.)

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 license inferences about their targets, or what leads them to license some inferences and not others? *Contessa* makes this point most stridently when he argues that [3.29, p. 61]:

> "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."

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 and 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 us with a concept of epistemic 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?

We turn to this question shortly. Before doing so we want to briefly outline how the above account fares with respect to the other problems introduced in Sect. 3.1. The account provides a neat explanation of the possibility of misrepresentation [3.32, p. 776]:

> "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."

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, Suárez illustrates his account with a large range of representations, including diagrams, equations, scientific models, and nonscientific representations such as artistic portraits. 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" [3.32, p. 777] and he mentions the example of Velázquez's portrait of Innocent X [3.32]. It is clear that the conditions of inferentialism 1 (Definition 3.9) are met by nonscientific as well as scientific epistemic representations. So, at least without sufficient conditions, there is no clear way of demarcating between the different kinds of epistemic representation.

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 (Definition 3.9) 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*]" [3.32, p. 774]. 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 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 (Definition 3.9). 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 analyzed. 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.

Let us now return to the question about the motivation for deflationism. As we have seen, 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* [3.172] distinguishes between the *redundancy* theory (associated with Frank Ramsey and also referred to as the *no theory* view), *abstract minimalism* (associated with Crispin Wright) and the *use theory* (associated with Paul Horwich). What all three are claimed to have in common is that they accept the disquotational schema – i. e., instances of the form: P is true iff P. Moreover they [3.172, p. 37]

> "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."

He claims that the redundancy theory of truth is characterized by the idea that [3.172, p. 39]:

> "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."

So, as Suárez characterizes the position, the redundancy theory denies that any necessary and sufficient conditions for application of the truth predicate case be given. He argues that [3.172]:

> "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."

This motivates n-deflationism (although one might ask why such a position would allow even necessary conditions. Suárez doesn't discuss this).

This approach faces a number of challenges. First, the argument is based on the premise that if deflationism 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 favor of any sort of deflationary account of epistemic 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 deflationism 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* [3.174]). But as long as it's not clear that deflationism about truth is a superior position, it's hard to see how one can muster support for deflationism about representation by appealing to deflationism about truth.

Moreover, a position that allows only necessary conditions on epistemic representation faces a serious problem. While such an account allows us to *rule out* 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 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 (Definition 3.9) we are never in position to assert that a particular model actually is a representation, which is an unsatisfactory situation.

The other two deflationary positions in the debate over truth are abstract minimalism and the use theory. Suárez characterizes 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" [3.172, p. 40]. Abstract minimalism is presented as the view that while truth is [3.172, p. 40]:

> "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."

Both positions imply that necessary and sufficient conditions for truth *can* be given [3.172]. But on either account, such conditions only capture nonexplanatory surface features. This motivates s-deflationism.

Since s-deflationism explicitly allows for necessary and sufficient conditions, inferentialism 1 (Definition 3.9) can be extended to an instance of the ER-scheme, providing necessary and sufficient conditions (which also seems to be in line with *Suárez* and *Solé* [3.171, p. 41] who provide a formulation of inferentialism with a biconditional):

Definition 3.10 Inferentialism 2

A scientific model M represents a target T iff (i) the representational force of M points towards T, and (ii) M allows competent and informed agents to draw specific inferences regarding T.

If one takes conditions (i) and (ii) to refer to "features of activates within a normative practice, [that] do not stand for relations between sources and targets" [3.172, p. 46], then we arrive at a *use-based* account of epistemic representation. In order to understand a particular instance of a model M representing a target T we have to understand 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.

Plausibly, such a focus on practice amounts to looking at the inferential rules employed in each instance, or type of instance, of epistemic representation. This, however, raises a question about the status of any such analysis vis-à-vis the general theory of representation as given in inferentialism 2 (Definition 3.10). There seem to be two options. The first is to affirm inferentialism 2's (Definition 3.10) 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 (Definition 3.10) as providing abstract conditions that require concretization in each instance of epistemic representation (abstraction can here be understood, for instance, in *Cartwright*'s [3.74] sense). Studying the concrete realizations of the abstract conditions is then an integral part of the theory. This approach seems plausible, but it renders deflationism obsolete. Thus understood, the view becomes indistinguishable from a theory that accepts the *surrogative reasoning condition* and the *requirement of directionality* as conditions of adequacy and analyzes them in pluralist spirit, that is, under the assumption that these conditions can have different concrete realizers in different contexts. But this program can be carried out without ever mentioning deflationism.

One might reply that the first option unfairly stacks the deck against inferentialism and point out that different inferential practices *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 – to inferential notions – such as the validity of argument [3.175, 176]. 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 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* [3.177] and seems like a fruitful route for future research, but for want of space we will not discuss it in detail here. There is no evidence that Suárez would endorse such an approach. And, more worrying for inferentialism 2 (Definition 3.10), it is not clear whether such an approach would satisfy *s*-deflationism. Each investigation into the inferential rules utilized in each instance, or type of instance of epistemic representation will likely be a substantial (possibly sociological or anthropological) project. Thus the *s*-deflationary credentials of the approach – at least if they are taken to require that nothing substantial can be said about scientific representation in each instance, as well as in general – are called into question.

Finally, if the conditions in inferentialism 2 (Definition 3.10) are taken to be abstract platitudes then we arrive at an abstract minimalism. Although inferentialism 2 (Definition 3.10) defines the concept of epistemic representation, the definition does not suffice to explain the use of any particular instance of epistemic representation for ([3.172, p. 48], cf. [3.171]):

> "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."

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" ([3.171, p. 45], [3.32, p. 773], [3.172, p. 43]). 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 ([3.23, p. 230], [3.171, p. 43], [3.172, p. 46], [3.178, pp. 93–94]). We are told that the means cannot be reduced to the constituents but that [3.171, p. 43]:

> "all representational means (such as isomorphism and similarity) are concrete instantiations, or realizations, of one of the basic platitudes that constitute representation"

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" [3.171, p. 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 concretization in each instance of epistemic representation. In this sense it is immune to Contessa's objection about the mysterious capacity that models have to license 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 concretize the abstract conditions. But we are now facing a similar problem as the above. Even if s-deflationism applies to epistemic representation in general, an investigation into each specific instance of will involve uncovering substantial relations that hold between models and their targets, which again conflicts with Suárez's adherence to a deflationist approach.

3.5.2 Inflating Inferentialism: Interpretation

In response to difficulties like the above *Contessa* claims that "it is not clear why we should adopt a deflationary attitude *from the start*" [3.29, p. 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 [3.29, p. 48]:

> "[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."

To explain in virtue of what the inferences can be drawn, *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 ([3.29, p. 57], [3.179, pp. 126–127]):

Definition 3.11 Interpretation
A scientific 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 M in terms of T.

Contessa offers a detailed formal characterization of an interpretation, which we cannot repeat here for want of space (see [3.29, pp. 57–62] for details). The leading idea is that the model user first identifies a set of relevant objects in the model, and a set of properties and relations these objects instantiate, along with a set of relevant objects in the target and a set of properties and relations these objects instantiate. The user then:

1. Takes M to denote T.
2. Takes every identified object in the model to denote exactly one object in the target (and every relevant object in the target has to be so denoted and as a result there is a one-to-one correspondence between relevant objects in the model and relevant objects in the target).
3. Takes every property and relation in the model to denote a property or relation of the same arity in the target (and, again, and every property and relation in the target has to be so denoted and as a result there will be one-to-one correspondence between relevant properties and relations in the model and target).

A formal rendering of these conditions is what Contessa calls an *analytic interpretation* (he also includes an additional condition pertaining to functions in the model and target, which we suppress for brevity). The relationship between interpretations and the surrogative reasoning mentioned above is that it is in virtue of the user adopting an analytic interpretation that a model licenses inferences about its target.

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, and relations can be interpreted as full-fledged relations rather than purely extensionally specified sets of tuples.

Interpretation (Definition 3.11) is a nondeflationary account of scientific representation: most (if not all) instances of scientific representation involve a model user adopting 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 [3.29, p. 58]

> "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."

Even with this in mind, it is clear that he intends that *some* interpretation is a necessary condition on epistemic representation.

Let's now turn to how interpretation fares with respect to our questions for an account of epistemic representation as set out in Sect. 3.2. Modulo the caveat about nonanalytical interpretations, interpretation (Definition 3.11) provides necessary and sufficient conditions on epistemic representation and hence answers the ER-problem. Furthermore, it does so in a way that explains the directionality of representation: 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 with the structuralist account an emphasis on relations and one-to-one model-target correspondence, Contessa can appeal to the same account of the applicability of mathematics as structuralist.

With respect to the demarcation problem, *Contessa* is explicit that "[p]ortraits, photographs, maps, graphs, and a large number of other representational devices" perform inferential functions [3.29, p. 54]. Since nothing in the notion of an interpretation seems restricted to scientific models, it is plausible to regard interpretation (Definition 3.11) as a universal theory of epistemic representation (a conclusion that is also supported by the fact that *Contessa* [3.29] uses the example of the London Underground map to motivate his account; see also [3.179]). As such, interpretation (Definition 3.11) seems to deny the existence of a substantial distinction between scientific and nonscientific epistemic representations (at least in terms of their representational properties). It remains unclear how interpretation (Definition 3.11) addresses the problem of style. As we have seen earlier, in particular visual representations fall into different categories. It is a question for future research how these can be classified within the interpretational framework.

With respect to the question of ontology, interpretation (Definition 3.11) itself places few constraints on what scientific models are, ontologically speaking. All it requires is that they consist of objects, properties, relations, and functions. For this reason our discussion in Sect. 3.3.3 above rears its head again here. As before, how to apply interpretation (Definition 3.11) to physical models can be understood relatively easily. But how to apply it to nonphysical models is less straightforward. *Contessa* [3.180] distinguishes between mathematical models and fictional models, where fictional models are taken to be fictional objects. We briefly return to his ontological views in Sect. 3.6.

In order to deal with the possibly of misrepresentation, *Contessa* notes that "a user does not need to believe that every object in the model denotes some object in the system in order to interpret the model in terms of the system" [3.29, p. 59]. He illustrates this claim with an example of contemporary scientists using the Aristotelian model of the cosmos to represent the universe, pointing out that "in order to interpret the model in terms of the universe, we do not need to assume that the sphere of fixed stars itself [...] denotes anything in the universe" [3.29].

From this example it is clear that the relevant sets of objects, properties and functions isolated in the construction of the analytic interpretation do not need to exhaust the objects, properties, relations, and functions of either the model or the target. The model user can identify a relevant *proper* subset in each instance. This allows interpretation (Definition 3.11) to capture the common practice of abstraction in scientific models: a model need only represent some features of its target, and moreover, the model may have the sort of *surplus* features are not taken to represent anything in the target, i. e., that not all of a model's features need to play a direct representational role.

This suggestion bears some resemblance to partial structures, and it suffers from the same problem too. In particular distortive idealisations are a source of problems for interpretation (Definition 3.11), as several commentators have observed (see *Shech* [3.181] and *Bolinska* [3.28]). Contessa is aware of this problem and illustrates it with the example of a massless string. His response to the problem is to appeal to a user's corrective abilities [3.29, p. 60]:

> "since models often misrepresent some aspect of the system or other, it is usually up to the user's competence, judgment, and background knowledge to use the model successfully in spite of the fact that the model misrepresents certain aspects of the system."

This is undoubtedly true, but it is unclear how such a view relates, or even derives from, interpretation (Definition 3.11). An appeal to the competence of users seems to be an ad hoc move that has no systematic grounding in the idea of an interpretation, and it is an open question how the notion of an interpretation could be amended to give distortive idealizations a systematic place.

Ducheyne [3.182] provides a variant of interpretation (Definition 3.11) that one might think could be used to accommodate these distortive idealizations. The details of the account, which we won't state precisely here for want of space, can be found in [3.182, pp. 83–86]. The central idea is that each relevant relation specified in the interpretation holds precisely in the model, and corresponds to the same relation 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 and relations 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 succeed in explaining how distortive idealizations are scientific representations, then more needs to be said about how a target relation can *approximate* a model relation. *Ducheyne* implicitly relies on the fact that relations are such that "we can determine *the extent to which* [they hold] empirically" [3.182, p. 83] (emphasis added). This suggests that he has quantifiable relations in mind, and that what it means for a relation *r* in the target to approximate a relation *r*' 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* [3.183], *Laymon* [3.184], *Liu* [3.185], *Norton* [3.186], and *Ramsey* [3.187], among others, offer discussions of different kinds of idealizations 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 (now derogatively referred to as the *plum pudding model*) is an approximation of what the quantum mechanical shell 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. But this does not mean that it is not a scientific representation of the atom; it's just an incorrect one. 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 license falsehoods about their targets).

3.5.3 The Denotation, Demonstration, and Interpretation Account

Our final account is *Hughes*' *denotation, demonstration, and interpretation* (DDI) account of scientific representation [3.188] and [3.189, Chap. 5]. This account has inspired both the inferential (see *Suárez* [3.32, p. 770] and [3.172]) and the interpretational account (see *Contessa* [3.179, p. 126]) discussed in this section.

Quoting directly from *Goodman* [3.64, p. 5], *Hughes* takes a model of a physical system to "be a symbol for it, stand for it, refer to it" [3.188, p. 330]. Presumably the idea is that a model denotes its target it the same way that a proper name denotes its bearer, or, stretching the notion of denotation slightly, a predicate denote elements in its extension. (*Hughes* [3.188, p. 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 Sect. 3.8). 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" [3.188, p. 331] (that models have an *internal dynamic* is all that Hughes has to say about the problem of ontology). 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" [3.188], the other is a physical ripple chamber where they are supplied by "the natural processes involved in the propagation of water waves" [3.188, p. 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 [3.188, p. 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 unanalyzed) 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 [3.188, p. 339]

> "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."

We are 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 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 epistemic 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 discussion of deflationism in Sect. 3.5.1: perhaps the very act of using a model, with all the user intentions and practices that brings with it, constitutes the epistemic representation relationship itself. And as such, perhaps the DDI conditions could be taken as an answer to the ER-problem:

Definition 3.12 DDI-ER
A scientific model M represents a target T iff 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 of thereof) to be about T.

This account comes very close to interpretation (Definition 3.11) as discussed in Sect. 3.5.2. And as such it serves to answer the questions we set out in Sect. 3.1 above in the same way. But in this instance, the notion of what it means to *exploit an internal dynamic* and *interpret the results* of this to be about T need further explication. If the notion of an 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 characterizes as connecting models to the target with a theoretical hypothesis [3.190, p. 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" [3.188, p. 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 utilizing 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 realizations of the abstract conditions are discussed. Despite Hughes' claims to the contrary, such an account could prove a viable answer the ER-problem, and it seems to capture much of what is valuable about both the abstract minimalist version of inferentialism 2 (Definition 3.10) as well as interpretation (Definition 3.11) discussed above.

3.6 The Fiction View of Models

In this section we discuss a number of recent attempts to analyze scientific modeling by drawing an analogy with literary fiction. We begin by introducing the leading ideas and differentiating between different strands of argument. We then examine a number of accounts that analyze epistemic representation against the backdrop of literary fiction. We finally discuss criticisms of the fiction view.

3.6.1 Models and Fiction

Scientific discourse is rife with passages that appear to be descriptions of systems in a particular discipline, and the pages of textbooks and journals are filled with discussions of the properties and the behavior of those systems. Students of mechanics investigate at length the dynamical properties of a system consisting of two or three spinning spheres with homogeneous mass distributions gravitationally interacting only with each other. Population biologists study the evolution of one species that reproduces at a constant rate in an unchanging environment. And when studying the exchange of goods, economists consider a situation in which there are only two goods, two perfectly rational agents, no restrictions on available information, no transaction costs, no money, and dealings are done immediately. Their surface structure notwithstanding, no one would mistake descriptions of such systems as descriptions of an *actual* system: we know very well that there are no such systems (of course some models are actual systems – a scale model of a car in a wind tunnel for example – but in this section we focus on models that are not of this kind). Scientists sometimes express this fact by saying that they talk about *model land* (for instance [3.191, p.135]).

Thomson-Jones [3.98, p. 284] refers to such a description as a "description of a missing system". These descriptions are embedded in what he calls the "face value practice" [3.98, p. 285]: the practice of talking and thinking about these systems as if they were real. We observe that the amplitude of an ideal pendulum remains constant over time in much the same way in which we say that the Moon's mass is approximately 7.34×10^{22} kg. Yet the former statement is about a point mass suspended from a massless string – and there is no such thing in the world.

The face value practice raises a number of questions. What account should be given of these descriptions and what sort of objects, if any, do they describe? How should we analyze the face value practice? Are we putting forward truth-evaluable claims when putting forward descriptions of missing systems? An answer to these questions emerges from the following passage by *Peter Godfrey-Smith* [3.152, p. 735]:

> "[. . .] I take at face value the fact that modelers often *take* themselves to be describing imaginary biological populations, imaginary neural networks, or imaginary economies. [. . .] Although these imagined entities are puzzling, I suggest that at least much of the time they might be treated as similar to something that we are all familiar with, the imagined objects of literary fiction. Here I have in mind entities like Sherlock Holmes' London, and Tolkein's Middle Earth. [. . .] the model systems of science often work similarly to these familiar fictions."

This is the core of the fiction view of models: models are akin to places and characters in literary fiction. When modeling the solar system as consisting of ten perfectly spherical spinning tops physicists describe (and *take themselves* to describe) an imaginary physical system; when considering an ecosystem with only one species biologists describe an imaginary population; and when investigating an economy without money and transaction costs economists describe an imaginary economy. These imaginary scenarios are tellingly like the places and characters in works of fiction like Madame Bovary and Sherlock Holmes.

Although hardly at the center of attention, the parallels between certain aspects of science and literary fiction have not gone unnoticed. Maxwell discussed in great detail the motion of "a purely imaginary fluid" in order to understand the electromagnetic field [3.192, pp. 159–160]. The parallel between science and fiction occupied center stage in *Vaihinger*'s [3.193] philosophy of the *as if*. More recently, the parallel has also been drawn specifically between models and fiction. *Cartwright* observes that "a model is a work of fiction" [3.194, p. 153] and later suggests an analysis of models as fables [3.73, Chap. 2]. *McCloskey* [3.195] emphasises the importance of narratives and stories in economics. *Fine* notes that modeling natural phenomena in every area of science involves fictions in Vaihinger's sense [3.196, p. 16], and *Sklar* highlights that describing systems *as if* they were systems of some other kind is a royal route to success [3.197, p. 71]. *Elgin* [3.198, Chap. 6] argues that science shares important epistemic practices with artistic fiction. *Hartmann* [3.199] and *Morgan* [3.200] emphasize that stories and narratives play an important role in models, and *Morgan* [3.201] stresses the importance of imagination in model building. *Sugden* [3.202] points out that economic models describe "counterfactual worlds" constructed by the modeler. *Frigg* [3.30, 203] suggests that models are imaginary objects, and *Grüne-Yanoff* and *Schweinzer* [3.204] emphasize the importance of stories in the application of game theory. *Toon* [3.48, 205] has formulated an account of representation based on a theory of literary fiction. *Contessa* [3.180] provides a fictional ontology of models and *Levy* [3.43, 206] discusses models as fictions.

But simply likening modeling to fiction does not solve philosophical problems. Fictional discourse and fictional entities face well-known philosophical questions, and hence explaining models in terms of fictional characters seems to amount to little more than to explain *obscurum per obscurius*. The challenge for proponents of the fiction view is to show that drawing an analogy between models and fiction has heuristic value.

A first step towards making the analogy productive is to get clear on what the problem is that the appeal to fiction is supposed to solve. This issue divides proponents of the fiction view into two groups. Authors belonging to the first camp see the analogy with fiction as providing an answer to the problem of ontology. Models, in that view, are *ontologically* on par with literary fiction while there is no productive parallel between models and fiction as far as the ER-problem (or indeed any other problem of representation) is concerned. Authors belonging to the second group hold the opposite view. They see the analogy with fiction first and foremost as providing an answer to the ER-problem (although, as we have seen, this may place restrictions on the ontology of models). Scientific representation, in this view, has to be understood along the lines of how literary fiction relates to reality. Positions on ontology vary. Some authors in this group also adopt a fiction view of ontology; some remain agnostic about the analogy's contribution to the matters of ontology; and some reject the problem of ontology altogether.

This being a review of models and representation, we refer the reader to *Gelfert*'s contribution to this book for an in-depth discussion of the ontology of models, Chap. 1, and focus on the fiction view's contribution to semantics. Let us just note that those who see fiction as providing an ontology of models are spoiled for choice. In principle every option available in the extensive literature on fiction is a candidate for an ontology of models; for reviews of these options see *Friend* [3.207] and *Salis* [3.208]. Different authors have made different choices, with proposals being offered by *Contessa* [3.180], *Ducheyne* [3.72], *Frigg* [3.203], *Godfrey-Smith* [3.209], *Levy* [3.43], and *Sugden* [3.210]. *Cat* [3.211], *Liu* [3.212, 213], *Pincock* [3.214, Chap. 12], *Thomson-Jones* [3.98] and *Toon* [3.205] offer critical discussions of some of these approaches.

Even if these ontological problems were settled in a satisfactory manner, we would not be home and dry yet. *Vorms* [3.215, 216] argues that what's more important than the entity itself is the format in which the entity is presented. A fiction view that predominantly focuses on understanding the fictional entities themselves (and, once this task is out of the way, their relation to the real-world targets), misses an important aspect, namely how agents draw inferences from models. This, Vorms submits, crucially depends on the format under which they are presented to scientists, and different formats allow scientists to draw different inferences. This ties in with *Knuuttila*'s insistence that we ought to pay more attention to the "medium of representation" when studying models [3.9, 217].

One last point stands in need of clarification: the meaning of the term *fiction*. Setting aside subtleties that are irrelevant to the current discussion, the different uses of *fiction* fall into two groups: fiction as falsity and fiction as imagination [3.218]. Even though not mutually exclusive, the senses should be kept separate. The first use of *fiction* characterizes something as deviating from reality. We brand Peter's account of events a fiction if he does not report truthfully how things have happened. In the second use, *fiction* refers to a kind of literature, *literary fiction*. Rife prejudice notwithstanding, the defining feature of literary fiction is not falsity. Neither is everything that is said in, say, a novel untrue (novels like War and Peace contain correct historical information); nor does every text containing false reports qualify as fiction (a wrong news report or a faulty documentary do not by that token turn into fiction – they remain what they are, namely wrong factual statements). What makes a text fictional is the attitude that the reader is expected to adopt towards it. When reading a novel we are not meant to take the sentences we read as reports of fact; rather we are supposed to imagine the events described.

It is obvious from what has been said so far that the fiction view of models invokes the second sense of *fiction*. Authors in this tradition do not primarily intend to brand models as false; they aim to emphasize that models are presented as something to ponder. This is not to say the first sense of fiction is irrelevant in science. Traditionally fictions in that sense have been used as calculational devices for generating predictions, and recently *Bokulich* [3.14] emphasized the explanatory function of fictions. The first sense of fiction is also at work in philosophy where antirealist positions are described as fictionalism. For instance, someone is a fictionalist about numbers if she thinks that numbers don't exist (see *Kalderon* [3.219] for a discussion of several fictionalisms of this kind). Scientific antirealists are fictionalists about many aspects of scientific theories, and hence *Fine* characterizes fictionalism as an "antirealist position in the debate over scientific realism" [3.196, 220, 221], a position echoed in *Winsberg* [3.222] and *Suárez* [3.223]. *Morrison* [3.224] and *Purves* [3.225] and offer critical discussions of this approach, which the latter calls fiction as "truth conducive falsehood" [3.225, p. 236]; *Woods* [3.226] offers a critical assessment of fictionalism in general. Although there are interesting discussions to be had about the role that this kind of fictions play in the philosophy of science, it is not our interest here.

Part A | 3.6

3.6.2 Direct Representation

In this subsection and the next we discuss proposals that have used the analogy between models and fiction to elucidate representation.

Most theories of representation we have encountered so far posit that there are model systems and construe epistemic representation as a relation between two entities, the model system and the target system. *Toon* calls this the *indirect* view of representation [3.205, p. 43]; *Levy*, speaking specifically about the fiction view of models, refers to it as the *whole-cloth fiction* view [3.206, p. 741]. Indeed, *Weisberg* views this indirectness as the defining feature of modeling [3.227]. This view faces the problem of ontology because it has to say what kind of things model systems are. This view contrasts with what *Toon* [3.205, p. 43] and *Levy* [3.43, p. 790] call a *direct* view of representation (*Levy* [3.206, p. 741] earlier also referred to it as the *worldly fiction* view). This view does not recognize model systems and aims instead to explain epistemic representation as a form of direct description. Model descriptions (like the description of an ideal pendulum) provide an "imaginative description of real things" [3.206, p. 741] such as actual pendula, and there is no such thing as a model system of which the pendulum description is literally true [3.205, pp. 43–44]. In what follows we use Toon's terminology and refer to this approach as *direct representation*.

Toon and Levy both reject the indirect approach because of metaphysical worries about fictional entities, and they both argue that the direct view has the considerable advantage that it does not have to deal with the vexed problem of the ontology of model systems and their comparison with real things at all. *Levy* [3.43, p. 790] sees his approach as "largely complimentary to Toon's". So we first discuss Toon's approach and then turn to Levy's.

Toon [3.48, 205, 228] takes as his point of departure *Walton*'s [3.229] theory of representation in the arts. 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 [3.229, p. 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 *props*, and the rule that we imagine a bear when we see a stump is a *principle of generation*. Together a prop and a principle of generation prescribe what is to be imagined. If a proposition is so prescribed to be imagined, then the proposition is *fictional* in the relevant game. The term *fictional* has nothing to do with falsity; on the contrary, it indicates that the proposition is *true in the game*. The set of propositions actually imagined by someone need not coincide with the set of all fictional propositions in 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 fictional in the game.

Walton considers a vast variety of different props. In the current context two kinds of props are particularly important. The first are objects like statues. Consider a statue showing Napoleon on horseback [3.205, p. 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 literary fictions of a certain kind, generates fictional truths by prescribing readers to imagine certain things. For instance, when reading *The War of the Worlds* [3.205, p. 39] 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.

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 [3.229, pp. 35,39]. 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 [3.205, p. 37]: the model is the prop and the bridge is the object of the representation. The same observation applies to theoretical models, such as a mechanical model of a bob bouncing on a spring. The model portrays the bob as a point mass and the spring as perfectly elastic. The model description represents the real ball and spring system in the same way in which a literary text represents its objects [3.205, pp. 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.

We now see why Toon's account is a direct view of modeling. Theoretical model descriptions represent actual concrete objects: the Forth Road Bridge and the bob 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 in.

This is an elegant account of representation, but it is not without problems. The first issue is that it does not offer an answer to the ER-problem. Imagining that the target has a certain feature does not tell us how the imagined feature relates to the properties the target actually has, 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 of the real bob. *Toon* mentions this problem briefly. His response is that [3.205, pp. 68–69]:

> "Principles of generation often link properties of models to properties of the system they represent in a rather direct way. If the model has a certain property then we are to imagine that system does too. If the model is accurate, then the model and the system will be similar in this respect. [. . .] [But] not all principles of generation are so straightforward. [. . .] In some cases similarity seems to play no role at all."

In as far as the transfer mechanism is similarity, the view moves close to the similarity view, which brings with it both some of the benefits and the problems we have discussed in Sect. 3.3. The cases in which similarity plays no role are left unresolved and it remains unclear how surrogative reasoning with such models is supposed to happen.

The next issue is that not all models have a target system, which is a serious problem for a view that analyzes representation in terms of imagining something *about* a target. *Toon* is well aware of this issue and calls them *models without objects* [3.205, p. 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 [3.205, p. 76]. 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 [3.205, p. 76].

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 object. Passages from *Dracula*, for instance, "do not represent any actual, concrete object but are instead about fictional characters" [3.205, p. 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 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 [3.205, p. 54].

Toon immediately admits that models without targets "give rise to all the usual problems with fictional characters" [3.205, p. 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. He offers the following summative statement of his account [3.205, p. 62]:

Definition 3.13 Direct Representation

A scientific model *M* represents a target system *T* iff *M* functions as prop in game of make believe.

This definition takes it to be understood that the imaginings prescribed are about the target *T* if there is a target, and about a fictional character if there isn't because there need not be any object that the model prescribes imaginings about [3.205, p. 81].

This bifurcation of imaginative activities raises questions. The first is whether the bifurcation squares with the face value practice. Toon's presentation would suggest that the imaginative practices involved in models with targets are very different from the ones involved in models without them. Moreover, they require a different analysis because imagining something about an existing object is different from imagining something about a fictional entity. This, however, does not seem to sit well with scientific practice. In some cases we are mistaken: we think that the target exists but then find out that it doesn't (as in the case of 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 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. This does not preclude different philosophical analyzes being given of modeling with and without a target, but any such analysis will have to make clear the commonalities between the two.

Let us now turn to a few other aspects of direct representation (Definition 3.13). The view successfully solves the problem of asymmetry. Even if it uses similarity in response to the ER-problem, the imaginative process is clearly directed towards the target. An appeal to imagination also solves the problem of misrepresentation because there is no expectation that our imaginations are correct when interpreted as statements about the target. Given its roots in a theory of representation in art, it's natural to renounce any attempts to demarcate scientific representation from other kinds of representation [3.205, p. 62]. The problem of ontology is dispelled for representations with an object, but it remains unresolved for representations without one. However, direct representation (Definition 3.13) offers at best a partial answer to the ER-problem, and nothing is said about either the problem of style and/or standards of accuracy. Similarly, Toon remains silent about the applicability of mathematics.

Levy also rejects an indirect view primarily because of the unwieldiness of its ontology and endorses a direct view of representation ([3.43, pp. 780–790], [3.206, pp. 744–747]). Like Toon, he develops his version of the direct view by appeal to Walton's notion of prop-oriented make believe. When, for instance, we're asked where in Italy the town of Crotone lies, we can be told that it's in the arch of the Italian boot. In doing so we are asked to imagine something about the shape of Italy and this imagination is used to convey geographical information. *Levy* then submits that "we treat models as games of prop-oriented make believe" [3.206, p. 791]. Hence modeling consists in imagining something directly about the target.

Levy pays careful attention to the ER-problem. In his [3.206, p. 744] he proposed that the problem be conceptualized in analogy with metaphors, but immediately added that this was only a beginning which requires substantial elaboration. In his [3.43, pp. 792–796] he takes a different route and appeals to *Yablo*'s [3.230] theory 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" [3.43, p. 792]. *Levy* submits that this will also work for a number of cases of modeling, but immediately adds that there are other sorts of cases that don't fit the mold [3.43, p. 794]. Such cases often are ones in which distortive idealizations are crucial and cannot be set aside. These require a different treatment and it's an open question what this treatment would be.

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 [3.43, pp. 796–797]. To this end he appeals to Godfrey-Smith's notion of *hub-and-spoke* cases: families of models where only some have a target (which makes them the hub models) and the others are connected to them via conceptual links (spokes) but don't have a specific target. Levy points out that such cases should be understood as having a *generalized target*. If something that looks like a model doesn't meet the requirement of having even a generalized target, then it's not a model at all. *Levy* mentions structures like the game of life and observes that they are "bits of mathematics" rather than models [3.43, p. 797]. This eliminates the need for fictional characters in the case of targetless models.

This is a heroic act of liberation, but questions about it remain. The direct view renders fictional entities otiose by positing that a model is nothing but an act of imagining something about a concrete actual thing. But generalized targets are not concrete actual things, and often not even classes of such things. There is a serious question whether one can still reap the (alleged) benefits of a view that analyzes modeling as imaginings about concrete things, if the things about which we imagine something are no longer concrete. Population growth or complex behavior are not concrete things like rabbits and stumps, and this would seems to pull the rug from underneath a direct approach to representation. Likewise, the claim that models without target are just mathematics stands in need of further elucidation. Looking back at Toon's examples of such models, a view that considers them just mathematics does not come out looking very natural.

3.6.3 Parables and Fables

Cartwright [3.231] focuses on highly idealized models such as *Schelling*'s model of social segregation [3.232] and *Pissarides*' model of the labor market [3.233].The problem with these models is that the objects and situations we find in such models are not at all like the things in the world that we are interested in. Cities aren't organized as checkerboards and people don't move according to simple algorithmic rules (as they do in Schelling's model), and there are no laborers who are solely interested in leisure and income (as is the case in Pissarides' model). Yet we are supposed to learn something about the real world from these models. The question is how.

Cartwright submits that an answer to this question emerges from a comparison of models with narratives, in particular fables and parables. An example of a fable is the following: "A marten eats the grouse; a fox throttles the marten; the tooth of the wolf, the fox. Moral: the weaker are always prey to the stronger" [3.231, p. 20]. The characters in the fable are highly idiosyncratic, and typically we aren't interested in them per se – we don't read fables to learn about foxes and martens. What we are interested in is the fable's general and more abstract conclusion, in the above example that the weaker are always prey to the stronger. In the case of the fable the moral is typically built in the story and explicitly stated [3.231].

Cartwright then invites us to consider the parable of the laborers in the vineyard told in the Gospel of Matthew [3.231]. A man goes to the market to hire day laborers. He hires the first group early in the morning, and then returns several times during the day to hire more laborers, and he hires the last group shortly before dusk. Some worked all day, while some hardly started when the day ended. Yet he pays the same amount to all of them. Like in a fable, when engaging with a parable the reader takes no intrinsic interest in the actors and instead tries to extract a more general moral. But unlike in fables, in parables no moral appears as part of the parable itself [3.231, p. 29]. Hence parables need interpretation, and alternative interpretations are possible. The above fable is often interpreted as being about the entry to God's kingdom, but, as Cartwright observes, it can just as well be interpreted as making the market-based capitalist point that you get what you contract for, and should not appeal to higher forms of justice [3.231, p. 21].

These are features models share with fables and parables: "like the characters in the fable, the objects in the model are highly special and do not in general resemble the ones we want to learn about" [3.231, p. 20] and the "lesson of the model is, properly, more abstract than what is seen to happen in the model" [3.231, p. 28]. This leaves the question whether models are fables or parables. Some models are like fables in that they have the conclusion explicitly stated in them. But most models are like parables [3.231, p. 29]: their lesson is not written in the models themselves [3.231, p. 21], and worse: "a variety of morals can be attributed to the models" [3.231, p. 21]. A model, just like a parable, is interpreted against a rich background of theory and observation, and the conclusion we draw depends to a large extent on the background [3.231, p. 30].

So far the focus was on deriving a conclusion about the model *itself*. Cartwright is clear that one more step is needed: "In many cases we want to use the results of these models to inform our conclusions about a range of actually occurring (so-called *target*) situations" [3.231, p. 22] (original emphasis). In fact, making this transfer of model results to the real world is the ER-problem. Unfortunately she does not offer much by way of explaining this step and merely observes that "a description of what happens in the model that does not fit the target gets recast as one that can" [3.231, p. 20]. This gestures in the right direction, but more would have to be said about how exactly a model description is recast to allow for transfer of model results to target systems. In earlier work *Cartwright* observed that what underlies the relationship between models and their targets is a "loose notion of resemblance" [3.73, pp. 192–193] and [3.74, pp. 261–262]. This could be read as suggesting that she would endorse some kind of similarity view of representation. Such a view, however, is independent of an appeal to fables and parables.

In passing we would like to mention that the same kind of models is also discussed in *Sugden* [3.202, 210]. However, his interest is in induction rather than representation, and if reframed in representational terms then his account becomes a similarity account like Giere's. See *Grüne-Yanoff* [3.234] and *Knuuttila* [3.235] for a discussion.

3.6.4 Against Fiction

The criticisms we have encountered above were intrinsic criticisms of particular versions of the fiction view, and as such they presuppose a constructive engagement with the view's point of departure. Some critics think that any such engagement is misplaced because the view got started on the wrong foot entirely. There are five different lines of attack. The first criticism is driven by philosophical worries about fiction. Fictions, so the argument goes, are intrinsically dubious and are beset with so many serious problems that one should steer away from them whenever possible. So it could be claimed that assigning them a central role in science is a manifestation of philosophical masochism. This, however, overstates the problems with fictions. Sure enough, there is controversy about fictions. But the problems pertaining to fictions aren't more devastating than those surrounding other items on the philosophical curriculum, and these problems surely don't render fictions off limits.

The second criticism, offered for example by *Giere* [3.97, p. 257], is that the fiction view – involuntarily – plays into the hands of irrationalists. Creationists and other science skeptics will find great comfort, if not powerful rhetorical ammunition, in the fact that philosophers of science say that scientists produce fiction. This, so the argument goes, will be seen as a justification of the view that religious dogma is on par with, or even superior to, scientific knowledge. Hence the fiction view of models undermines the authority of science and fosters the cause of those who wish to replace science with religious or other unscientific worldviews.

Needless to say, we share Giere's concerns about creationism. In order not to misidentify the problem it is important to point out that Giere's claim is not that the view itself – or its proponents – support creationism; his worry is that the view is a dangerous tool when it falls into the wrong hands. What follows from this, however, is not that the fiction view itself should be abandoned; but rather that some care is needed when dealing with the press office. As long as the fiction view of models is discussed in informed circles, and, when popularized, is presented carefully and with the necessary qualifications, it is no more dangerous than other ideas, which, when taken out of context, can be put to uses that would (probably) send shivers down the spines of their progenitors (think, for instance, of the use of Darwinism to justify eugenics).

The third objection, also due to *Giere*, has it that the fiction view misidentifies the aims of models. Giere agrees that from an *ontological* point of view scientific models and works of fictions are on par, but emphasizes that "[i]t is their differing function in practice that makes it inappropriate to regard scientific models as works of fiction" [3.97, p. 249]. *Giere* identifies three functional differences [3.97, pp. 249–252]. First, while fictions are the product of a single author's individual endeavors, scientific models are the result of a public effort because scientists discuss their creations with their colleagues and subject them to public scrutiny. Second, there is a clear distinction between fiction and nonfiction books, and even when a book classified as nonfiction is found to contain false claims, it is not reclassified as fiction. Third, unlike works of fiction, whose prime purpose is to entertain (although some works can also give insight into certain aspects of human life), scientific models are representations of certain aspects of the world.

These observations, although correct in themselves, have no force against the fiction view of models. First, whether a fiction is the product of an individual or a collective effort has no impact on its status as a fiction; a collectively produced fiction is just a different kind of fiction. Even if *War and Peace* (to take Giere's example) had been written in a collective effort by all established Russian writers of Tolstoy's time, it would still be a fiction. Vice versa, even if Newton had never discussed his model of the solar system with anybody before publishing it, it would still be science. The history of production is immaterial to the fictional status of a work. Second, as we have seen in Sect. 3.6.1, falsity is not a defining feature of fiction. We agree with Giere that there is a clear distinction between texts of fiction and nonfiction, but we deny that this distinction is defined by truth or falsity; it is the attitude that we are supposed to adopt towards the text's content that makes the difference. Once this is realized, the problem fades away. Third, many proponents of the fiction view (those belonging to the first group mentioned in Sect. 3.6.1) are clear that problems of ontology should be kept separate from function and agree that it is one of the prime function of models to represent. This point has been stressed by *Godfrey-Smith* [3.209, pp. 108–111] and it is explicit in other views such as *Frigg*'s [3.203].

The fourth objection is due to *Magnani*, who dismisses the fiction view for misconstruing the role of models in the process of scientific discovery. The fundamental role played by models, he emphasizes [3.236, p. 3]:

> "is the one we find in the core conceptual discovery processes, and that these kinds of models cannot be indicated as fictional at all, because they are constitutive of new scientific frameworks and new empirical domains."

This criticism seems to be based on an understanding of fiction as falsity because falsities can't play a constitutive role in the constitution of new empirical domains. We reiterate that the fiction view is not committed to the *fiction as falsity* account and hence is not open to this objection.

The fifth objection is that fictions are superfluous and hence should not be regarded as forming part of (let alone *being*) scientific models because we can give a systematic account of how scientific models work without invoking fictions. This point has been made in different ways by *Pincock* [3.214, Chap. 12] and *Weisberg* [3.33, Chap. 4] (for a discussion of Weisberg's arguments see *Odenbaugh* [3.237]). We cannot do justice to the details of their sophisticated arguments here, and will concern ourselves only with their main conclusion. They argue that scientific models are mathematical objects and that they relate to the world due to the fact that there is a relationship between the mathematical properties of the model and the properties found in the target system (in Weisberg's version similarity relations to a parametrized version of the target). In other words, models are mathematical structures and they represent due to there being certain mathematical relations between these structures and a mathematical rendering of the target system. (Weisberg includes fictions as convenient *folk ontology* that may serve as a crutch when thinking about the model, but takes them to be ultimately dispensable when it comes to explaining how models relate to the world.) This, however, brings us back to a structuralist theory of representation, and this theory, as we have seen in Sect. 3.4, is far from unproblematic. So it is at best an open question whether getting rid of fiction provides an obvious advantage.

3.7 Representation-as

In this section we discuss approaches that depart from *Goodman*'s notion of *representation-as* [3.64]. In his account of aesthetic representation the idea is that a work of art does not just denote its subject, but moreover it represents it as being thus or so. *Elgin* [3.34] further developed this account and, crucially, suggested that it also applies to scientific representations. This is a vital insight and it provides the entry point to what we think of as the most promising account of epistemic representation.

In this section we present Goodman and Elgin's notion of *representation-as*, and outline how it is a complex type of reference involving a mixture of denotation and what they call exemplification. We introduce the term of art *representation-as* to indicate that we are talking about the specific concept that emerges from Goodman's and Elgin's writings. We then discuss how the account needs to be developed in the context of scientific representation. And finally we present our own answer to the ER-problem, and demonstrate how it answers the questions laid out in Sect. 3.1.

3.7.1 Exemplification and Representation-as

Many instances of epistemic representation are instances of representation-as. Caricatures are paradigmatic examples: Churchill is represented as a bulldog, Thatcher is represented as a boxer, and the Olympic Stadium is represented as a UFO. Using these caricatures we can attempt to learn about their targets: attempt to learn about a politician's personality or a building's appearance. The notion applies beyond caricatures. Holbein's *Portrait of Henry VIII* represents Henry as imposing and powerful and Stoddart's statue of David Hume represents him as thoughtful and wise. The leading idea is that scientific representation works in much the same way. A model of the solar system represents the sun as perfect sphere; the logistic model of growth represents the population as reproducing at fixed intervals of time; and so on. In each instance, models can be used to attempt to learn about their targets by determining what the former represent the latter as being. So representation-as relates, in a way to be made more specific below, to the surrogative reasoning condition discussed in Sect. 3.1.

The locution of representation-as functions in the following way: An object X (e.g., a picture, statue, or model) represents a subject Y (e.g., a person or target system) as being thus or so (Z). The question then is what establishes this sort of representational relationship? The answer requires presenting some of the tools Goodman and Elgin use to develop their account of representation-as.

One of the central posits of *Goodman*'s account is that denotation is "the core of representation" [3.64, p. 5]. Stoddart's statue of David Hume denotes Hume and a model of the solar system denotes the solar system. In that sense the statue and the model are representations *of* their respective targets. To distinguish representation of something from other notions of representation we introduce the technical term *representation-of*. Denotation is what establishes representation-of. (For a number of qualifications and caveats about denotation see our [3.238, Sect. 2]).

Not all representations are a representation-of. A picture showing a unicorn is not a representation-of a unicorn because things that don't exist can't be denoted. Yet there is a clear sense in which such a picture is a representation. *Goodman* and *Elgin*'s solution to this is to distinguish between being a representation-of something and being a something-representation ([3.34, pp. 1–2], [3.64, pp. 21–26]). What makes a picture a something-representation (despite the fact it may fail to denote anything) is that it is the sort of symbol that denotes. *Elgin* argues [3.34, pp. 1–2]:

> "A picture that portrays a griffin, a map that maps the route to Mordor [...] are all representations, although they do not represent anything. To be a representation, a symbol need not itself denote, but it needs to be the sort of symbol that denotes. Griffin pictures are representations then because they are animal pictures, and some animal pictures denote animals. Middle Earth maps are representations because they are maps and some maps denote real locations. [...] So whether a symbol is a representation is a question of what kind of symbol it is."

These representations can be classified into genres, in a way that does not depend on what they are

Part A | 3.7

representations-of (since some may fail to denote), but instead on what they portray. In the case of pictures, this is fairly intuitive (how this is to be developed in the case of scientific models is discussed below). If a picture portrays a man, it is a man-representation, if it portrays a griffin it is a griffin-representation and so on. In general, a picture X is Z-representation if it portrays Z. The crucial point is that this does not presuppose that X be a representation-of Z; indeed X can be Z-representation without denoting anything. A picture must denote a man to be a representation-of a man. But it need not denote anything to be a man-representation.

The next notion we need to introduce is *exemplification*. An item exemplifies a property if it at once instantiates the property and refers to it [3.64, p. 53]:

> "Exemplification is possession plus reference. To have without symbolizing is merely to possess, while to symbolize without having is to refer in some other way than by exemplifying."

Exemplification is a mode of reference that holds between items and properties. In the current context properties are to be understood in the widest possible sense. An item can exemplify one-place properties, multi-place properties (i. e., relations), higher order properties, structural properties, etc. Paradigmatic examples of exemplification are samples. A chip of paint on a manufacturer's sample card both instantiates a certain color, and at the same time refers to that color [3.239, p. 71].

But although exemplification requires instantiation, not every property instantiated by an object is exemplified by it. The chip of paint does not, for example, exemplify its shape or its location on the card. In order to exemplify a property, an object must both instantiate the property and the property itself must be made epistemically salient. How saliency is established will be determined on a case-by-case basis, and we say more about this below.

We can now turn to the conditions under which X represents Y as Z. A first stab would be to say that X represents Y as Z if X is a Z-representation and denotes Y. This however, is not yet good enough. It is important that properties of Z are *transferred* to Y. *Elgin* makes this explicit [3.34, p. 10]:

> "[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 gives a name to the crucial step: imputation. This step can be analyzed in terms of stipulation by a user of a representation. When someone uses X as a representation-as, she has to stipulate that certain properties that are exemplified in X be imputed to Y. We emphasize that imputation does not imply truth: Y may or may not have the properties imputed to it by X. So the representation can be seen as generating a claim about Y that can be true or false; it should not be understood as producing truisms.

Applied to scientific models, the account of epistemic representation that emerges from Goodman and Elgin's discussion of representation can then be summarized as follows:

Definition 3.14 Representation-As

A scientific model M represents a target system T iff:

1. M denotes T
2. M is a Z-representation exemplifying properties $P_1, \ldots, P_n$
3. $P_1, \ldots, P_n$, or related properties, are imputed to T.

It should be added that the first condition can easily be extended to include part-part denotation. In a family portrait the entire portrait denotes the family; at the same time a part of the portrait can denote the mother and another part the father. This is obvious and unproblematic.

We think that this account is on the right track, but all three conditions need to be further developed to furnish a full-fledged account of epistemic representation (at least as applied to scientific models). The developments needed are of different kinds, though. The first condition needs more specificity. How is denotation characterized? What different ways of establishing denotation are there? And how is denotation established in particular cases? These are but some of the questions that a complete account of epistemic representation will have answer. In many cases epistemic representation seems to borrow denotation from linguistic descriptions in which they are embellished and denotation is in effect borrowed from language. So the philosopher of science can turn to the philosophy of language to get a deeper understanding of denotation. This is an interesting project, but it is not one we can pursue here.

In contrast with denotation the other two conditions need to be reformulated because an account molded on visual representations is only an imperfect match for scientific representations. This is the task for the next section.

3.7.2 From Pictures to Models: The Denotation, Exemplification, Keying-up and Imputation Account

According to Goodman and Elgin, for a picture to be a Z-representation it has to be the kind of symbol that denotes. On the face of it, there is a mismatch between pictures and scientific models in this regard. The Schelling model represents social segregation with a checkerboard; billiard balls are used to represent molecules; the Phillips–Newlyn model uses a system of pipes and reservoirs to represent the flow of money through an economy; and the worm *Caenorhabditis elegans* is used as a model of other organisms. But neither checkerboards, billiard balls, pipes, or worms seem to belong to classes of objects that typically denote. The same observation applies to scientific fictions (frictionless planes, utility maximizing agents, and so on) and the mathematical objects used in science. In fact, matrices, curvilinear geometries, Hilbert spaces etc. were all studied as mathematical objects before they became important in the empirical sciences.

Rather than relying on the idea that scientific models belong to classes of objects that typically denote we propose directly introducing an agent and ground representation in this agent's actions. Specific checkerboards, systems of pipes, frictionless places and mathematical structures, are epistemic representations because they are used by an agent to represent a system. When an agent uses an object as a representation, we call it a *base*.

What allows us to classify bases into Z-representations is also less clear in the case of scientific representation. We approach this issue in two steps. The first is to recognize the importance of the intrinsic constitution of the base. Pictures are typically canvases covered with paint. They are classified as Z-representations because under appropriate circumstances the canvas is recognized as portraying a Z. Much can be said about the canvas' material constitution (the thickness or chemical constitution of the paint, etc.), but these are generally of little interest to understanding what the picture portrays. By contrast, the properties of a scientific model – qua material object – do matter. How water flows through the pipes in the Phillips–Newlyn model is crucial to how it represents the movement of money in an economy. That *Caenorhabditis elegans* is a biological organism is of vital importance for how it is used representationally. In fact, models are frequently classified according to what their material base is. We talk about a pipe model of the economy or worm model of cell division because their bases are pipes and worms. Here we introduce a term of art to recognize that scientific models are generally categorized according to their material constitution. An O-object specifies the kind of object something is, qua physical object.

O-objects become representations when they are used as such. But how are they classified as Z-representations? How does the Phillips–Newyln machine become an economy-representation, or how does a collection of billiard balls become a gas-representation? (Again, recall that this is not because they denote economies or gases.) We suggest, and this is the second step, that this requires an act of *interpretation* (notice that we do not use *interpretation* in the same sense as Contessa). In the case of pictures, the nature of this interpretation has been the center of attention for a good while: how one sees a canvas covered with paint as showing a cathedral is regarded by many as one of the important problems of aesthetics. *Schier* [3.240, p. 1] dubbed it the "enigma of depiction", and an entire body of literature is been concerned with it (*Kulvicki* [3.241] provides a useful review). In the case of scientific models we don't think a simple and universal account of how models are interpreted as Z-representations can be given. Interpreting an O-object as a Z-representation requires attributing properties of Zs to the object. How this is done will depend on disciplinary traditions, research interests, background theory and much more. In fact, *interpretation* is a blank to be filled, and it will be filled differently in different cases.

Some examples should help elucidate what we mean by this. In the case of scale models the interpretation is *close* to the O-object in that it interprets the object in its *own* terms. The small car is interpreted as a car-representation and the small ship is interpreted as a ship-representation. Likewise, in the case of the Army Corps' model of the San Francisco bay [3.33], parts of the model bay are interpreted in terms of the real bay. In cases like these, the same predicates that apply to the base (qua O-object) are applied to the object in order to make it into a Z-representation (here $O = Z$). But this is not always the case. For example, the Phillips–Newlyn machine is a system of pipes and reservoirs, but it becomes an economy-representation only when the quantity and flow of water throughout the system are interpreted as the quantity and flow of money throughout an economy. The system is interpreted in terms of predicates that do not apply to the object (qua O-object), but turn it into a Z-representation (here O and Z come apart). In sum, an O-object that has been chosen as the base of a representation becomes a Z-representation if O is interpreted in terms of Z.

Next in line is exemplification. Much can be said about exemplification in general, but the points by and large carry over from the general discussion to the case of models without much ado. There is one difference,

Part A | 3.7

though, in cases like the Phillips–Newlyn machine. Recall that exemplification was defined as the instantiation of a property P by an object in such a way that the object thereby refers to P. How can the Phillips–Newlyn machine exemplify economic properties when it does not, strictly speaking, instantiate them? The crucial point is that nothing in the current account depends on instantiation being literal instantiation. On this point we are in agreement with Goodman and Elgin, whose account relies on nonliteral instantiation. The portrait of Henry cannot, strictly speaking, instantiate the property of being male, even if it represents him as such. *Goodman* and *Elgin* call this metaphorical instantiation ([3.64, pp. 50–51], [3.239, p. 81]).

What matters is that properties are epistemically accessible and salient, and this can be achieved with what we call *instantiation-under-an-interpretation* I, *I-instantiation* for short. An economic interpretation of the Phillips–Newlyn machine interprets amounts of water as amounts of money. It does so by introducing a clearly circumscribed rule of proportionality: x liters of water correspond to y millions of the model-economy's currency. This rule is applied without exception when the machine is interpreted as an economy-representation. So we say that under the economic interpretation I_e the machine I_e-instantiates money properties. With the notion of *I-instantiation* at hand, exemplification poses no problem.

The final issue to clear is the imputation of the model's exemplified properties to the target system. In particular, which properties are so imputed? Elgin describes this as the imputation of the properties exemplified by *M or related ones*. The observation that the properties exemplified by a scientific model and the properties imputed to its target system need not be identical is correct. In fact, few, if any, models in science portray their targets as exhibiting exactly the same features as the model itself. 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 other, and as long as no specific relation is specified it remains unclear which properties are imputed onto the system.

In the context of science, the relation between the properties exemplified and the ones ascribed to the system is sometimes described as one of simplification [3.198, p. 184], idealization [3.198, p. 184] and approximation [3.34, p. 11]. This could suggest that *related ones* means *idealized*, at least in the context of science (we are not attributing this claim to Elgin; we are merely considering the option), perhaps similar to the way in which Ducheyne's account discussed above took target properties to be approximations of model properties. But shifting from *related* to *idealized* or *approximated* (or any of their cognates) makes things worse, not better. For one, *idealization* can mean very different things in different contexts and hence describing the relation between two properties as *idealization* adds little specificity (see *Jones* [3.242] for a discussion of different kinds of idealization). For another, while the relationship between some representation-target properties may be characterized in terms of idealization, many cannot. A map of the world exemplifies a distance of 29 cm between the two points labeled *Paris* and *New York*; the distance between the two cities is 5800 km; but 29 cm is not an idealization of 5800 km. A scale model of a ship being towed through water is not an idealization of an actual ship, at least not in any obvious way. Or in standard representations of Mandelbrod sets the color of a point indicates the speed of divergence of an iterative function for certain parameter value associated with that point, but color is not an idealization of divergence speed.

For this reason it is preferable, in our view, to build a specification of the relationship between model properties and target properties directly into an account of epistemic representation. Let $P_1, \dots, P_n$ be the properties exemplified by M, and let $Q_1, \dots, Q_m$ be the *related* properties that M imputes to Y (where n and m are positive natural numbers that can but need not be equal). Then the representation M must come with a key K that specifies how exactly $P_1, \dots, P_n$ are converted into $Q_1, \dots, Q_m$ [3.50]. Borrowing notation from algebra (somewhat loosely) we can write $K(\langle P_1, \dots, P_n \rangle) = \langle Q_1, \dots, Q_m \rangle$. K can, but need not be, the identity function; any rule that associates a unique set $Q_1, \dots, Q_m$ with $P_1, \dots, P_n$ is admissible. The relevant clause in the definition of representation-as then becomes: M exemplifies $P_1, \dots, P_n$ and the representation imputes properties $Q_1, \dots, Q_m$ to T where the two sets of properties are connected to each other by a key K.

The above examples help illustrate what we have in mind. Let us begin with the example of the map (in fact the idea of a key is motivated by a study of maps; for a discussion of maps see *Galton* [3.243] and *Sismondo* and *Chrisman* [3.244]). P is a measured distance on the map between the point labeled *New York* and the point labeled *Paris*; Q is the distance between New York and Paris in the world; and K is the scale of the map (in the above case, 1 : 20000000). So the key allows us to translate a property of the map (the 29 cm distance) into a property of the world (that New York and Paris are 5800 km apart). But the key involved in the scale model of the ship is more complicated. One of the Ps 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 nonlinear relationship because smaller models encounter disproportionate effects due to the viscosity of the fluid. The exact form of the key is often highly nontrivial and emerges as the result of a thoroughgoing study of the situation; see *Sterrett* [3.245] for a discussion of fluid mechanics. In the representation of the Madelbrod set in [3.246, p. 660] a key is used that translates color into divergence speed [3.246, p. 695]. The square shown is a segment of the complex plane and each point represents a complex number. This number is used as parameter value for an iterative function. If the function converges for number c, then the point in the plane representing c is colored black. If the function diverges, then a shading from yellow over green to blue is used to indicate the speed of divergence, where yellow is slow, green is in the middle and blue is fast.

Neither of these keys is obvious or trivial. Determining how to move from properties exemplified by models to properties of their target systems can be a significant task, and should not go unrecognized in an account of scientific representation. In general K is a blank to be filled, and it depends on a number of factors: the scientific discipline, the context, the aims and purposes for which M is used, the theoretical backdrop against which M operates, etc. Building K into the definition of representation-as does not prejudge the nature of K, much less single out a particular key as the correct one. The requirement merely is that there must be *some* key for M to qualify as a representation-as.

With these modifications in place we can now formulate our own account of representation [3.238, 247]. Consider an agent who chooses an O-object as the base of representation and turns it into Z-representation by adopting an interpretation I. Let M refer to the package of the O-object together with the interpretation I that turns it into a Z-representation. Then:

Definition 3.15 DEKI

A scientific model M represents a target T iff:

1. M denotes T (and, possibly, parts of M denote parts of T)
2. M is a Z-representation exemplifying properties $P_1, \ldots, P_n$
3. M comes with a key, K, specifying how $P_1, \ldots, P_n$ are translated into a set of features $Q_1, \ldots, Q_m$: $K(\langle P_1, \ldots, P_n \rangle) = \langle Q_1, \ldots, Q_m \rangle$
4. The model imputes at least one of the properties $Q_1, \ldots, Q_m$ onto T.

We call this the DEKI account of representation to highlight its key features: denotation, exemplification, keying-up and imputation.

Before highlighting some issues with this account, let us clarify how the account answers the questions we laid out in Sect. 3.1. Firstly, as an answer to the ER-problem, DEKI (Definition 3.15) provides an abstract framework in which to think about epistemic representation. In general, what concretizes each of the conditions needs to be investigated on a case-by-case basis. But far from being a defect, this degree of abstractness is an advantage. *Epistemic representation*, and even the narrower *model-representation*, are umbrella terms 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. DEKI (Definition 3.15) occupies the right middle ground: it is general enough to cover a large array of cases and yet it highlights what all instances of scientific representation have in common. At the same time the account offers an elegant solution to the problem of models without targets: a model that apparently represents Z while there is no Z is a Z-representation but not representation of a Z.

It should be clear how we can use models to perform surrogative reasoning about their targets according to DEKI (Definition 3.15). The account requires that we investigate the properties that are exhibited by the model. These are then translated into a set of properties that are imputed onto the target. This act of imputation supplies a hypothesis about the target system: does it, or does it not, have those properties? This hypothesis does not have to be true, and as such DEKI (Definition 3.15) allows for the possibility of misrepresentation in a straightforward manner.

DEKI's (Definition 3.15) abstract character also allows us to talk about different styles of representation. Style, on the DEKI (Definition 3.15) account, is not a monolithic concept; instead it has several dimensions. Firstly, different O-objects can be chosen. In this way we may speak, say, of the *checkerboard style* and of the *cellular automaton style*. In each case a specific kind of object has been chosen for various modeling 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. Thirdly, different types of keys could be used to characterize different styles. In some instances the key might be the identity key, which would amount to a style of modeling that aims to construct replicas of target systems; in other cases the key incorporates different kinds of idealizations or abstractions, which gives rise to idealization and abstraction keys. But different keys may be associated with entirely different representational styles.

Similarly, DEKI (Definition 3.15) suggests that there is no significant difference between scientific representations and other kinds of epistemic representation, at least at the general level. However, this is not to say that the two cannot be demarcated whatsoever. The sorts of interpretations under which pictures portray *Z*s seem to be different to the sorts of interpretations that are adopted in the scientific framework. Whether or not this can be cashed of more specifically is an interesting question that we cannot investigate here.

Many details in DEKI (Definition 3.15) still need to be spelled out. But the most significant difficulty, perhaps, arises in connection with the problem of ontology. It is not by accident that we have illustrated the account with a physical model, the Phillips–Newlyn machine. Exemplification requires instantiation, which is easily understood for material models, but is highly problematic in the context of nonconcrete models. One option is to view models as fictional entities as discussed in Sect. 3.6. But whether, and if so how, fictional entities instantiate properties is controversially discussed and more philosophical work is needed to make sense of such a notion. It is therefore an open question how this account works for nonconcrete models; for a discussion and a proposal see *Frigg* and *Nguyen* [3.248].

Finally, the account provides us with resources with which to think about the applicability of mathematics. Like the problem of style, various options are available. Firstly, mathematical structures themselves can be taken to be *O*-objects and feature as bases of representation. They can be interpreted on their own terms and therefore exemplify strictly mathematical properties. If one were of a structuralist bent, then the appropriate mathematical properties could be *structural*, which could then be imputed onto the target system (although notice that this approach faces a similar problem to the question of target-end structure discussed in Sect. 3.4.4). Alternatively, the key could provide a translation of these mathematical properties into ones more readily applicable to physical systems. A third alternative would be to take scientific models to be fictional objects, and then adopt an interpretation towards them under which they exemplify mathematical properties. Again, these could be imputed directly onto the target system, or translated into an alternative set of properties. Finally, these fictional models could themselves exemplify physical properties, but in doing so exemplify structural ones as well. Whenever a physical property is exemplified, this provides an extensional relation defined over the objects that instantiate it. The pros and cons of each of these approaches demands further research, but for the purposes of this chapter we simply note that DEKI (Definition 3.15) puts all of these options on the table. Using the framework of *O*-objects, interpretations, exemplification, keys, and imputation provides a novel way in which to think about the applicability of mathematics.

3.8 Envoi

We reviewed theories of epistemic representation. That each approach faces a number of challenges and that there is no consensus on the matter will not have come as a surprise to anybody. We hope, however, that we managed to map the lay of the land and to uncover the fault lines, and thereby aid future discussions.

Acknowledgments. The authors are listed alphabetically; the chapter is fully collaborative. We would like to thank Demetris Portides and Fiora Salis for helpful comments on an earlier draft.

References

3.1 G. Boniolo: *On Scientific Representations: From Kant to a New Philosophy of Science* (Palgrave Macmillan, Hampsire, New York 2007)

3.2 L. Perini: The truth in pictures, Philos. Sci. **72**, 262–285 (2005)

3.3 L. Perini: Visual representation and confirmation, Philos. Sci. **72**, 913–926 (2005)

3.4 L. Perini: Scientific representation and the semiotics of pictures. In: *New Waves in the Philosophy of Science*, ed. by P.D. Magnus, J. Busch (Macmilan, New York 2010) pp. 131–154

3.5 J. Elkins: *The Domain of Images* (Cornell Univ. Press, Ithaca, London 1999)

3.6 K. Warmbrōd: Primitive representation and misrepresentation, Topoi **11**, 89–101 (1992)

3.7 C. Peirce: Principles of philosophy and elements of logic. In: *Collected Papers of Charles Sanders Peirce, Volumes I and II: Principles of Philosophy and Elements of Logic*, ed. by C. Hartshorne, P. Weiss (Harvard Univ. Press, Cambridge 1932)

3.8 E. Tal: Measurement in science. In: *Stanford Encyclopedia of Philosophy*, ed. by E.N. Zalta, http://

plato.stanford.edu/archives/sum2015/entries/measurement-science/ (Summer 2015 Edition)

3.9 T. Knuuttila: Models as Epistemic Artefacts: Toward a Non-Representationalist Account of Scientific Representation, Ph.D. Thesis (Univ. Helsinki, Helsinki 2005)

3.10 T. Knuuttila: Modelling and representing: An artefactual approach to model-based representation, Stud. Hist. Philos. Sci. **42**, 262–271 (2011)

3.11 M. Morgan, M. Morrison (Eds.): *Models as Mediators: Perspectives on Natural and Social Science* (Cambridge Univ. Press, Cambridge 1999)

3.12 S. Hartmann: Models as a tool for theory construction: Some strategies of preliminary physics. In: *Theories and Models in Scientific Processes*, Vol. 44, ed. by W.E. Herfel, W. Krajewski, I. Niiniluoto, R. Wojcicki (Rodopi, Amsterdam, Atlanta 1995) pp. 49–67

3.13 I. Peschard: Making sense of modeling: Beyond representation, Eur. J. Philos. Sci. **1**, 335–352 (2011)

3.14 A. Bokulich: Explanatory fictions. In: *Fictions in Science. Philosophical Essays on Modelling and Idealization*, ed. by M. Suárez (Routledge, London, New York 2009) pp. 91–109

3.15 A.G. Kennedy: A non representationalist view of model explanation, Stud. Hist. Philos. Sci. **43**, 326–332 (2012)

3.16 A.I. Woody: More telltale signs: What attention to representation reveals about scientific explanation, Philos. Sci. **71**, 780–793 (2004)

3.17 J. Reiss: The explanation paradox, J. Econ. Methodol. **19**, 43–62 (2012)

3.18 M. Lynch, S. Woolgar: *Representation in Scientific Practice* (MIT, Cambridge 1990)

3.19 R.N. Giere: No representation without representation, Biol. Philos. **9**, 113–120 (1994)

3.20 R. Frigg: *Models and Representation: Why Structures Are Not Enough*, Measurement in Physics and Economics Project Discussion Paper, Vol. DP MEAS 25/02 (London School of Economics, London 2002)

3.21 R. Frigg: Scientific representation and the semantic view of theories, Theoria **55**, 49–65 (2006)

3.22 M. Morrison: Models as representational structures. In: *Nancy Cartwright's Philosophy of Science*, ed. by S. Hartmann, C. Hoefer, L. Bovens (Routledge, New York 2008) pp. 67–90

3.23 M. Suárez: Scientific representation: Against similarity and isomorphism, Int. Stud. Philos. Sci. **17**, 225–244 (2003)

3.24 S. Laurence, E. Margolis: Concepts and cognitive science. In: *Concepts: Core Readings*, ed. by S. Laurence, E. Margolis (MIT, Cambridge 1999) pp. 3–81

3.25 C. Swoyer: Structural representation and surrogative reasoning, Synthese **87**, 449–508 (1991)

3.26 C. Callender, J. Cohen: There is no special problem about scientific representation, Theoria **55**, 7–25 (2006)

3.27 D.M. Bailer-Jones: When scientific models represent, Int. Stud. Philos. Sci. **17**, 59–74 (2003)

3.28 A. Bolinska: Epistemic representation, informativeness and the aim of faithful representation, Synthese **190**, 219–234 (2013)

3.29 G. Contessa: Scientific representation, interpretation, and surrogative reasoning, Philos. Sci. **74**, 48–68 (2007)

3.30 R. Frigg: Re-Presenting Scientific Represenation, Ph.D. Thesis (London School of Economics and Political Science, London 2003)

3.31 C. Liu: Deflationism on scientific representation. In: *EPSA11 Perspectives and Foundational Problems in Philosophy of Science*, ed. by V. Karakostas, D. Dieks (Springer, Dordrecht 2013) pp. 93–102

3.32 M. Suárez: An inferential conception of scientific representation, Philos. Sci. **71**, 767–779 (2004)

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

3.34 C.Z. Elgin: Telling instances. In: *Beyond Mimesis and Convention: Representation in Art and Science*, ed. by R. Frigg, M.C. Hunter (Springer, Berlin, New York 2010) pp. 1–18

3.35 S. French: A model-theoretic account of representation (or, I don't know much about art ... but I know it involves isomorphism), Philos. Sci. **70**, 1472–1483 (2003)

3.36 B.C. van Fraassen: *Scientific Representation: Paradoxes of Perspective* (Oxford Univ. Press, Oxford 2008)

3.37 A.I. Woody: Putting quantum mechanics to work in chemistry: The power of diagrammatic pepresentation, Philos. Sci. **67**, S612–S627 (2000)

3.38 S. Stich, T. Warfield (Eds.): *Mental Representation: A Reader* (Blackwell, Oxford 1994)

3.39 K. Sterelny, P.E. Griffiths: *Sex and Death: An Introduction to Philosophy of Biology* (Univ. Chicago Press, London, Chicago 1999)

3.40 E. Wigner: The unreasonable effectiveness of mathematics in the natural sciences, Commun. Pure Appl. Math. **13**, 1–14 (1960)

3.41 S. Shapiro: *Philosophy of Mathematics: Structure and Ontology* (Oxford Univ. Press, Oxford 1997)

3.42 O. Bueno, M. Colyvan: An inferential conception of the application of mathematics, Nous **45**, 345–374 (2011)

3.43 A. Levy: Modeling without models, Philos. Stud. **152**, 781–798 (2015)

3.44 I. Hacking: *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science* (Cambridge Univ. Press, Cambridge 1983)

3.45 A. Rosenblueth, N. Wiener: The role of models in science, Philos. Sci. **12**, 316–321 (1945)

3.46 R.A. Ankeny, S. Leonelli: What's so special about model organisms?, Stud. Hist. Philos. Sci. **42**, 313–323 (2011)

3.47 U. Klein (Ed.): *Tools and Modes of Representation in the Laboratory Sciences* (Kluwer, London, Dordrecht 2001)

3.48 A. Toon: Models as make-believe. In: *Beyond Mimesis and Convention: Representation in Art and Science*, ed. by R. Frigg, M. Hunter (Springer, Berlin 2010) pp. 71–96

3.49 A. Toon: Similarity and scientific representation, Int. Stud. Philos. Sci. **26**, 241–257 (2012)

3.50 R. Frigg: Fiction and scientific representation. In: *Beyond Mimesis and Convention: Representation in Art and Science*, ed. by R. Frigg, M. Hunter (Springer, Berlin, New York 2010) pp. 97–138

3.51 R.N. Giere: An agent-based conception of models and scientific representation, Synthese **172**, 269–281 (2010)

3.52 P. Teller: Twilight of the perfect model model, Erkenntnis **55**, 393–415 (2001)

3.53 O. Bueno, S. French: How theories represent, Br. J. Philos. Sci. **62**, 857–894 (2011)

3.54 A.F. MacKay: Mr. Donnellan and Humpty Dumpty on referring, Philos. Rev. **77**, 197–202 (1968)

3.55 K.S. Donnellan: Putting Humpty Dumpty together again, Philos. Rev. **77**, 203–215 (1968)

3.56 E. Michaelson: This and That: A Theory of Reference for Names, Demonstratives, and Things in Between, Ph.D. Thesis (Univ. California, Los Angels 2013)

3.57 M. Reimer, E. Michaelson: Reference. In: *Stanford Encycloledia of Philosophy*, ed. by E.N. Zalta, http://plato.stanford.edu/archives/win2014/entries/reference/ (Winter Edition 2014)

3.58 C. Abell: Canny resemblance, Philos. Rev. **118**, 183–223 (2009)

3.59 D. Lopes: *Understanding Pictures* (Oxford Univ. Press, Oxford 2004)

3.60 R.N. Giere: How models are used to represent reality, Philos. Sci. **71**, 742–752 (2004)

3.61 R.N. Giere: Visual models and scientific judgement. In: *Picturing Knowledge: Historical and Philosophical Problems Concerning the Use of Art in Science*, ed. by B.S. Baigrie (Univ. Toronto Press, Toronto 1996) pp. 269–302

3.62 B. Kralemann, C. Lattmann: Models as icons: Modeling models in the semiotic framework of Peirce's theory of signs, Synthese **190**, 3397–3420 (2013)

3.63 R. Frigg, S. Bradley, H. Du, L.A. Smith: Laplace's demon and the adventures of his apprentices, Philos. Sci. **81**, 31–59 (2014)

3.64 N. Goodman: *Languages of Art* (Hacket, Indianapolis, Cambridge 1976)

3.65 A. Yaghmaie: Reflexive, symmetric and transitive scientific representations, http://philsci-archive.pitt.edu/9454 (2012)

3.66 A. Tversky, I. Gati: Studies of similarity. In: *Cognition and Categorization*, ed. by E. Rosch, B. Lloyd (Lawrence Elbaum Associates, Hillside New Jersey 1978) pp. 79–98

3.67 M. Poznic: Representation and similarity: Suárez on necessary and sufficient conditions of scientific representation, J. Gen. Philos. Sci. (2015), doi:10.1007/s10838-015-9307-7

3.68 H. Putnam: *Reason, Truth, and History* (Cambridge Univ. Press, Cambridge 1981)

3.69 M. Black: How do pictures represent? In: *Art, Perception, and Reality*, ed. by E. Gombrich, J. Hochberg, M. Black (Johns Hopkins Univ. Press, London, Baltimore 1973) pp. 95–130

3.70 J.L. Aronson, R. Harré, E. Cornell Way: *Realism Rescued: How Scientific Progress is Possible* (Open Court, Chicago 1995)

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

3.72 S. Ducheyne: Towards an ontology of scientific models, Metaphysica **9**, 119–127 (2008)

3.73 N. Cartwright: *The Dappled World: A Study of the Boundaries of Science* (Cambridge Univ. Press, Cambridge 1999)

3.74 N. Cartwright: Models and the limits of theory: Quantum hamiltonians and the BCS models of superconductivity. In: *Models as Mediators: Perspectives on Natural and Social Science*, ed. by M. Morgan, M. Morrison (Cambridge Univ. Press, Cambridge 1999) pp. 241–281

3.75 L. Apostel: Towards the formal study of models in the non-formal sciences. In: *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences*, ed. by H. Freudenthal (Reidel, Dordrecht 1961) pp. 1–37

3.76 A.-M. Rusanen, O. Lappi: An information semantic account of scientific models. In: *EPSA Philosophy of Science: Amsterdam 2009*, ed. by H.W. de Regt, S. Hartmann, S. Okasha (Springer, Dordrecht 2012) pp. 315–328

3.77 B.C. van Fraassen: *The Empirical Stance* (Yale Univ. Press, New Haven, London 2002)

3.78 H. Putnam: *The Collapse of the Fact-Value Distinction* (Harvard Univ. Press, Cambridge 2002)

3.79 U. Mäki: Models and the locus of their truth, Synthese **180**, 47–63 (2011)

3.80 S.M. Downes: Models, pictures, and unified accounts of representation: Lessons from aesthetics for philosophy of science, Perspect. Sci. **17**, 417–428 (2009)

3.81 M. Morreau: It simply does not add up: The trouble with overall similarity, J. Philos. **107**, 469–490 (2010)

3.82 W.V.O. Quine: *Ontological Relativity and Other Essays* (Columbia Univ. Press, New York 1969)

3.83 N. Goodman: Seven strictures on similarity. In: *Problems and Projects*, ed. by N. Goodman (Bobbs-Merrill, Indianapolis, New York 1972) pp. 437–446

3.84 L. Decock, I. Douven: Similarity after Goodman, Rev. Philos. Psychol. **2**, 61–75 (2011)

3.85 R.N. Shepard: Multidimensional scaling, tree-fitting, and clustering, Science **210**, 390–398 (1980)

3.86 A. Tversky: Features of similarity, Psychol. Rev. **84**, 327–352 (1977)

3.87 M. Weisberg: Getting serious about similarity, Philos. Sci. **79**, 785–794 (2012)

3.88 M. Hesse: *Models and Analogies in Science* (Sheed Ward, London 1963)

3.89 W. Parker: Getting (even more) serious about similarity, Biol. Philos. **30**, 267–276 (2015)

3.90 I. Niiniluoto: Analogy and similarity in scientific reasoning. In: *In Analogical Reasoning: Perspectives of Artificial Intelligence, Cognitive Science, and Philosophy*, ed. by D.H. Helman (Kluwer, Dordrecht 1988) pp. 271–298

3.91 M. Weisberg: Biology and philosophy symposium on simulation and similarity: Using models to understand the world: Response to critics, Biol. Philos. **30**, 299–310 (2015)

3.92 A. Toon: Playing with molecules, Stud. Hist. Philos. Sci. **42**, 580–589 (2011)

3.93 M. Morgan, T. Knuuttila: Models and modelling in economics. In: *Philosophy of Economics*, ed. by U. Mäki (Elsevier, Amsterdam 2012) pp. 49–87

3.94 M. Thomson-Jones: Modeling without mathematics, Philos. Sci. **79**, 761–772 (2012)

3.95 G. Rosen: Abstract objects. In: *The Stanford Encyclopedia of Philosophy*, ed. by E.N. Zalta, http://plato.stanford.edu/archives/fall2014/entries/abstract-objects/ (Fall 2014 Edition)

3.96 S. Hale: Spacetime and the abstract–concrete distinction, Philos. Stud. **53**, 85–102 (1988)

3.97 R.N. Giere: Why scientific models should not be regarded as works of fiction. In: *Fictions in Science. Philosophical Essays on Modelling and Idealization*, ed. by M. Suárez (Routledge, London 2009) pp. 248–258

3.98 M. Thomson-Jones: Missing systems and face value practise, Synthese **172**, 283–299 (2010)

3.99 D.M. Armstrong: *Universals: An Opinionated Introduction* (Westview, London 1989)

3.100 P. Suppes: *Representation and Invariance of Scientific Structures* (CSLI Publications, Stanford 2002)

3.101 B.C. van Fraassen: *The Scientific Image* (Oxford Univ. Press, Oxford 1980)

3.102 N.C.A. Da Costa, S. French: *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning* (Oxford Univ. Press, Oxford 2003)

3.103 H. Byerly: Model-structures and model-objects, Br. J. Philos. Sci. **20**, 135–144 (1969)

3.104 A. Chakravartty: The semantic or model-theoretic view of theories and scientific realism, Synthese **127**, 325–345 (2001)

3.105 C. Klein: Multiple realizability and the semantic view of theories, Philos. Stud. **163**, 683–695 (2013)

3.106 D. Portides: Scientific models and the semantic view of theories, Philos. Sci. **72**, 1287–1289 (2005)

3.107 D. Portides: Why the model-theoretic view of theories does not adequately depict the methodology of theory application. In: *EPSA Epistemology and Methodology of Science*, ed. by M. Suárez, M. Dorato, M. Rédei (Springer, Dordrecht 2010) pp. 211–220

3.108 M.D. Resnik: *Mathematics as a Science of Patterns* (Oxford Univ. Press, Oxford 1997)

3.109 S. Shapiro: *Thinking About Mathematics* (Oxford Univ. Press, Oxford 2000)

3.110 M. Thomson-Jones: Structuralism about scientific representation. In: *Scientific Structuralism*, ed. by A. Bokulich, P. Bokulich (Springer, Dordrecht 2011) pp. 119–141

3.111 M. Machover: *Set Theory, Logic and Their Limitations* (Cambridge Univ. Press, Cambridge 1996)

3.112 W. Hodges: *A Shorter Model Theory* (Cambridge Univ. Press, Cambridge 1997)

3.113 C.E. Rickart: *Structuralism and Structure: A Mathematical Perspective* (World Scientific Publishing, Singapore 1995)

3.114 G.S. Boolos, R.C. Jeffrey: *Computability and Logic* (Cambridge Univ. Press, Cambridge 1989)

3.115 B. Russell: *Introduction to Mathematical Philosophy* (Routledge, London, New York 1993)

3.116 H.B. Enderton: *A Mathematical Introduction to Logic* (Harcourt, San Diego, New York 2001)

3.117 P. Suppes: A comparison of the meaning and uses of models in mathematics and the empirical sciences. In: *Studies in the Methodology and Foundations of Science: Selected Papers from 1951 to 1969*, ed. by P. Suppes (Reidel, Dordrecht 1969) pp. 10–23, 1960

3.118 B.C. van Fraassen: Structure and perspective: Philosophical perplexity and paradox. In: *Logic and Scientific Methods*, ed. by M.L. Dalla Chiara (Kluwer, Dordrecht 1997) pp. 511–530

3.119 M. Redhead: The intelligibility of the universe. In: *Philosophy at the New Millennium*, ed. by A. O'Hear (Cambridge Univ. Press, Cambridge 2001)

3.120 S. French, J. Ladyman: Reinflating the semantic approach, Int. Stud. Philos. Sci. **13**, 103–121 (1999)

3.121 N.C.A. Da Costa, S. French: The model-theoretic approach to the philosophy of science, Philos. Sci. **57**, 248–265 (1990)

3.122 P. Suppes: Models of data. In: *Studies in the Methodology and Foundations of Science: Selected Papers from 1951 to 1969*, ed. by P. Suppes (Reidel, Dordrecht 1969) pp. 24–35, 1962

3.123 P. Suppes: *Set-Theoretical Structures in Science* (Stanford Univ., Stanford 1970), lecture notes

3.124 B.C. van Fraassen: *Quantum Mechanics: An Empiricist View* (Oxford Univ. Press, Oxford 1991)

3.125 B.C. van Fraassen: A philosophical approach to foundations of science, Found. Sci. **1**, 5–9 (1995)

3.126 N.C.A. Da Costa, S. French: Models, theories, and structures: Thirty years on, Philos. Sci. **67**, 116–127 (2000)

3.127 M. Dummett: *Frege: Philosophy of Mathematics* (Duckworth, London 1991)

3.128 G. Hellman: *Mathematics Without Numbers: Towards a Modal-Structural Interpretation* (Oxford Univ. Press, Oxford 1989)

3.129 G. Hellman: Structuralism without structures, Philos. Math. **4**, 100–123 (1996)

3.130 O. Bueno, S. French, J. Ladyman: On representing the relationship between the mathematical and the empirical, Philos. Sci. **69**, 452–473 (2002)

3.131 S. French: Keeping quiet on the ontology of models, Synthese **172**, 231–249 (2010)

3.132 S. French, J. Saatsi: Realism about structure: The semantic view and nonlinguistic representations, Philos. Sci. **73**, 548–559 (2006)

3.133 S. French, P. Vickers: Are there no things that are scientific theories?, Br. J. Philos. Sci. **62**, 771–804 (2011)

3.134 E. Landry: Shared structure need not be shared set-structure, Synthese **158**, 1–17 (2007)

3.135 H. Halvorson: What scientific theories could not be, Philos. Sci. **79**, 183–206 (2012)

3.136 H. Halvorson: Scientific theories. In: *The Oxford Handbook of Philosophy of Science*, ed. by P. Humphreys (Oxford Univ. Press, Oxford 2016)

3.137 K. Brading, E. Landry: Scientific structuralism: Presentation and representation, Philos. Sci. **73**, 571–581 (2006)

3.138 C. Glymour: Theoretical equivalence and the semantic view of theories, Philos. Sci. **80**, 286–297 (2013)

3.139 J.B. Ubbink: Model, description and knowledge, Synthese **12**, 302–319 (1960)

3.140 A. Bartels: Defending the structural concept of representation, Theoria **21**, 7–19 (2006)

3.141 E. Lloyd: A semantic approach to the structure of population genetics, Philos. Sci. **51**, 242–264 (1984)

3.142 B. Mundy: On the general theory of meaningful representation, Synthese **67**, 391–437 (1986)

3.143 S. French: The reasonable effectiveness of mathematics: Partial structures and the application of group theory to physics, Synthese **125**, 103–120 (2000)

3.144 O. Bueno: Empirical adequacy: A partial structure approach, Stud. Hist. Philos. Sci. **28**, 585–610 (1997)

3.145 O. Bueno: What is structural empiricism? Scientific change in an empiricist setting, Erkenntnis **50**, 59–85 (1999)

3.146 F. Pero, M. Suárez: Varieties of misrepresentation and homomorphism, Eur. J. Philos. Sci. **6**(1), 71–90 (2016)

3.147 P. Kroes: Structural analogies between physical systems, Br. J. Philos. Sci. **40**, 145–154 (1989)

3.148 F.A. Muller: Reflections on the revolution at Stanford, Synthese **183**, 87–114 (2011)

3.149 E.W. Adams: The foundations of rigid body mechanics and the derivation of its laws from those of particle mechanics. In: *The Axiomatic Method: With Special Reference to Geometry and Physics*, ed. by L. Henkin, P. Suppes, A. Tarski (North-Holland, Amsterdam 1959) pp. 250–265

3.150 O. Bueno: Models and scientific representations. In: *New Waves in Philosophy of Science*, ed. by P.D. Magnus, J. Busch (Pelgrave MacMillan, Hampshire 2010) pp. 94–111

3.151 M. Budd: How pictures look. In: *Virtue and Taste*, ed. by D. Knowles, J. Skorupski (Blackwell, Oxford 1993) pp. 154–175

3.152 P. Godfrey-Smith: The strategy of model-based science, Biol. Philos. **21**, 725–740 (2006)

3.153 T. Harris: Data models and the acquisition and manipulation of data, Philos. Sci. **70**, 1508–1517 (2003)

3.154 B.C. van Fraassen: Theory construction and experiment: An empiricist view, Proc. Philos. Sci. **2**, 663–677 (1981)

3.155 B.C. van Fraassen: *Laws and Symmetry* (Clarendon, Oxford 1989)

3.156 B.C. van Fraassen: Empricism in the philosophy of science. In: *Images of Science: Essays on Realism and Empiricism with a Reply from Bas C. van Fraassen*, ed. by P.M. Churchland, C.A. Hooker (Univ. Chicago Press, London, Chicago 1985) pp. 245–308

3.157 J. Bogen, J. Woodward: Saving the phenomena, Philos. Rev. **97**, 303–352 (1988)

3.158 J. Woodward: Data and phenomena, Synthese **79**, 393–472 (1989)

3.159 P. Teller: Whither constructive empiricism, Philos. Stud. **106**, 123–150 (2001)

3.160 J.W. McAllister: Phenomena and patterns in data sets, Erkenntnis **47**, 217–228 (1997)

3.161 J. Nguyen: On the pragmatic equivalence between representing data and phenomena, Philos. Sci. **83**, 171–191 (2016)

3.162 M. Frisch: Users, structures, and representation, Br. J. Philos. Sci. **66**, 285–306 (2015)

3.163 W. Balzer, C.U. Moulines, J.D. Sneed: *An Architectonic for Science the Structuralist Program* (D. Reidel, Dordrecht 1987)

3.164 W. Demopoulos: On the rational reconstruction of our theoretical knowledge, Br. J. Philos. Sci. **54**, 371–403 (2003)

3.165 J. Ketland: Empirical adequacy and ramsification, Br. J. Philos. Sci. **55**, 287–300 (2004)

3.166 R. Frigg, I. Votsis: Everything you always wanted to know about structural realism but were afraid to ask, Eur. J. Philos. Sci. **1**, 227–276 (2011)

3.167 P. Ainsworth: Newman's objection, Br. J. Philos. Sci. **60**, 135–171 (2009)

3.168 S. Shapiro: Mathematics and reality, Philos. Sci. **50**, 523–548 (1983)

3.169 S. French: *The Structure of the World. Metaphysics and Representation* (Oxford Univ. Press, Oxford 2014)

3.170 M. Tegmark: The mathematical universe, Found. Phys. **38**, 101–150 (2008)

3.171 M. Suárez, A. Solé: On the analogy between cognitive representation and truth, Theoria **55**, 39–48 (2006)

3.172 M. Suárez: Deflationary representation, inference, and practice, Stud. Hist. Philos. Sci. **49**, 36–47 (2015)

3.173 A. Chakravartty: Informational versus functional theories of scientific representation, Synthese **172**, 197–213 (2010)

3.174 W. Künne: *Conceptions of Truth* (Clarendon, Oxford 2003)

3.175 R.B. Brandom: *Making it Explicit: Reasoning, Representing and Discursive Commitment* (Harvard Univ. Press, Cambridge 1994)

3.176 R.B. Brandom: *Articulating Reasons: An Introduction to Inferentialism* (Harvard Univ. Press, Cambridge 2000)

3.177 X. de Donato Rodriguez, J. Zamora Bonilla: Credibility, idealisation, and model building: An inferential approach, Erkenntnis **70**, 101–118 (2009)

3.178 M. Suárez: Scientific Representation, Philos. Compass **5**, 91–101 (2010)

3.179 G. Contessa: Scientific models and representation. In: *The Continuum Companion to the Philosophy of Science*, ed. by S. French, J. Saatsi (Continuum Press, London 2011) pp. 120–137

3.180 G. Contessa: Scientific models and fictional objects, Synthese **172**, 215–229 (2010)

3.181 E. Shech: Scientific misrepresentation and guides to ontology: The need for representational code and contents, Synthese **192**(11), 3463–3485 (2015)

3.182 S. Ducheyne: Scientific representations as limiting cases, Erkenntnis **76**, 73–89 (2012)

3.183 R.F. Hendry: Models and approximations in quantum chemistry. In: *Idealization IX: Idealization in Contemporary Physics*, ed. by N. Shanks (Rodopi, Amsterdam 1998) pp. 123–142

3.184 R. Laymon: Computer simulations, idealizations and approximations, Proc. Bienn. Meet. Philos. Sci. Assoc., Vol. 2 (1990) pp. 519–534

3.185 C. Liu: Explaining the emergence of cooperative phenomena, Philos. Sci. **66**, S92–S106 (1999)

3.186 J. Norton: Approximation and idealization: Why the difference matters, Philos. Sci. **79**, 207–232 (2012)

3.187 J.L. Ramsey: Approximation. In: *The Philosophy of Science: An Encyclopedia*, ed. by S. Sarkar, J. Pfeifer (Routledge, New York 2006) pp. 24–27

3.188 R.I.G. Hughes: Models and representation, Philos. Sci. **64**, S325–S336 (1997)

3.189 R.I.G. Hughes: *The Theoretical Practises of Physics: Philosophical Essays* (Oxford Univ. Press, Oxford 2010)

3.190 R.I.G. Hughes: Laws of nature, laws of physics, and the representational account of theories, ProtoSociology **12**, 113–143 (1998)

3.191 L.A. Smith: *Chaos: A Very Short Introduction* (Oxford Univ. Press, Oxford 2007)

3.192 W.D. Niven (Ed.): *The Scientific Papers of James Clerk Maxwell* (Dover Publications, New York 1965)

3.193 H. Vaihinger: *The Philosophy of as if: A System of the Theoretical, Practical, and Religious Fictions of Mankind* (Kegan Paul, London 1911) p. 1924, English translation

3.194 N. Cartwright: *How the Laws of Physics Lie* (Oxford Univ. Press, Oxford 1983)

3.195 D.N. McCloskey: Storytelling in economics. In: *Narrative in Culture. The uses of Storytelling in the Sciences, Philosophy, and Literature*, ed. by C. Nash (Routledge, London 1990) pp. 5–22

3.196 A. Fine: Fictionalism, Midwest Stud. Philos. **18**, 1–18 (1993)

3.197 L. Sklar: *Theory and Truth. Philosophical Critique Within Foundational Science* (Oxford Univ. Press, Oxford 2000)

3.198 C.Z. Elgin: *Considered Judgement* (Princeton Univ. Press, Princeton 1996)

3.199 S. Hartmann: Models and stories in hadron physics. In: *Models as Mediators. Perspectives on Natural and Social Science*, ed. by M. Morgan, M. Morrison (Cambridge Univ. Press, Cambridge 1999) pp. 326–346

3.200 M. Morgan: Models, stories and the economic world, J. Econ. Methodol. **8**, 361–384 (2001)

3.201 M. Morgan: Imagination and imaging in model building, Philos. Sci. **71**, 753–766 (2004)

3.202 R. Sugden: Credible worlds: The status of theoretical models in economics, J. Econ. Methodol. **7**, 1–31 (2000)

3.203 R. Frigg: Models and fiction, Synthese **172**, 251–268 (2010)

3.204 T. Grüne-Yanoff, P. Schweinzer: The roles of stories in applying game theory, J. Econ. Methodol. **15**, 131–146 (2008)

3.205 A. Toon: *Models as Make-Believe. Imagination, Fiction and Scientific Representation* (Palgrave Macmillan, Basingstoke 2012)

3.206 A. Levy: Models, fictions, and realism: Two packages, Philos. Sci. **79**, 738–748 (2012)

3.207 S. Friend: Fictional characters, Philos. Compass **2**, 141–156 (2007)

3.208 F. Salis: Fictional entities. In: *Online Companion to Problems in Aanalytical Philosophy*, ed. by J. Branquinho, R. Santos, doi:10.13140/2.1.1931.9040 (2014)

3.209 P. Godfrey-Smith: Models and fictions in science, Philos. Stud. **143**, 101–116 (2009)

3.210 R. Sugden: Credible worlds, capacities and mechanisms, Erkenntnis **70**, 3–27 (2009)

3.211 J. Cat: Who's afraid of scientific fictions?: Mauricio Suárez (Ed.): Fictions in Science. Philosophical Essays on Modeling and Idealization, J. Gen. Philos. Sci. **43**, 187–194 (2012), book review

3.212 C. Liu: A Study of model and representation based on a Duhemian thesis. In: *Philosophy and Cognitive Science: Western and Eastern studies*, ed. by L. Magnani, P. Li (Springer, Berlin, Heidelberg 2012) pp. 115–141

3.213 C. Liu: Symbolic versus modelistic elements in scientific modeling, Theoria **30**, 287–300 (2015)

3.214 C. Pincock: *Mathematics and Scientific Representation* (Oxford Univ. Press, Oxford 2012)

3.215 M. Vorms: Representing with imaginary models: Formats matter, Stud. Hist. Philos. Sci. **42**, 287–295 (2011)

3.216 M. Vorms: Formats of representation in scientific theorising. In: *Models, Simulations, and Representations*, ed. by P. Humphreys, C. Imbert (Routledge, New York 2012) pp. 250–274

3.217 T. Knuuttila, M. Boon: How do models give us knowledge? The case of Carnot's ideal heat engine, Eur. J. Philos. Sci. **1**, 309–334 (2011)

3.218 R. Frigg: Fiction in science. In: *Fictions and Models: New Essays*, ed. by J. Woods (Philiosophia, Munich 2010) pp. 247–287

3.219 M.E. Kalderon (Ed.): *Fictionalism in Metaphysics* (Oxford Univ. Press, Oxford 2005)

3.220 A. Fine: Fictionalism. In: *Routledge Encyclopedia of Philosophy*, ed. by E. Craig (Routledge, London 1998)

3.221 A. Fine: Science fictions: Comment on Godfrey-Smith, Philos. Stud. **143**, 117–125 (2009)

3.222 E. Winsberg: A function for fictions: Expanding the scope of science. In: *Fictions in Science: Philosophical Essays in on Modeling and Idealization*, ed. by M. Suárez (Routledge, New York 2009) pp. 179–191

3.223 M. Suárez: Scientific fictions as rules of inference. In: *Fictions in Science: Philosophical Essays in on Modeling and Idealization*, ed. by M. Suárez (Routledge, New York 2009) pp. 158–178

3.224 M. Morrison: Fictions, representations, and reality. In: *Fictions in Science: Philosophical Essays on Modeling and Idealization*, ed. by M. Suárez (Routledge, New York 2009) pp. 110–135

3.225 G.M. Purves: Finding truth in fictions: Identifying non-fictions in imaginary cracks, Synthese **190**, 235–251 (2013)

3.226 J. Woods: Against fictionalism. In: *Model-Based Reasoning in Science and Technology: Theoretical and Cognitive Issues*, ed. by L. Magnani (Springer, Berlin, Heidelberg 2014) pp. 9–42

3.227 M. Weisberg: Who is a modeler?, Br. J. Philos. Sci. **58**, 207–233 (2007)

3.228 A. Toon: The ontology of theoretical modelling: Models as make-believe, Synthese **172**, 301–315 (2010)

3.229 K.L. Walton: *Mimesis as Make-Believe: On the Foundations of the Representational Arts* (Harvard Univ. Press, Cambridge 1990)

3.230 S. Yablo: *Aboutness* (Princeton Univ. Press, Princeton 2014)

3.231 N. Cartwright: Models: Parables v fables. In: *Beyond Mimesis and Convention. Representation in Art and Science*, ed. by R. Frigg, M.C. Hunter (Springer, Berlin, New York 2010) pp. 19–32

3.232 T. Schelling: *Micromotives and Macrobehavior* (Norton, New York 1978)

3.233 C.A. Pissarides: Loss of skill during unemployment and the persistence of unemployment shocks, Q. J. Econ. **107**, 1371–1391 (1992)

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

3.235 T. Knuuttila: Isolating representations versus credible constructions? Economic modelling in theory and practice, Erkenntnis **70**, 59–80 (2009)

3.236 L. Magnani: Scientific models are not fictions: Model-based science as epistemic warfare. In: *Philosophy and Cognitive Science: Western and Eastern Studies*, ed. by L. Magnani, P. Li (Springer, Berlin, Heidelberg 2012) pp. 1–38

3.237 J. Odenbaugh: Semblance or similarity?, Reflections on simulation and similarity, Biol. Philos. **30**, 277–291 (2015)

3.238 R. Frigg, J. Nguyen: Scientific representation is representation as. In: *Philosophy of Science in Practice: Nancy Cartwright and the Nature of Scientific Reasoning*, ed. by H.-K. Chao, R. Julian, C. Szu-Ting (Springer, New York 2017), in press

3.239 C.Z. Elgin: *With Reference to Reference* (Hackett, Indianapolis 1983)

3.240 F. Schier: *Deeper in Pictures: An Essay on Pictorial Representation* (Cambridge Univ. Press, Cambridge 1986)

3.241 J. Kulvicki: Pictorial representation, Philos. Compass **1**, 535–546 (2006)

3.242 M. Jones: Idealization and abstraction: A framework. In: *Idealization XII: Correcting the Model-Idealization and Abstraction in the Sciences*, ed. by M. Jones, N. Cartwright (Rodopi, Amsterdam 2005) pp. 173–218

3.243 A. Galton: Space, time, and the representation of geographical reality, Topoi **20**, 173–187 (2001)

3.244 S. Sismondo, N. Chrisman: Deflationary metaphysics and the nature of maps, Proc. Philos. Sci. **68**, 38–49 (2001)

3.245 S.G. Sterrett: Models of machines and models of phenomena, Int. Stud. Philos. Sci. **20**, 69–80 (2006)

3.246 J.H. Argyris, G. Faust, M. Haase: *Die Erforschung des Chaos: Eine Einführung für Naturwissenschaftler und Ingenieure* (Vieweg Teubner, Braunschweig 1994)

3.247 R. Frigg, J. Nguyen: *The Turn of the Valve: Representing with Material Models*, Unpublished Manuscript

3.248 R. Frigg, J. Nguyen: The fiction view of models reloaded, forthcoming in The Monist, July 2016

4. Models and Explanation

Alisa Bokulich

Detailed examinations of scientific practice have revealed that the use of idealized models in the sciences is pervasive. These models play a central role in not only the investigation and prediction of phenomena, but also in their received scientific explanations. This has led philosophers of science to begin revising the traditional philosophical accounts of scientific explanation in order to make sense of this practice. These new model-based accounts of scientific explanation, however, raise a number of key questions: Can the fictions and falsehoods inherent in the modeling practice do real explanatory work? Do some highly abstract and mathematical models exhibit a noncausal form of scientific explanation? How can one distinguish an exploratory *how-possibly* model explanation from a genuine *how-actually* model explanation? Do modelers face tradeoffs such that a model that is optimized for yielding explanatory insight, for example, might fail to be the most predictively accurate, and vice versa? This chapter explores the various answers that have been given to these questions.

Explanation is one of the central aims of science, and the attempt to understand the nature of scientific explanation is at the heart of the philosophy of science. An explanation can be analyzed as consisting of two parts, a phenomenon or event to be explained, known as the *explanandum*, and that which does the job of explaining, the *explanans*. On the traditional approach, to explain a phenomenon is either to deduce the explanandum phenomenon from the relevant laws of nature and initial conditions, such as on the deductive-nomological (DN) account [4.1], or to trace the detailed causal chain leading up to that event, such as on the causal–mechanical account [4.2]. Underlying this traditional approach are the assumptions that, in order to genuinely explain, the explanans must be entirely true, and that the more complete and detailed the explanans is, the better the scientific explanation.

As philosophers of science have turned to more careful examinations of actual scientific practice, however, there have been three key observations that have challenged this traditional approach: first, many of the phenomena scientists seek to explain are incredibly complex; second, the laws of nature supposedly needed for explanation are either few and far between or entirely absent in many of the sciences; and third, a detailed causal description of the chain of events and interactions leading up to a phenomenon are often either beyond our grasp or not in fact what is most important for a scientific understanding of the phenomenon.

More generally, there has been a growing recognition that much of science is a model-based activity. (For an overview of many different types of models in science, and some of the philosophical issues regarding the nature and use of such models, refer to [4.3]).

Models are by definition incomplete and idealized descriptions of the systems they describe. This practice raises all sorts of epistemological questions, such as how can it be that false models lead to true insights? And most relevant to our discussion here, how might the extensive use of models in science lead us to revise our philosophical account of scientific explanation?

4.1 The Explanatory Function of Models

Model-based explanations (or model explanations, for short) are explanations in which the explanans appeal to certain properties or behaviors observed in an idealized model or computer simulation as part of an explanation for why the (typically real-world) explanandum phenomenon exhibits the features that it does. For example, one might explain why sparrows of a certain species vary in their feather coloration from pale to dark by appealing to a particular game theory model: although coloration is unrelated to fitness, such a polymorphism can be a badge of status that allows the sparrows to avoid unnecessary conflicts over resources; dark birds are dominant and displace the pale birds from food sources. The model demonstrates that such a strategy is stable and successful, and hence can be used as part of the explanation for why we find this polymorphism among sparrows (see [4.4, 5] for further discussion).

There are, of course, many perils in assuming that just because we see a phenomenon or pattern exhibited in a model that it therefore explains why we see it in the real world: the same pattern or phenomenon could be produced in multiple, very different ways, and hence it might be only a phenomenological model at best, useful for prediction, but not a genuine explanation. Explanation and the concomitant notion of understanding are what we call success terms: if the purported explanation is not, in fact, right (right in some sense that will need to be spelled out) and the understanding is only illusory, then it is not, in fact, a genuine explanation. Determining what the success conditions are for a genuine explanation is the central philosophical problem in scientific explanation.

Those who have defended the explanatory power of models have typically argued that further conditions must be met in order for a model's exhibiting of a salient pattern or phenomenon to count as part of a genuine explanation of its real-world counterpart. Not all models are explanatory, and an adequate account of model explanation must provide grounds for making such discriminations. As we will see, however, different approaches have filled in these further requirements in different ways.

One of the earliest defenses of the view that models can explain is *McMullin*'s [4.6] *hypothetico-structural* HS account of model explanations. In an HS explanation, one explains a complex phenomenon by postulating an underlying structural model whose features are causally responsible for the phenomenon to be explained. McMullin notes that such models are often tentative or metaphorical, but that a good model explanation will lay out a research program for the further refinement of the model. On his account, the justification of the model as genuinely explanatory involves a process known as de-idealization, where features that were left out are added back or a more realistic representation of those processes is given. More specifically he requires that one be able to give a theoretical justification for this de-idealization process, so that it is not merely an ad hoc fitting of the model to the data. He writes [4.7, p. 261]:

> "If techniques for which no theoretical justification can be given have to be utilized to correct a formal idealization, this is taken to count against the explanatory propriety of that idealization. The model itself in such a case is suspect, no matter how good the predictive results it may produce."

He further notes that a theoretical justification for the de-idealization process will only succeed if the original model has successfully captured the real structure of the phenomenon of interest.

As an example, *McMullin* [4.8] describes the fertility of the continental drift model in explaining why the continents seem to fit together like pieces of a puzzle and why similar fossils are found at distant locations. The continental drift model involved all sorts of idealizations and gaps: most notably, the chief proponent of this approach, Alfred Wegener, could offer no account of the forces or mechanisms by which the massive continents could move. Strictly speaking, we now know that the continental drift model is false, and has been supplanted by plate tectonics. But as McMullin notes, the continental drift model nonetheless captures key features of the real structure of the phenomenon of interest, and, hence, succeeds in giving genuine explanatory insight.

While McMullin's account of HS model explanations fits in many cases, there are other examples of model explanations in the sciences that do not seem to fit his account. First, there seem to be examples of model explanations where the idealizations are ineliminable, and, hence, they cannot be justified through anything like the de-idealization analysis that McMullin describes [4.9]. Second, not all models are related to their target phenomena via an idealization: some models represent through a fictionalization [4.10]. Third, insofar as McMullin's HS model explanations are a subspecies of causal explanations, they do not account for noncausal model explanations. These sort of cases will be discussed more fully in subsequent sections.

Another early account of the explanatory power of models is *Cartwright*'s [4.11] *simulacrum* account of explanation, which she introduces as an alternative to the DN account of explanation and elaborates in her book *How the Laws of Physics Lie*. Drawing on *Duhem*'s [4.12] theory of explanation, she argues [4.11, p. 152]:

> "To explain a phenomenon is to find a model that fits it into the basic framework of the theory and that thus allows us to derive analogues for the messy and complicated phenomenological laws which are true of it."

According to Cartwright, the laws of physics do not describe our real messy world, only the idealized world we construct in our models. She gives the example of the harmonic oscillator model, which is used in quantum mechanics to describe a wide variety of systems. One describes a real-world helium-neon laser as if it were a van der Pol oscillator; this is how the phenomenon becomes tractable and we are able to make use of the mathematical framework of our theory. The laws of quantum mechanics are true in this model, but this model is just a simulacrum of the real-world phenomenon. By *model*, *Cartwright* means "an especially prepared, usually fictional description of the system under study" [4.11, p. 158]. She notes that while some of the properties ascribed to the objects in the models are idealizations, there are other properties that are pure fictions; hence, one should not think of models in terms of idealizations alone.

Although Cartwright's simulacrum account is highly suggestive, it leaves unanswered many key questions, such as when a model should or should not be counted as explanatory. *Elgin* and *Sober* [4.13] offer a possible emendation to Cartwright's account that they argue discriminates which sorts of idealized causal models can explain. The key, according to their approach, is to determine whether or not the idealizations in the model are what they call *harmless*. A harmless idealization is one that if corrected "wouldn't make much difference in the predicted value of the effect variable" [4.13, p. 448]. They illustrate this approach using the example of optimality models in evolutionary biology. Optimality models are models that determine what value of a trait maximizes fitness (is optimal) for an organism given certain constraints (e.g., the optimal length of a bear's fur, given the benefits of longer fur and the costs of growing it, or the optimal height at which crows should drop walnuts in order to crack open the shells, given the costs of flying higher, etc.). If organisms are indeed fitter the closer a trait is to the optimal value, and if natural selection is the only force operating, then the optimal value for that trait will evolve in the population. Thus, optimality models are used to explain why organisms have trait values at or near the optimal value (e.g., why crows drop walnuts from an average of 3 m high [4.14]).

As *Elgin* and *Sober* note, optimality models contain all sorts of idealizations: "they describe evolutionary trajectories of populations that are infinitely large in which reproduction is asexual with offspring always resembling their parents, etc." [4.13, p. 447]. Nonetheless, they argue that these models are genuinely explanatory when it can be shown that the value described in the explanandum is close to the value predicted by the idealized model; when this happens we can conclude that the idealizations in the model are harmless [4.13, p. 448]. Apart from this concession about *harmless* idealizations, Elgin and Sober's account of explanation remains close to the traditional DN account in that they further require:

1. The explanans must cite the cause of the explanandum
2. The explanans must cite a law
3. All of the explanans propositions must be true [4.13, p. 446]

though their condition 3 might better be stated as all the explanans propositions are *either* true *or harmlessly false*.

As a general account of model explanations, however, one might argue that the approaches of Cartwright, Elgin, and Sober are too restrictive. As noted before, this approach still depends on there being laws of nature from which the phenomenon is to be derived, and such laws just might not be available. Moreover, it is not clear that explanatory models will contain only harmless idealizations. There may very well be cases in which the idealizations make a difference (are not harmless) and yet are essential to the explanation (e.g., [4.15, 16]).

While the simulacrum approach of Cartwright, especially as further developed by Elgin and Sober, largely draws its inspiration from the traditional DN approach to explanation, there are other approaches to model explanation that are tied more closely to the traditional causal–mechanical approach to explanation. *Craver* [4.17], for example, has argued that models are explanatory when they describe mechanisms. He writes "[...] the distinction between explanatory and nonexplanatory models is that the [former], and not the [latter] describe mechanisms" [4.17, p. 367]. The central notion of mechanism, here, can be understood as consisting of the various components or parts of the phenomenon of interest, the activities of those components, and how they are organized in relation to each other.

Craver imposes rather strict conditions on when such mechanistic models can be counted as explanatory; he writes, "To characterize the phenomenon correctly and completely is the first restrictive step in turning a model into an acceptable mechanistic explanation" [4.17, p. 369]. (Some have argued that if one has a complete and accurate description of the system or phenomenon of interest, then it is not clear that one has a model [4.18]). Craver analyzes the example of the Hodgkin–Huxley mathematical model of the action potential in an axon (nerve fiber). Despite the fact that this model allowed Hodgkin and Huxley to derive many electrical features of neurons, and the fact that it was based on a number of fundamental laws of physics and chemistry, Craver argues that it was not in fact an explanatory model. He describes it instead as merely a phenomenological model because it failed to accurately describe the details of the underlying mechanism.

A similar mechanistic approach to model explanation has been developed by *Kaplan* [4.19], who introduces what he calls the mechanism–model–mapping (or 3M) constraint. He defines the 3M constraint as follows [4.19, p. 347]:

> "A model of a target phenomenon explains that phenomenon to the extent that (a) the variables in the model correspond to identifiable components, activities, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) variables in the model correspond to causal relations among the components of the target mechanism."

Kaplan takes this 3M constraint to provide a demarcation line between explanatory and nonexplanatory models. He further notes that [4.19, p. 347]

> "3M aligns with the highly plausible assumption that the more accurate and detailed the model is for a target system or phenomenon the better it explains that phenomenon."

Models that do not comply with 3M are rejected as nonexplanatory, being at best phenomenological models, useful for prediction, but giving no explanatory insight. In requiring that, explanatory models describe the *real* components and activities in the mechanism that are *in fact* responsible for producing the phenomenon ([4.17, p. 361], [4.19, p. 353]). Craver and Kaplan rule out the possibility that fictional, metaphorical, or strongly idealized models can be explanatory.

One of the most comprehensive defenses of the explanatory power of models is given by *Bokulich* [4.18, 20–22], who argues that model explanations such as the three discussed previously (McMullin, Cartwright–Elgin–Sober, and Craver–Kaplan), can be seen as special cases of a more general account of the explanatory power of models. Bokulich's approach draws on *Woodward*'s counterfactual account of explanation, in which [4.23, p. 11]

> "the explanation must enable us to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways."

She argues that model explanations typically share the following three features: first, the explanans makes essential reference to a scientific model, which, as is the case with all models, will be an idealized, abstracted, or fictionalized representation of the target system. Second, the model explains the explanandum by showing how the elements of the model correctly capture the patterns of counterfactual dependence in the target system, enabling one to answer a wide range of what Woodward calls *what-if-things-had-been-different* questions. Finally, there must be what *Bokulich* calls a *justificatory step*, specifying the domain of applicability of the model and showing where and to what extent the model can be trusted as an adequate representation of the target for the purpose(s) in question [4.18, p. 39]; see also [4.22, p. 730]. She notes that this justificatory step can proceed bottom-up through something like a de-idealization analysis (as McMullin, Elgin, and Sober describe), top-down through an overarching theory (such as in the semiclassical mechanics examples *Bokulich* [4.20, 21] discusses), or through some combination.

Arguably one of the advantages of Bokulich's approach is that it is not tied to one particular conception of scientific explanation, such as the DN or mechanistic accounts. By relaxing Woodward's manipulationist construal of the counterfactual condition, Bokulich's approach can even be extended to highly abstract, structural, or mathematical model explanations. She argues that the various *subspecies* of model explanation can be distinguished by noting what she calls the *origin* or ground of the counterfactual dependence. She explains, it could be either [4.18, p. 40]

> "the elements represented in the model *causally producing* the explanandum (in the case of causal

> model explanations), the elements of the model *being the mechanistic parts which make up* the explanandum-system whole (in the case of mechanistic model explanations), or the explanandum being a consequence of the laws cited in the model (in the case of covering law model explanations)."

She goes on to identify a fourth type of model explanation, which she calls structural model explanation, in which the counterfactual dependence is grounded in the typically mathematical structure of the theory, which limits the sorts of objects, properties, states, or behaviors that are admissible within the framework of that theory [4.18, p. 40]. Bokulich's approach can be thought of as one way to flesh out *Morrison*'s suggestive, but unelaborated, remark that "the reason models are explanatory is that in representing these systems, they exhibit certain kinds of structural dependencies" [4.24, p. 63].

More recently, *Rice* [4.25] has drawn on Bokulich's account to develop a similar approach to the explanatory power of models that likewise uses Woodward's counterfactual approach without the manipulation condition. He writes [4.25, p. 20]:

> "The requirement that these counterfactuals must enable one to, in principle, *intervene* in the system restricts Woodward's account to specifically causal explanations. However, I think it is a mistake to require that all scientific explanations must be causal. Indeed, if one looks at many of the explanations offered by scientific modelers, causes are not mentioned."

Compare this to *Bokulich*'s statement [4.18, p. 39]:

> "I think it is a mistake to construe all scientific explanation as a species of causal explanation, and more to the point here, it is certainly not the case that all model explanations should be understood as causal explanations. Thus while I shall adopt Woodward's account of explanation as the exhibiting of a pattern of counterfactual dependence, I will not construe this dependence narrowly in terms of the possible causal manipulations of the system"

Rice rightly notes that the question of causation is conceptually distinct from the question of what explains. He further requires on this approach that model explanations provide two kinds of counterfactual information, namely both what the phenomenon depends on and what sorts of changes are irrelevant to that phenomenon. Following *Batterman* [4.9, 15, 26], he notes that for explanations of phenomena that exhibit a kind of universality, an important part of the explanation is understanding that the particular causal details or processes are irrelevant – the same phenomenon would have been reproduced even if the causal details had been different in certain ways.

As an illustration, Rice discusses the case of optimality modeling in biology. He notes that optimality models are not only highly idealized, but also can be understood as a type of equilibrium explanation, where "most of the explanatory work in these models is done by *synchronic mathematical representations of structural features of the system*" [4.25, p. 8]. He connects this to the counterfactual account of model explanation as follows [4.25, p. 17]:

> "Optimality models primarily focus on noncausal counterfactual relations between structural features and the system's equilibrium point. Moreover, these features can sometimes explain the target phenomenon without requiring any additional causal claims about the relationships represented in the model."

These causal details are irrelevant because the structural features cited in the model are multiply realizable; indeed, this is what allows optimality models to be used in explaining a wide variety of features across a diversity of biological systems.

In the approaches to model explanations discussed here, two controversial issues have arisen that merit closer scrutiny: first, whether the fictions or falsehoods in models can themselves do real explanatory work (i. e., even when they are neither harmless, deidealizable, nor eliminable), and second, whether many model explanations illustrate an important, but often overlooked, noncausal form of explanation. These issues will be taken up in turn in the next two sections.

4.2 Explanatory Fictions: Can Falsehoods Explain?

Models contain all sorts of falsehoods, from omissions, abstractions, and idealizations to outright fictions. One of the most controversial issues in model explanations is whether these falsehoods, which are inherent in the modeling practice, are compatible with the explanatory aims of science. *Reiss* in the context of explanatory models in economics has called this tension the *explanation paradox*: he writes [4.27, p. 43]:

> "Three mutually inconsistent hypotheses concerning models and explanation are widely held: (1) economic models are false; (2) economic models are nevertheless explanatory; and (3) only true accounts explain. Commentators have typically resolved the paradox by rejecting either one of these hypotheses. I will argue that none of the proposed resolutions work and conclude that therefore the paradox is genuine and likely to stay."

(This paradox, and some criticisms to Reiss's approach (such as [4.28] are explored in a special issue of the *Journal of Economic Methodology* (volume 20, issue 3).)

The field has largely split into two camps on this issue: those who think it is only the true parts of models that do explanatory work and those who think the falsehoods play an essential role in the model explanation. Those in the former camp rely on things like de-idealization and harmless analyses to show that the falsehoods do not get in the way of the true parts of the model that do the real explanatory work. Those in the latter camp have the challenging task of showing that some idealizations are essential and some fictions yield true insights.

The *received view* is that the false parts of models only concern those things that are explanatorily irrelevant. Defenders of the received view include Strevens, who in his book detailing his kairetic account of scientific explanation (*Strevens* takes the term kairetic from the ancient Greek word *kairos*, meaning crucial moment [4.29, p. 477].)), writes, "No causal account of explanation – certainly not the kairetic account – allows nonveridical models to explain" [4.29, p. 297]. He spells out more carefully how such a view is to be reconciled with the widespread use of idealized models to explain phenomena in nature, by drawing the following distinction [4.29, p. 318]:

> "The content of an idealized model, then, can be divided into two parts. The first part contains the difference-makers for the explanatory target. [. . .] The second part is all idealization; its overt claims are false but its role is to point to parts of the actual world that do not make a difference to the explanatory target."

In other words, it is only the true parts of the model that do any explanatory work. The false parts are harmless, and hence should be able to be de-idealized away without affecting the explanation.

On the other side, a number of scholars have argued for the counterintuitive conclusion that sometimes it is in part *because* of their falsehoods – not despite them – that models explain. *Batterman* [4.9, 15, 26], for example, has argued that some idealizations are explanatorily ineliminable, that is, the idealizations or falsehoods themselves do real explanatory work. Batterman considers continuum model explanations of phenomena such shocks (e.g., compressions traveling through a gas in a tube) and breaking drops (e.g., the shape of water as it drips from a faucet). In order to explain such phenomena, scientists make the idealization that the gas or fluid is a continuum (rather than describing it veridically as a collection of discrete gas or water molecules). These false continuum assumptions are essential for obtaining the desired explanation. In the breaking drop case, it turns out that different fluids of different viscosities dripping from faucets of different widths will all exhibit the same shape upon breakup. The explanation depends on a singularity that exists only in the (false) continuum model; such an explanation does not exist on the de-idealized molecular dynamics approach [4.15, pp. 442–443]). Hence, he concludes [4.15, p. 427],

> "continuum idealizations are explanatorily ineliminable and [. . .] a full understanding of certain physical phenomena cannot be obtained through completely detailed, nonidealized representations."

If such analyses are right, then they show that not all idealizations can be de-idealized, and, moreover, those falsehoods can play an essential role in the explanation.

Bokulich [4.10, 20–22] has similarly defended the view that it is not just the true parts of models that can do explanatory work, arguing that in some cases even fictions can be explanatory. She writes, "some fictions can give us genuine insight into the way the world is, and hence be genuinely explanatory and yield real understanding" [4.10, p. 94]. She argues that some fictions are able to do this by capturing in their fictional representation real patterns of structural dependencies in the world. As an example, she discusses semiclassical models whereby fictional electron orbits are used to explain peculiar features of quantum spectra. Although, according to quantum mechanics, electrons do not follow definite trajectories or orbits (i. e., such orbits are fictions), physicists recognized that puzzling peaks in the recurrence spectrum of atoms in strong magnetic fields have a one-to-one correspondence with particular closed classical orbits [4.30, pp. 2789–2790] (quoted in [4.10, p. 99]):

> "The resonances [...] form a series of strikingly simple and regular organization, not previously anticipated or predicted. [...] The regular type resonances can be physically rationalized and explained by classical periodic orbits of the electron on closed trajectories starting at and returning to the proton as origin."

As she explains, at no point are these physicists challenging the status of quantum mechanics as the true, fundamental ontological theory; rather, they are deploying the fiction with the express recognition that it is indeed a literally false representation (interestingly this was one of the *Vaihinger*'s criteria for a *scientific* fiction [4.31, p. 98]). Nonetheless, it is a representation that is able to yield true physical insight and understanding by carefully capturing in its fictional representation the appropriate patterns of counterfactual dependence of the target phenomenon.

Bokulich [4.10, 20–22] offers several such examples of explanatory fictional models from semiclassical mechanics, where the received explanation of quantum phenomena appeals to classical structures, such as the Lyapunov (stability) exponents of classical trajectories, that have no clear quantum counterpart. Moreover, she notes that these semiclassical models with their fictional assumption of classical trajectories are valued not primarily as calculation tools (often they require calculations that are just as complicated), but rather are valued as models that provide an unparalleled level of physical insight into the structure of the quantum phenomena. Bokulich is careful to note that not just any fiction can do this kind of explanatory work; indeed, most fictions cannot. She shows more specifically how these semiclassical examples meet the three criteria of her account of model-based explanation, discussed earlier (e.g., [4.10, p. 106]).

A more pedestrian example of an explanatory fiction, and one that brings out some of the objections to such claims, is the case of light rays postulated by the ray (or geometrical) theory of optics. Strictly speaking, light rays are a fiction. The currently accepted fundamental theory of wave optics denies that they exist. Yet, light rays seem to play a central role in the scientific explanation of lots of phenomena, such as shadows and rainbows. The physicists *Kleppner* and *Delos*, for example, note [4.32, p. 610]:

> "When one sees the sharp shadows of buildings in a city, it seems difficult to insist that light-rays are merely calculational tools that provide approximations to the full solution of the wave equation."

Similarly, *Batterman*, argues [4.33, pp. 154–155]:

> "One cannot explain various features of the rainbow (in particular, the universal patterns of intensities and fringe spacings) without ultimately having to appeal to the structural stability of ray theoretic structures called caustics – focal properties of families of rays."

Batterman is quite explicit that he does not think that an explanatory appeal to these ray-theoretic structures requires reifying the rays; they are indeed fictions.

Some, such as Belot, want to dismiss ray-optics models as nothing but a mathematical device devoid of any physical content outside of the fundamental (wave) theory. He writes [4.34, p. 151]:

> "The mathematics of the less fundamental theory is definable in terms of that of the more fundamental theory; so the requisite mathematical results can be proved by someone whose repertoire of interpreted physical theories included only the latter."

The point is roughly this: it looks like in Batterman's examples that one is making an explanatory appeal to fictional entities from a *less fundamental* theory that has been superseded (e.g., ray optics or classical mechanics). However, all one needs from that superseded theory is the mathematics – one does not need to give those bits of mathematics a physical interpretation in terms of the fictional entities or structures. Moreover, that mathematics appears to be definable in terms of the mathematics of the true *fundamental* theory. Hence, those fictional entities are not, in fact, playing an explanatory role.

Batterman has responded to these objections, arguing that in order to have an explanation, one does, in fact, need the fictional physical interpretation of that mathematics, and hence the explanatory resources of the nonfundamental theory. He explains [4.33, p. 159]:

> "Without the physical interpretation to begin with, we would not know *what* boundary conditions to join to the differential equation. Neither, would we know *how* to join those boundary conditions to the equation. Put another way, we must examine the physical details of the *boundaries* (the shape, reflective and refractive details of the drops, etc.) in order to set up the *boundary conditions* required for the mathematical solution to the equation."

In other words, without appealing to the fictional rays, we would not have the relevant information we need to appropriately set up and solve the mathematical model that is needed for the explanation.

In a paper with *Jansson*, Belot has raised similar objections against Bokulich's arguments that classical structures can play a role in explaining quantum phenomena. They write [4.35, p. 82]:

> "Bokulich and others see explanations that draw on semiclassical considerations as involving elements of classical physics as well as of quantum physics. [...] But there is an alternative way of thinking of semiclassical mechanics: [...] starting with the formalism of quantum mechanics one proves theorems about approximate solutions – theorems that happen to involve some of the mathematical apparatus of classical mechanics. But this need not tempt us to think that there is [classical] physics in our explanations."

Once again, we see the objection that it is just the bare mathematics, not the mathematics with its physical interpretation that is involved in the explanation. On Bokulich's view, however, it is precisely by connecting that *mathematical apparatus* to its physical interpretation in terms of classical mechanics, that one gains a deeper physical insight into the system one is studying. On her view, explanation is importantly about advancing understanding, and for this the physical interpretation is important. (*Potochnik* [4.5, Chap. 5] has also argued for a tight connection between explanation and understanding, responding to some of the traditional objections against this association. More broadly, she emphasizes the communicative function of explanation over the ontological approach to explanation, which makes more room for nonveridical model explanations than the traditional approach.) Even though classical mechanics is not the true fundamental theory, there are important respects in which it gets things right, and hence reasoning with fictional classical structures within the well-established confines of semiclassical mechanics, can yield explanatory insight and deepen our understanding.

As we have seen, these claims that fictions can explain (in special cases such as ray optics and classical structures) remain controversial and involve subtle issues. These debates are not entirely new, however, and they have some interesting historical antecedents, for example, in the works of Niels Bohr and James Clerk Maxwell. More specifically, when Bohr is articulating his widely misunderstood *correspondence principle*, (for an accessible discussion see [4.36]) he argues that one can explain why only certain quantum transitions between stationary states in atoms are allowed by appealing to which harmonic components appear in the Fourier decomposition of the electron's classical orbit (see [4.20, Sect. 4.2] and references therein). He does this even long after he has conceded to the new quantum theory that classical electron trajectories in the atom are impossible (i. e., they are a fiction). Although *Heisenberg* used this formulation of the correspondence principle to construct his matrix mechanics, he argued that "it must be emphasized that this correspondence is a purely formal result" [4.37, p. 83], and should not be thought of as involving any physical content from the other theory. Bohr, by contrast, was dissatisfied with this interpretation of the correspondence principle as pure mathematics, arguing instead that it revealed a deep *physical* connection between classical and quantum mechanics. Even earlier, we can see some of these issues arising in the work of *Maxwell*, who, in exploiting the utility of fictional models and physical analogies between disparate fields, argued ([4.38, p. 187]; for a discussion, see [4.39]):

> "My aim has been to present the mathematical ideas to the mind in an embodied form [...] not as mere symbols, which convey neither the same ideas, nor readily adapt themselves to the phenomena to be explained."

Three other challenges have been raised against the explanatory power of fictional models. First, there is a kind of slippery-slope worry that, once we admit some fictional models as explanatory, we will not have any grounds on which to dismiss other fictional models as nonexplanatory. *Bokulich* [4.22] introduces a framework for addressing this problem. Second, *Schindler* [4.40] has raised what he sees as a tension in Bokulich's account. He claims that on one hand she says semiclassical explanations of quantum phenomena are autonomous in the sense that they provide more insight than the quantum mechanical ones. Yet, on the other hand, she notes that semiclassical models are justified through semiclassical theory, which connects these representations as a kind of approximation to the full quantum mechanics. Hence, they cannot be autonomous. This objection seems to trade on an equivocation of the term *autonomous*: in the first case, *autonomous* is used to mean "a representation of the phenomenon that yields more physical insight" and in the second case *autonomous* is used to mean "cannot be mathematical justified through various approximation methods". These seem to be two entirely different concepts, and, hence, not really in tension with each other. Moreover, Bokulich never uses the term *autonomous* to describe either, so this seems to be a misleading reading of her view.

Schindler also rehearses the objection, raised by *Belot* and *Jansson* [4.35], that by eliminating the interventionist condition in Woodward's counterfactual approach to explanation she loses what he calls "the asymmetry-individuating function", by which he means her account seems susceptible to the traditional problem of asymmetry that plagued the DN account of explanation (e.g., that falling barometers could be used to explain impending storms or shadows could used to explain the height of flag poles, to recall Sylvain Bromberger's well-known examples). This problem was taken to be solved by the causal approach to explanation, whereby one secures the explanatory asymmetry simply by appealing to the asymmetry of causation. It is important to note, however, that this is not an objection specifically to Bokulich's account of structural model explanation, but rather is a challenge for any noncausal account of explanation (*Bokulich* outlines a solution to the problem of asymmetry for her account in [4.22]). Since many examples of explanatory models purport to be noncausal explanations, we will examine this topic more fully in the next section.

Another context in which this issue about the explanatory power of fictional models arises is in connection with cognitive models in psychology and cognitive neuroscience. *Weiskopf*, for example, discusses how psychological capacities are often understood in terms of cognitive models that functionally abstract from the underlying real system. More specifically, he notes [4.41, p. 328]:

> "In attempting to understand the high level dynamics of complex systems like brains, modelers have recourse to many techniques for constructing such indirect accounts [...] *reification, functional abstraction*, and *fictionalization*."

By reification, he means "positing something with the characteristics of a more or less stable and enduring object, where in fact no such thing exists" [4.41, p. 328]. He gives as an example the positing of symbolic representations in classical computational systems, even though he notes that nothing in the brain seems to *stand still* or be manipulable in the way symbols do. Functional abstraction, he argues occurs when we [4.41, p. 329]

> "decompose a modeled system into subsystems and other components on the basis of what they do, rather than their correspondence with organizations and groupings in the target system."

He notes that this occurs when there are cross-cutting functional groupings that do not map onto the structural or anatomical divisions of the brain. He notes that this strategy emphasizes *networks, not locations* in relating cognition to neural structures. Finally, there is also fictionalization, which, as he describes [4.41, p. 331],

> "involves putting components into a model that are known not to correspond to any element of the modeled system, but which serve an essential role in getting the models to operate correctly."

He gives as an example of a fiction in cognitive modeling what are called *fast enabling links* (FELs), which are independent of the channels by which cells actually communicate and are assumed to have functionally infinite propagation speeds, allowing two cells to fire in synchrony [4.41, p. 331]. Despite being false in these ways, some modelers take these fictions to be essential to the operation of the model and not likely to be eliminated in future versions.

Weiskopf concludes that models involving reifications, functional abstractions, and fictions, can nonetheless in some cases succeed in "meeting the general normative constraints on explanatory models perfectly well" [4.41, p. 332], and hence such models can be counted as genuinely explanatory. Although Weiskopf recognizes the many great successes of mechanistic explanations in biological and neural systems, he wants to resist an *imperialism* that attempts to reduce all cases of model explanations in these fields to mechanistic model explanations.

More recently, *Buckner* [4.42] has criticized Weiskopf's arguments that functionalist models involving fictions, abstractions, and reification can be explanatory and defended the mechanist's maxim (e.g., as articulated by Craver and Kaplan) that only mechanistic models can genuinely explain. Buckner employs two strategies in arguing against Weiskopf: first, in cases where the models do explain, he argues that they are really just mechanism sketches, and where they cannot be reconstructed mechanistically, he dismisses them as impoverished explanations. He writes [4.42, p. 3]:

> "Concerning fictionalization and reification, I concede that models featuring such components cannot be interpreted as mechanism sketches, but argue that interpreting their nonlocalizable components as natural kinds comes with clear costs in terms of those models' counterfactual power. [...] Functional abstraction, on the other hand, can be considered a legitimate source of kinds, but only on the condition that the functionally abstract models be interpreted as sketches that could be elaborated into a more complete mechanistic model."

An essential feature of mechanistic models seems to be that their components are localizable. Weiskopf argues, however, that his functional kinds are multi-

Part A | 4.2

ply realizable, that is, they apply to many different kinds of underlying mechanisms, and that in some cases, they are distributed in the sense that they ascribe to a given model component capacities that are distributed amongst distinct parts of the physical system. Hence, without localization, such models cannot be reconstructed as mechanistic models.

What of Buckner's claim that fictional models will be impoverished with regard to their counterfactual power? Consider again Weiskopf's example of the fictional FELs, which are posited in the model to allow the cells to achieve synchrony. Buckner argues explanations involving models with FELs are impoverished in that if one had a true account of synchrony, that model explanation would support *more* counterfactual knowledge. It is not clear, however, that this objection undermines the explanatory power of models involving FELs per se; rather it seems only to suggest that if we knew more and had the true account of synchrony we might have a *deeper* explanation (at least on the assumption that this true account of synchrony would allow us to answer a wider range of what-if-things-had-been-different questions) (For an account of explanatory depth, see [4.43]). However, the explanation involving the fiction might still be perfectly adequate for the purpose for which it is being deployed, and hence it need not even be counted as impoverished. For example, there might be some explananda (ones other than the explanadum of *how do cells achieve synchrony*) for which it simply does not matter *how* cells achieve synchrony; the fact that they *do* achieve synchrony might be all that is required for some purposes.

Weiskopf is not alone in trying to make room for nonmechanistic model explanations; *Irvine* [4.44] and *Ross* [4.45] have also recently defended nonmechanistic model explanations in cognitive science and biology. Their approaches argue for noncausal forms of model explanation, which we will turn to next.

4.3 Explanatory Models and Noncausal Explanations

Recently, there has been a growing interest in noncausal forms of explanation. Similar to *Bokulich*'s [4.20, 21] approach, many of these seek to understand noncausal explanations within the context of *Woodward*'s [4.23] counterfactual approach to explanation without the interventionist criterion that restricts his account specifically to causal explanation [4.25, 46]. Noncausal explanations are usually defined negatively as explaining by some means *other than* citing causes, though this is presumably a heterogeneous group. We have already seen one type of noncausal model-based explanation: [4.20, 21] structural model explanations in physics. More recently, examples have been given in fields ranging from biology to cognitive science. Highly mathematical model explanations are another type of noncausal explanation, though not all mathematical models are noncausal. A few recent examples are considered here.

In the context of biology and cognitive science, *Irvine* [4.44] has argued for the need to go beyond the causal-mechanical account of model explanation and defends what she calls a noncausal structural form of model explanation. She focuses specifically on reinforcement learning (RL) models in cognitive science and optimality models in biology. She notes that although RL and optimality models can be construed as providing causal explanations in some contexts, there are other contexts in which causal explanations miss the mark. She writes [4.44, p. 11]:

> "In the account developed here, it is not the presence of idealisation or abstraction in models that is important, nor the lack of description of causal dynamics or use of robustness analyses to test the models. Instead, it is the bare fact that some models and target systems have equilibrium points [that] are highly O-robust with respect to initial conditions and perturbations. [...] This alone can drive a claim about noncausal structural explanations."

By O-robustness, Irvine means a robust convergence to an optimal state across a range of interventions, whether it be an optimization of fitness or an optimization of decision-making strategies. Her argument is that since interventions (in the sense of Woodward) do not make a difference to the convergence on the optimal state, that convergence cannot be explained causally, and is instead due to structural features of the model and target system it explains.

Another recent approach to noncausal model explanation is *Batterman* and *Rice*'s [4.47] minimal model explanations. Minimal models are models that explain patterns of macroscopic behavior for systems that are heterogeneous at smaller scales. Batterman and Rice discuss two examples of minimal models in depth: the Lattice Gas Automaton model, which is used to explain large-scale patterns in fluid flow, and Fisher's Sex Ratio model, which is used to explain why one typically finds a 1 : 1 ratio of males to females, across diverse populations of species. In both cases, they argue [4.47, p. 373]:

> "these minimal models are explanatory because there is a detailed story about why the myriad details that distinguish a class of systems are irrelevant

> to their large-scale behavior. This story demonstrates, rather than assumes, a kind of stability or robustness of the large-scale behavior we want to explain under drastic changes in the various details of the system."

They make two further claims about these minimal model explanations. First, they argue that these explanations are "distinct from various causal, mechanical, difference making, and so on, strategies prominent in the philosophical literature" [4.47, p. 349]. Second, they argue that the explanatory power of minimal models cannot be accounted for by any kind of mirroring or mapping between the model and target system (what they call the *common features* account). Instead, these noncausal explanations work by showing that the minimal model and diverse real-world systems fall into the same universality class. This latter claim has been challenged by *Lange* [4.48] who, though sympathetic to their claim that minimal models are a noncausal form of model explanation, argues that their explanatory power does in fact derive from the model sharing features in common with the diverse systems it describes (i. e., the *common features* account Batterman and Rice reject).

Ross [4.45] has applied the minimal models account to dynamical model explanations in the neurosciences. More specifically, she considers as an explanandum phenomenon the fact that a diverse set of neural systems (e.g., rat hippocampal neurons, crustacean motor neurons, and human cortical neurons *Ross* [4.45, p. 48]), which are quite different at the molecular level, nonetheless all exhibit the same *type I* excitability behavior. She shows that the explanation for this involves applying mathematical abstraction techniques to the various detailed models of each particular type of neural system and then showing that all these diverse systems converge on one and the same canonical model (known as the Ermentrout–Kopell model). After defending the explanatory power of these canonical models, *Ross* then contrasts this kind of noncausal model explanation with the causal–mechanical model approach [4.45, p. 46]:

> "The canonical model approach contrasts with Kaplan and Craver's claims because it is used to explain the shared behavior of neural systems without revealing their underlying causal–mechanical structure. As the neural systems that share this behavior consist of differing causal mechanisms [. . .] a mechanistic model that represented the causal structure of any single neural system would no longer represent the entire class of systems."

Her point is that a noncausal explanation is called for in this case because the particular causal details are irrelevant to the explanation of the universal behavior of class I neurons. The minimal models approach, as we saw above, is designed precisely to capture these sort explanations involving universality.

More generally, many highly abstract or highly mathematical model explanations also seem to fall into this general category of noncausal model explanations. *Pincock*, for example, identifies a type of explanation that he calls *abstract explanation*, which could be extended to model-based explanations. He writes "the best recent work on causal explanation is not able to naturally accommodate these abstract explanations" [4.49, p. 11]. Although some of the explanations Pincock cites, such as the topological (graph theory) explanation for why one cannot cross the seven bridges of Königsberg exactly once in a nonbacktracking circuit, seem to be genuinely noncausal explanations, it is not clear that all *abstract* explanations are necessarily noncausal. *Reutlinger* and *Andersen* [4.50] have recently raised this objection against Pincock's account, arguing that an explanation's being abstract is not a sufficient condition for it being noncausal. They argue that many causal explanations can be abstract too and so more work needs to be done identifying what makes an explanation truly noncausal. This is a particularly pressing issue in model-based explanations, since many scientific models are abstract in this sense of leaving out microphysical or concrete causal details about the explanandum phenomenon.

Lange [4.51] has also identified a kind of noncausal explanation that he calls a *distinctively mathematical* explanation. Lange considers a number of candidate mathematical explanations, such as why one cannot divide 23 strawberries evenly among three children, why cicadas have life-cycle periods that are prime, and why honeybees build their combs on a hexagonal grid. Lange notes that whether these are to count as distinctively mathematical explanations depends on precisely how one construes the explanandum phenomenon. If we ask why honeybees divide the honeycomb into hexagons, rather than other polygons, and we cite that it is selectively advantageous for them to minimize the wax used, together with the mathematical fact that a hexagonal grid has the least total perimeter, then it is an ordinary causal explanation (it works by citing selection pressures). If, however [4.50, p. 500]:

> "we narrow the explanandum to the fact that in any scheme to divide their combs into regions of equal area, honeybees would use at least the amount of wax they would use in dividing their combs into hexagons. [. . .] this fact has a distinctively mathematical explanation."

As *Lange* explains more generally [4.51, p. 485]:

> "These explanations are noncausal, but this does not mean that they fail to cite the explanandum's causes, that they abstract away from detailed causal histories, or that they cite no natural laws. Rather, in these explanations, the facts doing the explaining are modally stronger than ordinary causal laws. "

The key issue is not whether the explanans cite the explanandum's causes, but whether the explanation works *by virtue of* citing those causes. Distinctively mathematical (noncausal) explanations show the explanandum to be necessary to a stronger degree than would result from the causal powers alone.

As this literature makes clear, distinguishing causal from noncausal explanations is a subtle and open problem, but one crucial for understanding the wide-spread use of abstract mathematical models in many scientific explanations.

4.4 How-Possibly versus How-Actually Model Explanations

Models and computer simulations can often generate patterns or behaviors that are strikingly similar to the phenomenon to be explained. As we have seen, however, that is typically not enough to conclude that the model thereby explains the phenomenon. An important distinction here is that between a *how-possibly* model explanation and a *how-actually* model explanation.

The notion of a how-possibly explanation was first introduced in the 1950s by Dray in the context of explanations in history. *Dray* conceived of how-possibly explanations as a rival to the DN approach, which he labeled *why-necessarily* explanations [4.52, p. 161]. *Dray* interpreted how-possibly explanations as ones that merely aim to show why a particular phenomenon or event "need not have caused surprise" [4.52, p. 157]; hence, they are answers to a different kind of question and can be considered complete explanations in themselves. Although Dray's approach was influential, subsequent authors have interpreted this distinction in different ways. Brandon, in the context of explanations in evolutionary biology, for example, writes [4.53, p. 184]:

> "A how-possibly explanation is one where one or more of the explanatory conditions are speculatively postulated. But if we gather more and more evidence for the postulated conditions, we can move the how-possibly explanation along the continuum until finally we count it as a how-actually explanation. "

On this view, the distinction is a matter of the degree of confirmation, not a difference of kind: as we get more evidence that the processes cited in the model are the processes operating in nature, we move from a how-possibly to how-actually explanation.

Forber [4.54], however, rejects this interpretation of the distinction as marking a degree of empirical support, and instead defends Dray's original contention that they mark different kinds of explanations. More specifically, *Forber* distinguishes two kinds of how-possibly explanations that he labels *global how-possibly* and *local how possibly* explanations [4.54, p. 35]:

> "The global how-possibly explanations have theory, mathematics, simulations, and analytical techniques as the resources for fashioning such explanations. [...] The local how-possibly explanations draw upon the models of evolutionary processes and go one step further. They speculate about the biological possibilities relative to an information set enriched by the specific biology of a target system. [...] How-actually explanations, carefully confirmed by empirical tests, aim to identify the correct evolutionary processes that did, in fact, produce the target outcome."

Although Forber's distinction is conceptually helpful, it is not clear whether global versus local how-possibly explanations should, in fact, be seen as two distinct categories, rather than simply two poles of a spectrum.

Craver draws a distinction between how-possibly models and how-actually models that is supposed to track the corresponding two kinds of explanations. He notes that how-possibly models purport to explain (unlike phenomenological models, which do not purport to explain), but they are only loosely constrained conjectures about the mechanism. How-actually models, by contrast, describe the detailed components and activities that, in fact, produce the phenomenon. He writes [4.17, p. 361]:

> "How-possibly models are [...] not adequate explanations. In saying this I am saying not merely that the description must be true (or true enough) but further, that the model must correctly characterize the details of the mechanism."

Craver seems to see the distinction resting not just on the degree of confirmation (truth) but also on the degree of detail.

Bokulich [4.55] defends another construal of the how-possibly/how-actually distinction and applies it to model-based explanations more specifically. She considers, as an example, model-based explanations of a puzzling ecological phenomenon known as tiger bush. Tiger bush is a striking periodic banding of vegetation in semi-arid regions, such as southwest Niger. A surprising feature of tiger bush is that it can occur for a wide variety of plants and soils, and it is not induced by any local heterogeneities or variations in topography. By tracing how scientists use various idealized models (e.g., Turing models or differential flow models) to explain phenomena such as this, Bokulich argues a new insight into the how-possibly/how-actually distinction can be gained.

The first lesson she draws is that there are different levels of abstraction at which the explanandum phenomenon can be framed, which correspond to different explanatory contexts [4.55, p. 33]. These different explanatory contexts can be clarified by considering the relevant contrast class of explanations (for a discussion of contrast classes and their importance in scientific explanation, see [4.56, Chap. 5]). Second, she argues *pace* Craver that the how-possibly/how-actually distinction does not track how detailed the explanation is. She explains [4.55, p. 334]:

> "It is not the amount of detail that is relevant, but rather whether the mechanism represented in the model is the mechanism operating in nature. Indeed as we saw in the tiger bush case, the more abstractly the explanatory mechanism is specified, the easier it is to establish it as a how-actually explanation; whereas the more finely the explanatory mechanism is specified, the less confident scientists typically are that their particular detailed characterization of the mechanism is the actual one."

Hence, somewhat counterintuitively, model explanations at a more fine-grained level are more likely to be how-possibly model explanations, even when they are nested within a higher level how-actually model explanation of a more abstract characterization of the phenomenon. She concludes that when assessing model explanations, it is important to pay attention to what might be called the scale of resolution at which the explanandum phenomenon is being framed in a particular explanatory context.

4.5 Tradeoffs in Modeling: Explanation versus Other Functions for Models

Different scientists will often create different models of a given phenomenon, depending on their particular interests and aims. Following *Giere*, we might note that "there is no best scientific model of anything; there are only models more or less good for different purposes" [4.57, p. 1060]. If this is right, then it raises the following questions: What are the features that make a model particularly good for the purpose of explanation? Are there tradeoffs between different modeling aims, such that if one optimizes a model for explanation, for example, then that model will fail to be optimized for some other purpose, such as prediction?

One of the earliest papers to explore this theme of tradeoffs in modeling is Levins' paper *The Strategy of Model Building in Population Biology*. *Levins* writes [4.58, p. 422]:

> "It is of course desirable to work with manageable models which maximize generality, realism, and precision toward the overlapping but not identical goals of understanding, predicting, and modifying nature. But this cannot be done."

Levins then goes on to describe various modeling strategies that have evolved among modelers, such as sacrificing realism to generality and precision, or sacrificing precision to realism and generality. Levins in his own work on models in ecology favored this latter strategy, where he notes his concern was primarily qualitative not quantitative results, and he emphasizes the importance of robustness analyses in assessing these models.

Although Levins's arguments have not gone unchallenged, Matthewson and Weisberg have recently defended the view that some tradeoffs in modeling are genuine. They focus on precision and generality, given the relevance of this tradeoff to the aim of explanatory power. After a technical demonstration of different kinds of tradeoffs between two different notions of generality and precision, they conclude [4.59, p. 189]:

> "These accounts all suggest that increases in generality are, ceteris paribus, associated with an increase in explanatory power. The existence of tradeoffs between precision and generality indicates that one way to increase an explanatorily valuable desideratum is by sacrificing precision. Conversely, increasing precision may lead to a decrease in explanatory power via its effect on generality."

Mapping out various tensions and tradeoffs modelers may face in developing models for various aims, such as scientific explanation, remains a methodologically important, though underexplored topic.

More recently, *Bokulich* [4.60] has explored such tradeoffs in the context of modeling in geomorphology, which is the study of how landscapes and coastlines change over time. Even when it comes to a single phenomenon, such as braided rivers (i. e., rivers in which there is a number of interwoven channels and bars that dynamically shift over time), one finds that scientists use different kinds of models depending on whether their primary aim is explanation or prediction. When they are interested explaining why rivers braid geomorphologists tend to use what are known as *reduced complexity models*, which are typically very simple cellular automata models with a highly idealized representation of the fluvial dynamics [4.61]. The goal is to try to abstract away and isolate the key mechanisms responsible for the production of the braided pattern. This approach is contrasted with an alternative approach to modeling in geomorphology known as *reductionist* modeling. Here one tries to simulate the braided river in as much accurate detail and with as many different processes included as is computationally feasible, and then tries to solve the relevant Navier–Stokes equations in three dimensions. These reductionist models are the best available tools for predicting the features of braided rivers [4.61, p. 159], but they are so complex that they yield very little insight into *why* the patterns emerge as they do.

Bokulich uses cases such as these to argue for what she calls a division of cognitive labor among models [4.60, p. 121]:

> "If one's goal is explanation, then reduced complexity models will be more likely to yield explanatory insight than simulation models; whereas if one's goal is quantitative predictions for concrete systems, then simulation models are more likely to be successful. I shall refer to this as the *division of cognitive labor among models*."

As Bokulich notes, however, one consequence of this division of cognitive labor is that a model that was designed to optimize explanatory insight might fail to make quantitatively accurate predictions (a different cognitive goal). She continues [4.60, p. 121]:

> "This failure in predictive accuracy need not mean that the basic mechanism hypothesized in the explanatory model is incorrect. Nonetheless, explanatory models need to be tested to determine whether the explanatory mechanism represented in the model is in fact the real mechanism operating in nature."

She argues for the importance of robustness analyses in assessing these explanatory models, noting that while robustness analyses cannot themselves function as a nonempirical mode of confirmation, they can be used to identify those *qualitative* predictions or trends in the model that can appropriately be compared with observations.

4.6 Conclusion

There is a growing realization that the use of idealized models to explain phenomena is pervasive across the sciences. The appreciation of this fact has led philosophers of science to begin to introduce model-based accounts of explanation in order to bring the philosophical literature on scientific explanation into closer agreement with actual scientific practice.

A key question here has been whether the idealizations and falsehoods inherent in modeling are *harmless* in the sense of doing no real explanatory work, or whether they have an essential – maybe even ineliminable – role to play in some scientific explanations. Are such fictions compatible with the explanatory aims of science, and if so, under what circumstances? While some inroads have been made on this question, it remains an ongoing area of research. As we saw, yet another controversial issue concerns the fact that many highly abstract and mathematical models seem to exemplify a noncausal form of explanation, contrary to the current orthodoxy in scientific explanation. Determining what is or is not to count as a causal explanation turns out to be a subtle issue.

Finally, just because a model or computer simulation can reproduce a pattern or behavior that is strikingly like the phenomenon to be explained, does not mean that it thereby explains that phenomenon. An important distinction here is that between a how-possibly model explanation and a how-actually model explanation. Despite the wide agreement that such a distinction is important, there has been less agreement concerning how precisely these lines should be drawn.

Although significant progress has been made in recent years in understanding the role of models in scientific explanation, there remains much work to be done in further clarifying many of these issues. However, as the articles reviewed here reveal, exploring just how and when models can explain is a rich and fruitful area of philosophical investigation and one essential for understanding the nature of scientific practice.

References

4.1 C. Hempel: *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science* (Free Press, New York 1965)

4.2 W. Salmon: *Scientific Explanation and the Causal Structure of the World* (Princeton Univ. Press, Princeton 1984)

4.3 R. Frigg, S. Hartmann: Models in science. In: *The Stanford Encyclopedia of Philosophy*, ed. by E.N. Zalta (Stanford Univ., Stanford 2012)

4.4 J. Maynard Smith: *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge 1982)

4.5 A. Potochnik: *Idealization and the Aims of Science* (Univ. Chicago Press, forthcoming)

4.6 E. McMullin: Structural explanation, Am. Philos. Q. **15**(2), 139–147 (1978)

4.7 E. McMullin: Galilean idealization, Stud. Hist. Philos. Sci. **16**(3), 247–273 (1985)

4.8 E. McMullin: A case for scientific realism. In: *Scientific Realism*, ed. by J. Leplin (Univ. California Press, Berkeley 1984)

4.9 R. Batterman: Critical phenomena and breaking drops: Infinite idealizations in physics, Stud. Hist. Philos. Modern Phys. **36**, 25–244 (2005)

4.10 A. Bokulich: Explanatory Fictions. In: *Fictions in Science: Philosophical Essays on Modeling and Idealization*, ed. by M. Suárez (Routledge, London 2009) pp. 91–109

4.11 N. Cartwright: *How the Laws of Physics Lie* (Clarendon Press, Oxford 1983)

4.12 P. Duhem: *The Aim and Structure of Physical Theory* (Princeton Univ. Press, Princeton 1914/ 1954)

4.13 M. Elgin, E. Sober: Cartwright on explanation and idealization, Erkenntnis **57**, 441–450 (2002)

4.14 D. Cristol, P. Switzer: Avian prey-dropping behavior. II. American crows and walnuts, Behav. Ecol. **10**, 220–226 (1999)

4.15 R. Batterman: Idealization and modeling, Synthese **169**, 427–446 (2009)

4.16 A. Kennedy: A non representationalist view of model explanation, Stud. Hist. Philos. Sci. **43**(2), 326–332 (2012)

4.17 C. Craver: When mechanistic models explain, Synthese **153**, 355–376 (2006)

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

4.19 D.M. Kaplan: Explanation and description in computational neuroscience, Synthese **183**, 339–373 (2011)

4.20 A. Bokulich: *Reexamining the Quantum-Classical Relation: Beyond Reductionism and Pluralism* (Cambridge Univ. Press, Cambridge 2008)

4.21 A. Bokulich: Can classical structures explain quantum phenomena?, Br. J. Philos. Sci. **59**(2), 217–235 (2008)

4.22 A. Bokulich: Distinguishing explanatory from nonexplanatory fictions, Philos. Sci. **79**(5), 725–737 (2012)

4.23 J. Woodward: *Making Things Happen: A Theory of Causal Explanation* (Oxford University Press, Oxford 2003)

4.24 M. Morrison: Models as autonomous agents. In: *Models and Mediators: Perspectives on Natural and Social Science*, ed. by M. Morgan, M. Morrison (Cambridge Univ. Press, Cambridge 1999) pp. 38–65

4.25 C. Rice: Moving beyond causes: Optimality models and scientific explanation, Noûs **49**(3), 589–615 (2015)

4.26 R. Batterman: *Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence* (Oxford University Press, Oxford 2002)

4.27 J. Reiss: The explanation paradox, J. Econ. Methodol. **19**(1), 43–62 (2012)

4.28 U. Mäki: On a paradox of truth, or how not to obscure the issue of whether explanatory models can be true, J. Econ. Methodol. **20**(3), 268–279 (2013)

4.29 M. Strevens: *Depth: An Account of Scientific Explanation* (Harvard Univ. Press, Cambridge 2008)

4.30 J. Main, G. Weibusch, A. Holle, K.H. Welge: New quasi-Landau structure of highly excited atoms: The hydrogen atom, Phys. Rev. Lett. **57**, 2789–2792 (1986)

4.31 H. Vaihinger: *The Philosophy of 'As If': A System of the Theoretical, Practical, and Religious Fictions of Mankind*, 2nd edn. (Lund Humphries, London [1911] 1952), translated by C.K. Ogden

4.32 D. Kleppner, J.B. Delos: Beyond quantum mechanics: Insights from the work of Martin Gutzwiller, Found. Phys. **31**, 593–612 (2001)

4.33 R. Batterman: Response to Belot's "Whose Devil? Which Details?", Philos. Sci. **72**, 154–163 (2005)

4.34 G. Belot: Whose Devil? Which Details?, Philos. Sci. **52**, 128–153 (2005)

4.35 G. Belot, L. Jansson: Review of reexamining the quantum-classical relation, Stud. Hist. Philos. Modern Phys. **41**, 81–83 (2010)

4.36 A. Bokulich: Bohr's correspondence principle. In: *The Stanford Encyclopedia of Philosophy*, ed. by E.N. Zalta (Stanford Univ., Stanford 2014), http://plato.stanford.edu/archives/spr2014/entries/bohr-correspondence/

4.37 W. Heisenberg: In: *The Physical Principles of the Quantum Theory*, ed. by C. Eckart, F. Hoyt (Univ. Chicago Press, Chicago 1930)

4.38 J.C. Maxwell: On Faraday's Lines of Force. In: *The Scientific Papers of James Clerk Maxwell*, ed. by W. Niven (Dover Press, New York [1855/56] 1890) pp. 155–229

4.39 A. Bokulich: Maxwell, Helmholtz, and the unreasonable effectiveness of the method of physical analogy, Stud. Hist. Philos. Sci. **50**, 28–37 (2015)

4.40 S. Schindler: Explanatory fictions – For real?, Synthese **191**, 1741–1755 (2014)

4.41 D. Weiskopf: Models and mechanism in psychological explanation, Synthese **183**, 313–338 (2011)

4.42 C. Buckner: Functional kinds: A skeptical look, Synthese **192**, 3915–3942 (2015)

4.43 C. Hitchcock, J. Woodward: Explanatory generalizations: Part II. Plumbing explanatory depth, Noûs **37**(2), 181–199 (2003)

4.44 E. Irvine: Models, robustness, and non-causal explanation: A foray into cognitive science and biology, Synthese **192**, 3943–3959 (2015), doi:10.1007/s11229-014-0524-0

4.45 L. Ross: Dynamical models and explanation in neuroscience, Philos. Sci. **82**(1), 32–54 (2015)

4.46 J. Saatsi, M. Pexton: Reassessing Woodward's account of explanation: Regularities, counterfactuals, and noncausal explanations, Philos. Sci. **80**(5), 613–624 (2013)

4.47 R. Batterman, C. Rice: Minimal model explanations, Philos. Sci. **81**(3), 349–376 (2014)

4.48 M. Lange: On 'Minimal model explanations': A reply to Batterman and Rice, Philos. Sci. **82**(2), 292–305 (2015)

4.49 C. Pincock: Abstract explanations in science, Br. J. Philos. Sci. **66**(4), 857–882 (2015), doi:10.1093/bjps/axu016

4.50 A. Reutlinger, H. Andersen: Are explanations noncausal by virtue of being abstract?, unpublished manuscript

4.51 M. Lange: What makes a scientific explanation distinctively mathematical?, Br. J. Philos. Sci. **64**, 485–511 (2013)

4.52 W. Dray: *Law and Explanation in History* (Oxford Univ. Press, Oxford 1957)

4.53 R. Brandon: *Adaptation and Environment* (Princeton Univ. Press, Princeton 1990)

4.54 P. Forber: Confirmation and explaining how possible, Stud. Hist. Philos. Biol. Biomed. Sci. **41**, 32–40 (2010)

4.55 A. Bokulich: How the tiger bush got its stripes: 'How possibly' vs. 'How actually' model explanations, The Monist **97**(3), 321–338 (2014)

4.56 B. van Fraassen: *The Scientific Image* (Oxford University Press, Oxford 1980)

4.57 R. Giere: The nature and function of models, Behav. Brain Sci. **24**(6), 1060 (2001)

4.58 R. Levins: The Strategy of model building in population biology, Am. Sci. **54**(4), 421–431 (1966)

4.59 J. Matthewson, M. Weisberg: The structure of tradeoffs in model building, Synthese **170**(1), 169–190 (2008)

4.60 A. Bokulich: Explanatory models versus predictive models: Reduced complexity modeling in geomorphology. In: *EPSA11 Perspectives and Foundational Problems in Philosophy of Science*, ed. by V. Karakostas, D. Dieks (Springer, Cham, Heidelberg, New York, Dordrecht, London 2013)

4.61 A.B. Murray: Contrasting the goals, strategies, and predictions associated with simplified numerical models and detailed simulations. In: *Prediction in Geomorphology*, ed. by P. Wilcock, R. Iverson (American Geophysical Union, Washington 2003) pp. 151–165

5. Models and Simulations

Nancy J. Nersessian, Miles MacLeod

In this chapter we present some of the central philosophical issues emerging from the increasingly expansive and sophisticated roles computational modeling is playing in the natural and social sciences. Many of these issues concern the adequacy of more traditional philosophical descriptions of scientific practice and accounts of justification for handling computational science, particularly the role of theory in the generation and justification of physical models. However, certain novel issues are also becoming increasingly prominent as a result of the spread of computational approaches, such as nontheory-driven simulations, computational methods of inference, and the important, but often ignored, role of cognitive processes in computational model building.

Most of the philosophical literature on *models and simulations* focuses on computational simulation, and this is the focus of our review. However, we wish to note that the chief distinguishing characteristic between a model and a simulation (model) is that the latter is dynamic. They can be *run* either as constructed or under a range of experimental conditions. Thus, the broad class of simulation models should be understood as comprising dynamic physical models and mental models, topics considered elsewhere in this volume.

This chapter is organized as follows. First in Sect. 5.1 we discuss simulation in the context of well-developed theory (usually physics-based simulations). Then in Sect. 5.2 we discuss simulation in contexts where there are no over-arching theories of the phenomena, notably agent-based simulations and those in systems biology. We then turn to issues of whether and how simulation modeling introduces novel concerns for the philosophy of science in Sect. 5.3. Finally, we conclude in Sect. 5.4 by addressing the question of the relation between human cognition and computational simulation, including the relationship between the latter and thought experimenting.

5.1 Theory-Based Simulation

A salient aspect of computational simulation, and the one which has attracted the most substantial philosophical interest so far, is its ability to extend the power and reach of theories in modern science beyond what could be achieved by pencil and paper alone. Work on simulations has concentrated on simulations built from established background theories or theoretical models and the relations between these simulations and theory. Examples have been sourced mainly from the physical sciences, including simulations in astrophysics, fluid dynamics, nanophysics, climate science and meteorology. *Winsberg* has been foremost in studying theory-driven forms of simulation and promoting the importance of philosophical investigation of it by arguing that such simulations set a new agenda for philosophy of science [5.1–5]. He uses the case of simulation to challenge the longstanding focus of philosophy of science on theories, particularly on how they are justified [5.1, 3, 5]. Simulations, he argues, cannot simply be understood as novel ways to test theories. They are in fact rarely used to help justify theories, rather simulations apply existing theories in order to explore, explain and understand real and possible phenomena, or make predictions about how such phenomena will evolve in time. Simulations open up a whole new set of philosophical issues concerning the practices and reliability of much modern science.

Winsberg's analysis of theory-based simulation shares much with *Cartwright*'s [5.6] and *Morgan* and *Morrison*'s [5.7] challenges to the role of theories. Like them, he starts by strongly disputing the presupposition that simulations are somehow deductive derivations from theory. Simulations are applied principally in the physical sciences when the equations generated from a theory to represent a particular phenomenon are not analytically solvable. The path from a theory to a simulation requires processes of computerization, which transform equations into tractable computable structures by relying on practices of discretization and idealization [5.8]. These practices employ specific transformations and simplifications in combination with those used to make tractable the application of theoretical equations to a specific phenomenon such as boundary conditions and symmetry assumptions. As such simulations are, according to *Winsberg* [5.1], better construed as particular articulations of a theory rather than derivations from theory. They make use of theoretical information and the credibility, explanatory scope and depth, of well-established theories, to provide warrant to simulations of particular phenomena. Inferences drawn by computational simulations have several features in this regard; they are *downward*, *motley* and *autonomous* [5.9]. Inferences are downward because they move from theory to the real world (rather than from the real world to theory). They are motley because they depend not just on theory but on a large range of extra-theoretical techniques and resources in order to derive inferences, such as approximation and simplification techniques, numerical methods, algorithmic methods, computer languages and hardware, and much trial and error. Finally, simulations are autonomous, in the sense of being autonomous from both theory and data. Simulations, according to Winsberg, are principally used to study phenomena where *data is sparse* and unavailable. These three conditions on inference from simulation require a specific philosophical evaluation of their reliability.

Such evaluation is complicated by the fact that relations between theory and inferences drawn from the simulation model are unclear and difficult to untangle. As *Winsberg* [5.1, 9] suggests it is a complex task to unpack what role theories play in the final result given all these intervening steps. The fact that much validation of simulations is done through matching simulation outputs to the data, muddies the water further (see also [5.10]). A well-matched simulation constructed through a downward, motley and autonomous process from a nonetheless well-established theory raises the question of the extent to which the confirmation afforded to the theory flows down to the simulation [5.2]. For instance, although fitting a certain data set might well be the dominant mode of validation of a simulation model, the model could be considered to hold outside the range of that data because the model applies a well-accepted theory of the phenomenon thought to hold under very general conditions.

There is widespread agreement that untangling the relations between theories and simulations, and the reliability of simulations built from theories will require more in depth investigation of the actual practices scientists use to justify the steps they make when building a simulation model. In the absence of such investigations discussions of justification are limited to considerations about whether a simulation fits the observational data or not. Among other things, this limitation hides from view important issues about the warrant of the various background steps that transform theoretical information into simulations [5.10]. In general, what is required is an *epistemology of simulation* which can discover rigorous grounds upon which scientists can and do sanction their results, and more properly the role of theory in modern science.

The concern with practices of simulation has opened up a new angle on the older discussion about the structure of theories. *Humphreys* [5.11] has used the entanglement of theory and simulation in modern scientific practice to reflect more explicitly upon the proper philosophical characterization of the structure of physical theories. Simulations, as with other models, are not logical derivations from theory which is a central, but incorrect, feature of the syntactic view. Humphreys also argues, however, that the now dominant semantic view of theories, which treats theories as nonlinguistic entities, is not adequate either. On the semantic view a syntactical formulation of a theory, and whether different formulations might be solvable or not, is not important for philosophical assessment of relations of representations to the world. Relations of representation are only in fact sensibly held by models not theories. Both Humphreys and Winsberg construe the semantic view as dismissing the role of theories in both normative and descriptive accounts of science, in place of models. But as *Humphreys* [5.12, p. 620] puts it, "the specific syntactic representation used is often crucial to the solvability of a theory's equations", and thus, the solvability of models derived from it. Computational tractability, as well as choices of approximation and simplification techniques, will depend on the particular syntax of a theory. Hence both the semantic and syntactic views are inadequate for describing theory in ways that capture their role in science.

5.2 Simulation not Driven by Theory

Investigations, such as those by Winsberg and others discussed in the previous section, have illustrated the importance of close attention to scientific practice and discovery when studying simulations. Simulation manifests application-intensive, rather than theoretical, processes of scientific investigation. As *Winsberg* [5.1] suggests choices about how to model a phenomenon reliably are developed often in the course of the to and fro *blood, sweat and tears* of the model-building process itself. Abstract armchair points of view, distant from an understanding of the contingent, but also technical and technological nature of these practices and their affordances, will not put philosophers in a position to create relevant normative assessments of good simulation practices. What has thus far been established by the accounts of theory-based simulation is that even in the case where there is an established theory of the phenomena, simulation model-building has a degree of independence from theory and theory-building.

However, though the initial focus on theory-based simulation in the study of simulation is not unsurprising given the historical preference in philosophy of science for treating *theory* as the principal unit of philosophical investigation, simulations are not just a tool of theory-driven science alone. Pushing philosophical investigation into model-building practices outside the domain of theory-driven science reveals whole new practices of scientific model production using computational simulations that are not in fact theory-based, in the sense of traditional physical sciences. Some of the most compelling and innovative fields in science today, including, for instance, big-data biology, systems biology and neuroscience, and much modeling in the social sciences, are not theory-driven. As *Winsberg* [5.5] admits (in response to *Parker* [5.13]), his description of simulation modeling is theory-centric, and neither necessarily applicable to understanding the processes by which simulation models are built in the absence of theory, nor an appropriate framework for assessing the reliability and informativeness of models built that way. This is not to say that characteristics of theory-based simulation are irrelevant to simulations that are not. Both theory and nontheory-based simulations share an independence of theory and there are likely to be similarities between them, but there are also profound differences.

One kind of simulation that is important in this regard is agent-based modeling. *Keller* [5.14] has labeled much agent-based modeling as *modeling from above* in the sense that such models are not constructed using a mathematical theory that governs the motions of agents. Agents follow local interactions rules. In many fields in the social sciences and biology differential equations cannot be used to aggregate accurately agent or population behavior, but it is nonetheless possible to hypothesize or observe the structure of individual interactions. An agent-based model can be used to run those interactions over a large population to test whether the local structures can reproduce aggregate behavior [5.15]. As noted by *Grüne-Yanoff* and *Weirich* [5.16] agent-based modeling facilitates constructing remarkably complex models within computationally tractable constraints that often go well beyond what is possible with equation-based representations.

Agent-based models provide one exemplar of simulations that are not theory-driven. From an epistemological perspective, these simulations exhibit weak emergence [5.17]. The underlying mechanisms are thoroughly opaque to the users, and the way in which emergent properties come about can simply not be reassembled by studying the simulation processes. This opacity raises questions about the purpose and value of agent-based modeling. What kind of explanation and understanding does an agent-based simulation provide if the multiscale mechanisms produced in a simulation are cognitively inaccessible? Further, how is one to evaluate predictions and explanations from agent-based simulations which, in fields like ecology and economics, commonly simplify very complex interactions in order to create computationally tractable simulations. If a simplistic model captures a known behavior, can we trust its predictions? To address questions such as these we need an epistemology that can evaluate proposed techniques for establishing the robustness of agent-based models. One alternative is to argue that agent-based models require a novel epistemology that is able to rationalize their function as types of fictions rather than as representations [5.18, 19]. Another alternative, presented by *Grüne-Yanoff* and *Weirich* [5.16], is to argue that agent-based models provide in many cases functional rather than causal explanations of the phenomena they simulate [5.20]. Agent-based model simulations rarely control for all the potential explanatory factors that might be relevant to a given phenomenon, and any choice of particular interaction mechanism is usually thoroughly underdetermined. In practice, all possible mechanisms cannot be explored. But agent-based models can show reliably how particular lower-level capacities behave in certain ways, when modeled by suitably general interactions rules, and can constitute higher-level capacities no matter how multiply realized those interactions might be. Hence, such models, even though greatly simplified, can extract useful information despite a large space of potential explananda.

Nontheory-driven forms of simulation such as agent-based models provide a basis for reflecting more broadly on the role theory plays in the production of simulations, and the warrant a theory brings to simulations based on it. Comparative studies of the kinds of arguments used to justify relying on a simulation should expose the roles well-established theories play. Our investigations of integrative systems biology (ISB) have revealed that not all equation-based modeling is theory-driven, if theory is construed in terms of theory in the physical sciences. The canonical meaning based on the physical sciences is something like a background body of laws and principles of a domain.

In the case of systems biology, researchers generally do not have access to such theory and in fact the kinds of theory they do make use of have a function different from what is usually meant by *theory* in fields like physics [5.21]. There are certain canonical theories in systems biology of how to mathematically represent interactions among, for instance, metabolites, in the form of sets of ordinary differential equations. These posit particular canonical mathematical forms for representing a large variety of interactions (see Biochemical Systems Theory [5.22]). In principle, for any particular metabolic network, if all the interactions and reactants are known, the only work for the modeler is to write down the equations for a particular network and calculate the parameters. The mathematics will take care of the rest since the mathematical formulations of interactions are general enough that any potential nonlinear behaviors should be represented if parameters are correctly fixed.

For the most part, however, these canonical frameworks do not provide the basic ontological information from which a representation of a system is ultimately drawn, in the way say that the Navier-Stokes equations of fluid dynamics describe fluids and their component interactions in a particular way. In practice, modelers in systems biology need to assemble that information themselves in the form of pathway diagrams which more or less list the molecules involved and then make their own decisions about how to represent molecular interactions. A canonical framework is better interpreted as a theory of how to approximate and simplify the information that the systems biologist has assembled about a pathway in order to reliably simulate the dominant dynamics of a network given sparse data and complex nonlinear dynamics. Hence, there is no real *theory articulation* in Winsberg's terms. Researchers do not articulate a general theory for a particular application. The challenge for systems biologists is to build a higher level or system level representation out of the lower level information they possess. We have found that canonical templates mediate this process by providing a possible structure for gluing together this lower level information in a tractable way [5.21]. These theories do not offer any direct explanatory value by virtue of their use.

Theory can in fact be used not just to describe a body of laws and theoretical principles, but also to describe principles that instruct scientists on how to reliably build models of given classes of phenomena from a background theory. As Peck puts it [5.18, p. 393]:

> "In traditional mathematical modeling, there is a long established research program in which standard methods, such as those used for differential equation modeling, are used to bring about certain ends. Once the variables and parameters and their relationships are chosen for the representation of the model, standard formulations are used to complete the modeling venture."

If one talks about what physical scientists often start with it is not just the raw theory itself but well-established rules for formulating the theory and applying it with respect to a particular phenomenon. We might refer to this latter sense of *theory* as a theory of how to apply a background theory to reliably represent a phenomenon. The two senses of theory are exclusive. In the case of the canonical frameworks, what is meant by *theory* is something closer to this latter rather than former sense.

Additionally, the modelers we have studied are never in a position to rely on these frameworks uncritically and in fact no theory exists that specifies which representations to use that will reliably lead to a good representation in all data situations. In integrative systems biology the variety of data situations are very complex, and the data are often sparse and are rarely adequate for applying a set mathematical framework. This forces researchers in practice into much more intensive and adaptive model-building processes that certainly share much in common with the back and forth processes Winsberg talks about in the context of theory application. But these processes have the added and serious difficulty that the starting points for even composing the mathematical framework out of which a model should be built are open-ended and need to be decided based on thorough investigation of the possibilities with the specific data available.

Canonical frameworks are just an option for modelers and do not drive the model-building process in the way physical theories do. Currently, systems biology generally lacks effective theory of either kind. Modelers have many different choices about how to confront a particular problem that do not necessarily involve picking up a canonical framework or sticking to it. MacLeod and Nersessian [5.21] have documented how the nontheory-derived model-building processes work in these contexts. Models are strategic adaptations to a complex set of constraints system biologists are working under [5.23]. Among these constraints are:

- Constraints of the biological problem: A model must address the constraints of the biological problem, such as how the redox environment is maintained in a healthy cell. The system involved is often of considerable complexity.
- Informational/data constraints: There are constraints on the accessibility and availability of experimental data and molecular and system parameters for constructing models.
- Cost constraints: ISB is data-intensive and relies on data that often go beyond what are collected by molecular biologists in small scale experiments. However, data are very costly to obtain.
- Collaboration constraints: Constraints on the ability to communicate effectively with experimental collaborators with different backgrounds or in different fields in order to obtain expert advice or new data. Molecular biologists largely do not understand the nature of simulation modeling, do not understand the data needs of modeling, and do not see the cost-benefit of producing the particular data systems biologists ask from them.
- Time-scale constraints: Different time scales operate with respect to generating molecular experimental data versus computational model testing and construction.
- Infrastructure constraints: There is little in the way of standardized databases of experimental information or standardized modeling software available for systems biologists to rely upon.
- Knowledge constraints: Modelers' lack knowledge of biological systems and experimental methods limits their understanding of what is biologically plausible and what reliable extrapolations can be made from the data sets available.
- Cognitive constraints: Constraints on the ability to process and manipulate models because of their complexity, and thus constraints on the ability to comprehend biological systems through modeling.

Working with these constraints requires them to be *adaptive problem-solvers*. Given the complexity of the systems, lack of data, and the ever-present problem of computational tractability, researchers have to experiment with different mathematical formulations, different parameter-fixing algorithms and approximation techniques in highly intensive trial and error processes. They build models in nest-like fashion in which bits of biological information and data and mathematical and computational techniques, get combined to create stable models. These processes transform not only the shape of the solutions, but also the problems, as researchers figure out what actual problem can be solved with the data at hand. Simulation plays a central exploratory role in the process. This point goes further than Lenhard's idea of an explorative cooperation between experimental simulation and models [5.8]. Simulation in systems biology is not just for experimenting on systems in order to *sound out the consequences of a model* [5.8, p. 181], but plays a fundamental role in incrementally building the model and learning the relevant known and sometimes unknown features of a system and gaining an understanding of its dynamics. Simulation's roles as a cognitive resource make the construction of representations of complex systems without a theoretical basis possible (see also [5.24, 25]).

Similar conclusions have been drawn by Peck for ecology which shares with systems biology the complexity in its problems and a lack of generalizable theory. As Peck [5.18, p. 393] points out:

> "there are no formal methodological procedures for building these types of models suggesting that constructing an ecological simulation can legitimately be described as an art."

This situation promotes methodological pluralism and creative methodological exploration by modelers. Modelers in these contexts thus focus our attention on the deeper roles (sometimes called heuristic roles [5.5]) that simulation plays in the ability of researchers to explore potential solutions in order to solve complex problems.

These roles have added epistemological importance when it is realized that the *downward* character of simulation can be fact reversed in both senses we have mentioned above. This is a potentially significant difference between cases of theory and nontheory-driven simulation. Consider again systems biology. Firstly, the methodological exploration we witness amongst the researchers we have studied can be rationalized as precisely an attempt by the field to establish a good theory of how to build models of biological systems that work well given a variety of data situations. Since the complexities of these systems and computational constraints make this difficult to know at the outset, the field needs its freedom to explore the possibilities. Lab directors do encourage exploration, and part of the reason they do is to try to glean which practices work well and which do not given a lack of knowledge of what will work well for a given problem.

Secondly, systems biology aspires to a theory of biological systems which will detail general system-level characteristics of biological systems but also the design principles underlying biological networks [5.26, 27]. What is interesting about this theory, if it does emerge, is that it will in fact be theory generated *by* simulation rather than the other way around. Simulation makes possible the exploration of quite complex systems for generalities that can form the basis of a theory of systems biology. As such the use of simulations can also be upwards, not just downwards, to perhaps an unprecedented extent. Upward uses of simulation requires analysis that appears to fit better with more traditional philosophical analysis of how theories are in fact justified, only in this case robust simulation models will possibly be the more significant source of evidence rather than traditional experiment and observation. How this affects the nature and reliability of our inferences to theory, and what kind of resemblance such theory might have to theory in physics, is something that will need investigation. Thus, further exploration of nontheory-driven modeling practices stand to provide a rich ground for investigation of novel practices that are emerging with simulation, but also for exploring the roles and meanings of *theory*.

5.3 What is Philosophically Novel About Simulation?

The question of whether or not simulation introduces new issues into the philosophy of science has emerged as a substantial debate in discussions of computational simulation. *Winsberg* [5.1, 3–5] and *Humphreys* [5.11, 12] are the major proponents of the view that simulation requires its own epistemology. Winsberg, for instance, takes the view that simulations exhibit "distinct epistemological characteristics ... novel to the philosophy of science" [5.9, p. 443]. Winsberg and Humphreys make this assertion on the basis of the points we outlined in Sec. 5.2; namely, 1) the traditional limited concern of philosophy of science with the justification of theory, and 2) the relative autonomy of simulations and simulation-building from the theory. The steps involved in generating simulations, such as applying approximation methods designed to generate computational tractability, are novel to science. These steps do not gain their legitimacy from a theory but are "autonomously sanctioned" [5.1, p. 837]. Winsberg argues, for instance, that while idealization and approximation methods have been discussed in the literature it has mostly been from a representational perspective in terms of how idealized and approximate models represent or resemble the world and in turn justify the theories on which they are based. But since simulations are often employed where data are sparse, they cannot usually be justified by being compared with the world alone. Simulations must be assessed according to the reliability of the processes used to construct them, and these often distinct and novel techniques require separate philosophical evaluation. Mainstream philosophy of science with its focus on theoretical justification does not have the conceptual resources for accounting for applications using computational methods. Even where theory is concerned, both Humphreys and Winsberg maintain that neither of the established semantic and syntactic conception of theories, conceptions which focus on justification and representation, can account for how theories are applied or justified in simulation modeling.

However, *Frigg* and *Reiss* [5.28] have countered that these claims were overblown and in fact simulation raises no new questions or problems that are specific to simulation alone. Part of the disagreement might simply come down to whether one construes philosophy of science narrowly or broadly by limiting *philosophical questions* to in-principle and normative issues, while avoiding practical methodological ones. Another part of the disagreement is over how one construes *new issues* or *new questions* for philosophy, since certainly at some level the basic philosophical questions about how representations represent and what makes them reliably do so, are still the same questions.

To some extent, part of the debate might be construed as a disagreement over the relevance of contexts of discovery to philosophy of science. Classically contexts of discovery, the scientific contexts in which model-building takes place, are considered irrelevant to normative philosophical assessments of whether those models are justified or not. *Winsberg* [5.3] and *Humphreys* [5.12] seem willing to assert that one of the lessons for philosophy of science from simulation is that practical constraints on scientific discovery matter for constructing relevant normative principles – both in terms of evaluating current practice, which in the case of simulation-building is driven by all kinds of practical constraints, and in terms of normatively directing practice sensitively within those constraints.

Part of the motivation for using the discovery/justification distinction to define philosophical interest and relevance is the belief that there is a clear distinction between the two contexts. Arguably Frigg and Reiss are reinforcing the idea of a clear distinction by relying on widespread presupposition that validation and verification are distinct independent processes [5.4]. Validation is the process of establishing that a simulation is a good representation, a quintessential concept of justification. Verification is the process of ensuring that a computational simulation adequately captures the equations from which it is constructed. Verification, according to Frigg and Reiss, represents the only novel aspects of modeling that simulation introduces. Yet it is a purely mathematical exercise that is of no relevance to questions of validation. As such, simulations involve no new issues of justification beyond those of ordinary models. *Winsberg* [5.3, 4], however, counters that there is, in practice, no clear division between processes of verification and validation. The equations chosen to represent a system are not simply selected on the basis of how valid they are, but also on the basis of decisions about computational tractability. Much of what validates a representation in practice occurs at the end stage, after all the necessary techniques of numerical approximation and discretization have been applied, by comparing the results of simulations with the data. As such, [5.5]:

> "If we want to understand why simulation results are taken to be credible, we have to look at the epistemology of simulation as an integrated whole, not as clearly divided into verification and validation – each of which would look inadequate to the task."

Hence what would otherwise seem to be distinct discovery and justification processes are in the context computational simulation interwoven.

Frigg and Reiss are right at some level that simulations do not change basic epistemological questions connected to the justification of models. They are also right that Winsberg in his downward, motley and autonomous description of simulation, does not reveal any fundamentally new observations on model-building that have not already been identified as issues by philosophers discussing traditional modeling. However, what appears to be really new in the case of simulation is: 1) the complexity of the philosophical problems of representation and reliability, and 2) the different methodological and epistemological strategies that have become available to modelers as a result of simulation.

Winsberg, in reply to Frigg and Reiss, has clarified what he thinks as novel about theory-based simulation as the *simultaneous confluence* of downward, motley and autonomous features of model-building [5.4]. It is the reliability and validity of the complex modeling processes instantiated by these three features that must be accounted for by an epistemology of simulation, and no current philosophical approaches are adequate to do so, particularly not those within traditional philosophical boundaries of analysis.

As a first step in helping with this task of assessing reliability and validity of simulation, philosophers such as *Winsberg* [5.29] have drawn lessons from comparison with experimentation, which they argue shares much with simulation in both function (enabling, for instance, in silico experiments) and also in terms of how the reliability of simulations is generated. Scientific researchers try to control for error in their simulations, and fix parameters, in ways that seem analogous to how experimenters calibrate their devices. Simulations build up credibility over long time scales and may have lives of their own independent of developments in other parts of science. These observations suggest a potentially rich analogy between simulations and Hacking's account of experimentation [5.29]. In a normative step, based on these links, *Parker* [5.10] has suggested that in fact *Mayo*'s [5.30] rigorous error-statistical approach for experimentation should be an appropriate starting point for more thorough evaluation of the results of simulations. Simulations need to be evaluated by the degree to which they avoid false positives when it comes to testing hypotheses by successfully controlling for potential sources of error that creep in during the simulation process. At the same time a rather vigorous debate has emerged concerning the clarification of the precise epistemological dissimilarities or disanalogies between simulation and traditional experimentation (see for instance [5.31–36]). This question is in itself of independent philosophical interest for assessing the benefits and value of each as alternatives, but should also help define the limits of the relevance of experimentation as a model for understanding and assessing simulation practices.

From our perspective, however, the new methodological and epistemological strategies that modelers are introducing in order to construct and guarantee the reliability of simulation models could prove to be the most interesting and novel aspect of simulation with which philosophers will have to grapple. Indeed, while much attention has focused on the contrasts and similarities between simulations, experiments and simulation experiments, no one has called attention to the fact that real-world experiments and simulations are also being used in concert to enhance the ability of researchers to handle uncertain complex systems. One of the labs we have studied conducts *bimodal* modeling, where the modelers conduct their own experiments in the service of building their models. We have analyzed the case of one modeler's behavior in which model-building, simulation and experimentation were tightly interwoven [5.37]. She used a conjunction of experiment and simulation to triangulate on errors and uncertainties in her model, thus demonstrating that the two can be combined in practice in sophisticated ways.

Part A | 5.3

Her model-building would not have been possible without the affordances of both simulation and her ability to perform experimentation precisely adapted to test questions about the model as she was in the process of formulating it. Simulation and experiment closely coupled in this fashion offers the possibility of extending the capacity to produce reliable models of complex phenomena.

Bimodal modeling is relatively easy to characterize epistemologically since experimentation is used to validate and check the simulations as the model is being constructed. Simulations are not relied on independent of experimental verification. Often, however, experimental or any kind of observational data are hard to come by for practical or theoretical reasons. More philosophically challenging will be to evaluate the new epistemological strategies researchers are in fact developing for drawing inferences in these often deeply uncertain and complex contexts with the aid of computation. *Parker* [5.38, 39], for instance, identifies the practice in climate science and meteorology of ensemble modeling. No theory of model-building exists that tells climate and weather modelers how to go from physical theory to reliable models. Different formulations using different initial conditions, models structures and different parameterizations of those models that fit the observational data can be developed from the physical theory. In this situation modelers average over results from large collections of models, using different weighting schemas, and argue for the validity of these results on the basis that these models collectively represent the possibility space. However, considerable philosophical questions emerge as to the underlying justifiability of these ensemble practices and the probability weightings being relied upon. Background theory can provide little guidance in this context and in the case of climate modeling there is little chance for predictively testing performance. Further, the robustness of particular ensemble choices is often very low and justifications for picking out particular ensembles are rarely carefully formulated.

The ability to generate and compare large numbers of complex models in this way is a development of modern computational power. In our studies we have also come across novel argumentation, particularly connected with parameter-fixing [5.40]. Because the parameter spaces these modelers have to deal with are so complex, there is almost no chance of getting a best fit solution. Instead modelers produce multiple models often using Monte Carlo techniques that converge on similar behavior and output. These models have different parameterizations and ultimately represent the underlying mechanisms of the systems differently. However, modelers can nonetheless make specific arguments about network structure and dynamic relationships among specific variables. There is not usually any well-established theory that licenses these arguments. The fact that the models converge on the same relevant results is motivation for inferring that these models are right at least about those aspects of the system for which they are designed to account. Unfortunately, because access to real-world experimentation is quite difficult, it is hard to judge how reliable this technique is in producing robust models. What is novel about this kind of strategy is that it implicitly treats parameter-fixing as an opportunity, not just a problem, for modelers. If instead of trying to capture the dynamics of whole systems modelers just fix their goals on capturing robust properties and relations of a system, the potential of finding results that work within these constraints in large parameter-spaces increases, and from the multiple models obtained modelers can pare down to those that converge. The more complex problem thus seems to allow a pathway for solving a simpler one. Nonetheless, whether we should accept these kinds of strategies as reliable and the models produced as robust remains the fundamental question, and an overarching question for the field itself. It is a reasonable reaction to suspect that something important is being given up in the process, which will affect how well scientists can assess the reliability and importance of the models they produce. Whether the power computational processes can adequately compensate for the potential distortions or errors introduced is one of the most critical and novel epistemological questions for philosophy today.

The kinds of epistemological innovations we have been considering raise deeper questions about the purposes of simulation, particularly in terms of traditional epistemic categories like understanding, explanation and so on. Of course at one extreme some simulations of the purely data-driven kind is purely phenomenological. Theory plays no role in its generation, and is not sought as its outcome. However in other cases some form of understanding at least is sought. In many cases though, where theory might be thought the essential agent of understanding, the complexity of the equations and resulting complexity of the computational processes that instantiate them, simply block any way of decomposing the theory or theoretical model in order to understand how the theory might explain a phenomena and thus assess the accuracy and plausibility of the underlying mechanisms it might prescribe. *Humphreys* labels this *epistemic opacity* [5.11]. *Lenhard* [5.41] in turn identifies a form of pragmatic understanding that can replace theoretical understanding when a simulation model is epistemically opaque. This form of understanding is pragmatic in the sense of being an understanding of how to *control* and *manipulate* phenomena, rather explain them using background theoretical principles and laws. Settling for this form of understanding is a choice made by researchers in order to handle more complex problems and systems using simulations. But it is a novel one in the context of physics and chemistry. In systems biology we recognize something similar [5.40]. Researchers give up accurate mechanistic understanding of their systems for more pragmatic goals of gaining network control, at least over specific variables. To do so they use simplification and parameter-fitting techniques that obscure the extent to which their models capture the underlying mechanisms. Mechanistic explanation is thus given up, for some weaker form of understanding.

Finally, computational modeling and simulation in the situations we have been considering in this section are driving a profound shift in the nature and level of human cognitive engagement in scientific production processes and their outputs [5.12, 24, 25, 42, 43]. So much of philosophy of science has been based on intuitive notions of human cognitive abilities. Our concepts of explanation and understanding are constructed implicitly on the basis of what we can grasp as humans. With simulation and big-data science those kinds of characterizations may no longer be accurate or relevant [5.44].

5.4 Computational Simulation and Human Cognition

It is on this last point that we turn to consider the ways in which human cognitive processes are implicated in processes of simulation model-building. Computational science, of the nonbig data or nonmachine learning kind which we have focused on here, is as Humphrey's calls it, a "hybrid scenario" as opposed to an "automated scenario" [5.12, p. 616]. In his words:

> "This distinction is important because in the hybrid scenario, one cannot completely abstract from human cognitive abilities when dealing with representational and computational issues.... We are now faced with a problem, which we can call the *anthropocentric predicament*, of how we, as humans, can understand and evaluate computationally-based scientific methods that transcend our own abilities."

Unlike machine-learning contexts, computational modeling is in many cases a practice of using computation to extend traditional modeling practices and our own capabilities to draw insight out of low-data contexts and complex systems for which theory provides at best a limited guide. In this way cognitive capacities are often heavily involved. The *hybrid* nature of computational science thus motivates the need for understanding how human agents cognitively engage with and control opaque computational processes, and in turn draw information out of them. Evaluating these processes – their productiveness and reliability – requires in the first step having some understanding of them. As we will see, although computational calculation processes are beyond our abilities, at least in the case of systems biology the use of computation by modelers is often far more integrated with their own cognitive processes and understanding, and thus far more under their control, than we might think.

As we have seen there are several lines of philosophical research on computational simulation that underscore it is through the processes of model-building – taken to comprise the incremental and interwoven processes of constructing the model and investigating its dynamics through simulation – that the modeler comes to develop at least a pragmatic understanding of the phenomena under investigation. Complex systems, such as investigated in systems biology, present perhaps the extreme case in which these practices are the *primary* means through which modelers, mostly nonbiologists, develop understanding of the systems. In our investigations, modelers called the building and *running* of their models under various conditions *getting a feel for the model*, which enables them to get a feel for the dynamics of the system.

In our investigations we have witnessed that modelers (mainly engineers) with little understanding of biology have been able to provide novel insights and highly significant predictions, later confirmed by biological collaborators, for the systems they are investigating through simulation. How is it possible that engineers with little to no biological training can be making significant biological discoveries? A related question concerns how complete novices are making scientific discoveries through simulations crowdsourced by means of video games such as Foldit and EteRNA, which appear to enable nonscientists to quickly build accurate/veridical structures representing molecular entities they had no prior knowledge of [5.45, 46]. *Nersessian* and *Chadrasekharan*, individually and together [5.24, 25, 42, 47–49], have argued that the answer to this question lies in understanding how computational simulation enhances human cognition in discovery processes. Because of the visual and manipulative nature of the crowdsourcing cases, the answer points in the direction of the *coupling* of the human sensorimotor systems with simulation models. These crowdsourcing models rerepresent conceptual knowledge developed by the sci-

Part A | 5.4

entific community (e.g., structure of proteins) as computational representations with a control interface that can be manipulated through the gamer's actions. The interface enables these novices to build new representations drawing on tacit/implicit sensorimotor processes. Although the use of crowdsourcing simulations in scientific problem solving is new, the human sensorimotor system has been used explicitly to detect patterns, especially in dynamic data generated by computational models, since the dawn of computational modeling. Entire disciplines and methods have been built using visualized patterns on computer screens. Complexity theory [5.50, 51], artificial life [5.52, 53] and computational chemistry [5.54, 55] provide a few exemplars where significant discoveries have been made.

Turning back now to the computational simulations used by scientists that we have been discussing, all of the above suggests that the model-building processes facilitate a close coupling between the model and the researcher's mental modeling processes even in the absence of a dynamic visualization. The building process manipulates procedural and declarative knowledge in the imagination and in the representation, creating a *coupled cognitive system* of model and modeler [5.25, 42, 43, 48, 56, 57]. This coupling can lead to explicit understanding of the dynamics of the system under investigation. The notion of a coupled cognitive system is best understood in terms of the framework of distributed cognition [5.58, 59], which was developed to study cognitive processes in complex task environments, particularly where external representations and other cognitive artifacts and, possibly, groups of people, accomplish the task. The primary unit of analysis is the socio-technical system that generates, manipulates and propagates representations (internal and external to people). Research leading to the formation of the distributed cognition framework has focused largely on the use of existing representational artifacts and less so on the building/creation of the artifacts. The central metaphor is that of the human *offloading* complex cognitive processes such as memory to the artifact, which, for example, in the canonical exemplar of the speed bug that marks critical airspeeds for a particular flight, replaces complex cognitive operations with a perceptual operation and provides a publically available representation that is shared between pilot and co-pilot.

In the research cited above, we have been arguing that *offloading* is not the right metaphor for understanding the cognitive enhancements provided through the building of novel computational representations. Rather, the metaphor should be that of *coupling* between internal and external representations. Delving into the modifications needed of the distributed cognition framework to accommodate the notion of a coupled cognitive system would take use too far afield in this review (but see [5.25]). Instead, we will flesh out the notion a bit by noting some of the ways in which building and using simulation models enhance human cognitive capabilities and, in particular, extend the capability of the imagination system for simulative model-based reasoning.

A central, but yet not well-researched premise of distributed cognition is, as Hutchins has stated succinctly, that "humans create cognitive powers by creating the environments in which they exercise those powers" [5.58, p. 169]. Since building modeling-environments for problem solving is a major component of scientific research [5.49], scientific practices provide an especially good locus for examining the human capability to extend and create cognitive powers. In the case of simulation model-building, the key question is: *What are the cognitive changes involved in building a simulation model and how do these lead to discoveries?* The key cognitive change is that over the course of many iterations of model-construction and simulation, the model gradually becomes coupled with the modeler's imagination system (mental model simulation), which enables the modeler to explore different scenarios. The coupling allows *what if* questions in the mind of the modeler to be turned into detailed explorations of the system, which would not be possible in the mind alone. The computational model enables this exploration because as it is incrementally built using many data sets, the model's behavior, in the systems biology case, for instance, comes to parallel the dynamics of the pathway. Each replication of experimental results adds complexity to the model and the process continues until the model is judged to fit all available data well. This judgment is complex, as it is based on a large number of iterations where a range of factors such as sensitivity, stability, consistency, computational complexity and so forth are explored. As the model gains complexity it starts to reveal or expose many details of the system's behavior enabling the modeler to interrogate the model in ways that are not possible in the mind alone (thought experimenting) or in real-world experiments. It makes evident many details of the system's behavior that the modeler could not have imagined alone because of the fine grain and complexity of the details.

The parallel between computation simulation experimenting and thought experimenting is one philosophers have commented on, but the current framing of the discussion primarily centers on the issue of interpreting simulations and whether computational simulations should be construed as *opaque* thought experiments [5.60, 61]. *Di Paolo* et al. [5.60] have argued that computational models are more opaque than thought experiments, and as such, require more *system-*

atic enquiry through probing of the model's behavior. In a similar vein, *Lenhard* [5.61] has claimed that thought experiments are more *lucid* than computational models, though it is left unclear what is meant by *lucid* in this context, particularly given the extensive discussions around what specific thought experiments actually demonstrate. In the context of the discussion of the relation of thought experimenting and computational simulation, we have argued that the discussion should be shifted from issues of interpretation to a process-oriented analysis of modeling [5.47]. *Nersessian* [5.62] casts thought experimenting as a form of *simulative model-based reasoning*, the cognitive basis of which is the human capacity for mental modeling. Thought experiments (conceptual models), physical models [5.63] and computational models [5.47, 48] form a spectrum of simulative model-based reasoning in that all these types of modeling generate and test counterfactual situations that are difficult (if not impossible) to implement in the real world. Both thought experiments and computational models support simulation of counterfactual situations, however, while thought experiments are built using concrete elements, computational models are built using variables. Simulating counterfactual scenarios beyond the specific one constructed in the thought experiment is difficult and requires complex cognitive transformations to move away from the concrete case to the abstract, generic case. On the other hand, computational simulation constructs the abstract, generic case from the outset. Since computational models are made entirely of variables, they naturally support thinking about parameter spaces, possible variations to the design seen in nature, and why this variation occurs rather than the many others that are possible.

Thought experiments are a product of a resource environment in science where the only tools available were writing implements, paper (blackboards, etc.) and the brain. Computational models create cognitive enhancements that go well beyond those resources and enable scientists to study the complex, dynamic and nonlinear behaviors of the phenomena that are the focus of contemporary science.

Returning to the nature of the cognitive enhancements created, the coupling of the computational model with the modeler's imagination system significantly enhances the researcher's natural capacity for simulative model-based reasoning, particularly in the following ways:

- It allows running many more simulations, with many variables at gradients not perceivable or manipulable by the mind, which can be compared and contrasted.
- It allows testing what-if scenarios with changes among many variables that would be impossible to do in the mind.
- It allows stopping the simulation at various points and checking and tracking its states. If some desirable effect is seen, variables can be tweaked in process to get that effect consistently.
- It allows taking the system apart as modules, simulating them, and putting them together in different combinations.
- It allows changing the time in which intermediate processes kick in.

These complex manipulations expose the modeler to system-level behaviors that are not possible to examine in either thought alone or in real-world experimentation. The processes involved in building the distributed model-based reasoning system comprising simulation model and modeler enhance several cognitive abilities. Here we will conclude by considering three (for a fuller discussion see [5.25]). First, the model-building process brings together a range of experimental data. Given Internet search engines and online data bases, current models synthesize more data than even before and create a synthesis that exists nowhere in the literature and would not be possible for modelers or biologists to produce on their own. In effect, the model becomes a *running literature review*. Thus, modeling enhances the synthesizing and integrating capabilities of the modeler, which is an important part of the answer as to how a modeler with scant biological knowledge can make important discoveries. Second, an important cognitive effect of the model-building is to enhance the modeler's powers of abstraction. Most significantly, through the gradual process of thousands of runs of simulations and analyses of system dynamics for these, the modeler gains an external, global view of the system as a whole. Such a global view would not be possible to develop just from mental simulation, especially since the interactions among elements are complex and difficult to keep track of separately. The system view, together with the detailed understanding of the dynamics, provides the modeler with an intuitive sense (*a feeling for the model*) of the biological mechanisms that enables her to extend the pathway structure in a constrained fashion to accommodate experimental data that could not be accounted for by the current pathway from which the model started. Additionally, this intuitive sense of the mechanism built from interaction with the model helps to explain the success of the crowdsourcing models noted above (see also [5.64]).

Finally, the model enhances the cognitive capacity for counterfactual or possible-worlds thinking. As noted in our discussion of thought experimenting, the model-building process begins by capturing the reactions/interactions using variables. Variables provide a place-holder representation, which when interpreted with combinations of numbers for these variables, can generate model data that parallels the known experimental data. One interesting feature of the place-holder representation is that it provides the modeler with a flexible way of thinking about the reactions, as opposed to the experimentalist who works with only one set of values. Once the model is using the experimental values, the variables can take any set of values, as long as they generate a fit with the experimental data. The modeler is able to think of the real-world values as only *one possible scenario*, to examine why this scenario is commonly seen in nature, and envision other scenarios that fit. Thinking in variables supports both the objective modelers often have of altering or redesigning a reaction (such as the thickness of lignin in plant wall for biofuels) and the objective of developing generic design patterns and principles. More broadly, the variable representation significantly expands the imagination space of the modeler, enabling counterfactual explorations of possible worlds that far outstrip the potential of thought experimenting alone.

A more microscopic focus like this one on the actual processes by which computational simulation is coupled with the cognitive processes of the modeler begins to help break down some of the mystery and seeming inscrutability surrounding computation conveyed by the idea that computational processes are offloaded automated processes from which inferences are derived. The implications of this research into hybrid nature of simulation modeling are that modelers might often have more control over and insight into their models and their alignment with the phenomena than philosophers have realized. Given the emphasis placed in published scientific literature on fitting the data and predictive success for validating simulations, we might be missing out on the important role that these processes internal to the model-building or discovery context appear to be playing (from a microanalysis of practice) in support of the models constructed. Indeed, the ability of computational modeling to support highly exploratory investigative processes makes it particularly relevant for philosophers to have fine-grained knowledge of model-building processes in order to begin to understand why models work as well as they do and how reliable they can be considered to be.

Acknowledgments. We gratefully acknowledge the support of the US National Science Foundation grant DRL097394084. Our analysis has benefited from collaboration with members of the Cognition and Learning in Interdisciplinary Cultures (CLIC) Research Group at the Georgia Institute of Technology, especially with Sanjay Chandrasekharan. Miles MacLeod's participation was also supported by a postdoctoral fellowship at the TINT Center, University of Helsinki.

References

5.1 E. Winsberg: Sanctioning models: The epistemology of simulation, Sci. Context **12**(2), 275–292 (1999)

5.2 E. Winsberg: Models of success vs. the success of models: Reliability without truth, Synthese **152**, 1–19 (2006)

5.3 E. Winsberg: Computer simulation and the philosophy of science, Philos. Compass **4/5**, 835–845 (2009)

5.4 E. Winsberg: *Science in the Age of Computer Simulation* (Univ. of Chicago Press, Chicago 2010)

5.5 E. Winserg: Computer simulations in science. In: *The Stanford Encyclopedia of Philosophy*, ed. by E.N. Zalta (Stanford Univ., Stanford 2014), http://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=simulations-science

5.6 N. Cartwright: *The Dappled World: A Study of the Boundaries of Science* (Cambridge Univ. Press, Cambridge 1999)

5.7 M.S. Morgan, M. Morrison: Models as mediating instruments. In: *Models as Mediators: Perspectives on Natural and Social Science*, ed. by M.S. Morgan, M. Morrison (Cambridge Univ. Press, Cambridge 1999)

5.8 J. Lenhard: Computer simulation: The cooperation between experimenting and modeling, Philos. Sci. **74**(2), 176–194 (2007)

5.9 E. Winsberg: Simulations, models, and theories: Complex physical systems and their representations, Philos. Sci. **68**(3), 442–454 (2001)

5.10 W. Parker: Computer simulation through an error-statistical lens, Synthese **163**(3), 371–384 (2008)

5.11 P. Humphreys: *Extending Ourselves: Computational Science, Empiricism, and Scientific Method* (Oxford Univ. Press, New York 2004)

5.12 P. Humphreys: The philosophical novelty of computer simulation methods, Synthese **169**, 615–626 (2009)

5.13 W. Parker: Computer simulation. In: *The Routledge Companion to Philosophy of Science*, ed. by S. Psillos, M. Curd (Routledge, London 2013) pp. 135–145

5.14 E. Fox Keller: Models, simulation, and computer experiments. In: *The Philosophy of Scientific Experimentation*, ed. by H. Radder (Univ. of Pittsburgh Press, Pittsburgh 2003) pp. 198–215

5.15 S. Peck: Agent-based models as fictive instantiations of ecological processes, Philos. Theory Biol. **4**, 1–12 (2012)

5.16 T. Grüne-Yanoff, P. Weirich: Philosophy of simulation, simulation and gaming, Interdiscip. J. **41**(1), 1–31 (2010)

5.17 M.A. Bedau: Weak emergence and computer simulation. In: *Models, Simulations, and Representations*, ed. by P. Humphreys, C. Imbert (Routledge, New York 2011) pp. 91–114

5.18 S. Peck: The Hermeneutics of ecological simulation, Biol. Philos. **23**(3), 383–402 (2008)

5.19 R. Frigg: Models and fiction, Synthese **172**(2), 251–268 (2010)

5.20 T. Grüne-Yanoff: The explanatory potential of artificial societies, Synthese **169**(3), 539–555 (2009)

5.21 M. MacLeod, N.J. Nersessian: Building simulations from the ground-up: Modeling and theory in systems biology, Philos. Sci. **80**(4), 533–556 (2013)

5.22 E.O. Voit: *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists* (Cambridge Univ. Press, Cambridge 2000)

5.23 M. MacLeod, N.J. Nersessian: The creative industry of systems biology, Mind Soc. **12**, 35–48 (2013)

5.24 S. Chandrasekharan, N.J. Nersessian: Building cognition: The construction of external representations for discovery, Cogn. Sci. **39**(8), 1727–1763 (2015), doi:10.1111/cogs.12203

5.25 S. Chandrasekharan, N.J. Nersessian: Building cognition: The construction of computational representations for scientific discovery, Cogn. Sci. **39**(8), 1727–1763 (2015)

5.26 H. Kitano: Looking beyond the details: A rise in system-oriented approaches in genetics and molecular biology, Curr. Genet. **41**(1), 1–10 (2002)

5.27 H.V. Westerhoff, D.B. Kell: The methodologies of systems biology. In: *Systems Biology: Philosophical Foundations*, ed. by F.C. Boogerd, F.J. Bruggeman, J.S. Hofmeyr, H.V. Westerhoff (Elsevier, Amsterdam 2007) pp. 23–70

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

5.29 E. Winsberg: Simulated experiments: Methodology for a virtual world, Philos. Sci. **70**(1), 105–125 (2003)

5.30 D.G. Mayo: *Error and the Growth of Experimental Knowledge* (Univ. of Chicago Press, Chicago 1996)

5.31 N. Gilbert, K. Troitzsch: *Simulation for the Social Scientist* (Open Univ. Press, Philadelphia 1999)

5.32 F. Guala: Models, simulations, and experiments. In: *Model-based reasoning: Science, technology, values*, ed. by L. Magani, N.J. Nersessian (Kluwer Academic/Plenum Publishers, New York 2002) pp. 59–74

5.33 F. Guala: Paradigmatic experiments: The ultimatum game from testing to measurement device, Philos. Sci. **75**, 658–669 (2008)

5.34 M. Morgan: Experiments without material intervention: Model experiments, virtual experiments and virtually experiments. In: *The Philosophy of Scientific Experimentation*, ed. by H. Radder (University of Pittsburgh Press, Pittsburgh 2003) pp. 216–235

5.35 W. Parker: Does matter really matter? Computer simulations, experiments and materiality, Synthese **169**(3), 483–496 (2009)

5.36 E. Winsberg: A tale of two methods, Synthese **169**(3), 575–592 (2009)

5.37 M. MacLeod, N.J. Nersessian: Coupling simulation and experiment: The bimodal strategy in integrative systems biology, Stud. Hist. Philos. Sci. Part C **44**, 572–584 (2013)

5.38 W.S. Parker: Predicting weather and climate: Uncertainty, ensembles and probability, Stud. Hist. Philos. Sci. Part B **41**(3), 263–272 (2010)

5.39 W.S. Parker: Whose probabilities? Predicting climate change with ensembles of models, Philos. Sci. **77**(5), 985–997 (2010)

5.40 M. MacLeod, N.J. Nersessian: Modeling systems-level dynamics: Understanding without mechanistic explanation in integrative systems biology, Stud. Hist. Philos. Sci. Part C **49**(1), 1–11 (2015)

5.41 J. Lenhard: Surprised by a nanowire: Simulation, control, and understanding, Philos. Sci. **73**(5), 605–616 (2006)

5.42 N.J. Nersessian: *Creating Scientific Concepts* (MIT Press, Cambridge 2008)

5.43 N.J. Nersessian: How do engineering scientists think? Model-based simulation in biomedical engineering research laboratories, Top. Cogn. Sci. **1**, 730–757 (2009)

5.44 W. Callebaut: Scientific perspectivism: A philosopher of science's response to the challenge of big data biology, Stud. Hist. Philos. Sci. Part C **43**(1), 69–80 (2012)

5.45 J. Bohannon: Gamers unravel the secret life of protein, Wired **17** (2009), http://www.wired.com/medtech/genetics/magazine/17-05/ff_protein, Last accessed 06-06-2016

5.46 F. Khatib, F. DiMaio, Foldit Contenders Group, Foldit Void Crushers Group, S. Cooper, M. Kazmierczyk, M. Gilski, S. Krzywda, H. Zabranska, I. Pichova, J. Thompson, Z. Popovic, M. Jaskolski, D. Baker: Crystal structure of a monomeric retroviral protease solved by protein folding game players, Nat. Struct. Mol. Biol. **18**(10), 1175–1177 (2011)

5.47 S. Chandrasekharan, N.J. Nersessian, V. Subramanian: Computational modeling: Is this the end of thought experiments in science? In: *Thought Experiments in Philosophy, Science and the Arts*, ed. by J. Brown, M. Frappier, L. Meynell (Routledge, London 2013) pp. 239–260

5.48 S. Chandrasekharan: Building to discover: A common coding model, Cogn. Sci. **33**(6), 1059–1086 (2009)

5.49 N.J. Nersessian: Engineering concepts: The interplay between concept formation and modeling practices in bioengineering sciences, Mind Cult. Activ. **19**, 222–239 (2012)

5.50 C.G. Langton: Self-reproduction in cellular automata, Physica D **10**, 135–144 (1984)

5.51 C.G. Langton: Computation at the edge of chaos: Phase transitions and emergent computation, Physica D **42**, 12–37 (1990)

5.52 C. Reynolds: Flocks, herds, and schools: A distributed behavioral model, Comp. Graph. **21**(4), 25–34 (1987)

5.53 K. Sims: Evolving 3D morphology and behavior by competition, Artif. Life **1**(4), 353–372 (1994)

5.54 W. Banzhaf: Self-organization in a system of binary strings. In: *Artificial Life IV*, ed. by R. Brooks, P. Maes (MIT Press, Cambridge MA 2011) pp. 109–119

5.55 L. Edwards, Y. Peng, J. Reggia: Computational models for the formation of protocell structure, Artif. Life **4**(1), 61–77 (1998)

5.56 N.J. Nersessian, E. Kurz-Milcke, W.C. Newstetter, J. Davies: Research laboratories as evolving distributed cognitive systems, Proc. 25th Annu. Conf. Cogn. Sci. Soc. (2003) pp. 857–862

5.57 L. Osbeck, N.J. Nersessian: The distribution of representation, J. Theor. Soc. Behav. **36**, 141–160 (2006)

5.58 E. Hutchins: *Cognition in the Wild* (MIT Press, Cambridge 1995)

5.59 E. Hutchins: How a cockpit remembers its speeds, Cogn. Sci. **19**(3), 265–288 (1995)

5.60 E.A. Di Paolo, J. Noble, S. Bullock: Simulation models as opaque thought experiments. In: *Artificial Life VII*, ed. by M.A. Bedau, J.S. McCaskill, N.H. Packard, S. Rasmussen (MIT Press, Cambridge 2000) pp. 497–506

5.61 J. Lenhard: When experiments start. Simulation experiments within simulation experiments, Int. Workshop Thought Exp. Comput. Simul. (2010)

5.62 N.J. Nersessian: In the theoretician's laboratory: Thought experimenting as mental modeling, Proc. Philos. Assoc. Am., Vol. 2 (1992) pp. 291–301

5.63 N.J. Nersessian, C. Patton: Model-based reasoning in interdisciplinary engineering. In: *Handbook of the Philosophy of Technology and Engineering Sciences*, ed. by A. Meijers (Elsevier, Amsterdam 2009) pp. 687–718

5.64 S. Chandrasekharan: Becoming knowledge: Cognitive and neural mechanisms that support scientific intuition. In: *Rational Intuition: Philosophical Roots, Scientific Investigations*, ed. by L.M. Osbeck, B.S. Held (Cambridge University Press, Cambridge 2014) pp. 307–337