SPRINGER BRIEFS IN PHILOSOPHY

Axel Gelfert

How to Do Science with Models

A Philosophical Primer

Springer

SpringerBriefs in Philosophy

More information about this series at http://www.springer.com/series/10082

"This is a truly excellent book. Not only does it provide insightful analysis of contemporary philosophical accounts of modelling, but it draws our attention to important yet unexplored questions related to the exploratory function of models and their connection to issues in the philosophy of technology. By focusing our attention on a broad range of examples it provides the best systematic treatment of scientific modelling to appear in many years. Highly recommended!" **Margaret Morrison**, *University of Toronto*.

Axel Gelfert

How to Do Science with Models

A Philosophical Primer

Springer

Axel Gelfert
Department of Philosophy
National University of Singapore
Singapore
Singapore

ISSN 2211-4548 ISSN 2211-4556 (electronic)
SpringerBriefs in Philosophy
ISBN 978-3-319-27952-7 ISBN 978-3-319-27954-1 (eBook)
DOI 10.1007/978-3-319-27954-1

Library of Congress Control Number: 2015958321

Printed on acid-free paper

This Springer imprint is published by SpringerNature
The registered company is Springer International Publishing AG Switzerland

Preface

After volcanic ash from the eruption of the Icelandic volcano Eyjafjallajökull in 2010 had shut down air traffic across the Atlantic Ocean for several days in a row, an angry airline CEO appeared in a television interview with the BBC and blamed civil aviation authorities for basing their decision to close the transatlantic airspace on 'mere models.' While the CEO's frustration may have been understandable from a business point of view, from the viewpoint of science it was a rather disingenuous way of reacting: After all, modern aircraft, too, are designed on the basis of models of flow, turbulence, atmospheric motion, and material behavior, which embody basically the same fundamental theoretical principles, whether one is dealing with the distribution of volcanic ash or its effects on jet engines. Modern science and technology are saturated with models—so much so that it is difficult to imagine what the modern scientific world would look like without the use of models. The ubiquity of models in contemporary science and technology is hardly news to any working scientist or engineer, but the realization that scientific inquiry and technological innovation are inextricably intertwined with scientific models has not yet sunk in with the general public and its representatives. Consider the case of climate change: Even today, it is not uncommon to come across pundits and politicians who dismiss the carefully cross-checked predictions of climate scientists on the grounds that they are 'just based on models'—yet the very same people then happily go on to make policy on the basis of (model-based) forecasts of economic growth. Models, then, are all around us, whether in the natural or social sciences, and any attempt to understand how science works had better account for, and make sense of, this basic fact about scientific practice.

This book is an attempt to come to philosophical terms with the ubiquity and indispensability of models in contemporary science and technology. As such, it is a contribution to a growing body of work by scholars in the history and philosophy of science. Historians and sociologists of science, over the past twenty-odd years or so, have amassed a vast number of case studies that describe and analyze specific scientific models in great detail. At the same time, a lively philosophical debate has developed, which focuses on general questions concerning the nature of models and the possibility of model-based representation. Yet, too often, these two projects—the in-depth study of specific cases of scientific models and the abstract concern for model-based representation—have stood side by side with one another, without entering into a true dialogue. By contrast, one of the guiding methodological assumptions of this book is that descriptive adequacy and normative–theoretical ambition need not be mutually exclusive: As I hope to show, careful attention to scientific modeling as a practice may itself be a source of insight about what gives model-based science its cohesion and makes it successful—and about what its limitations are. At the heart of this approach is the thought that the key to answering any of the more general philosophical questions about scientific models lies in the diversity of their varied uses and functions.

The structure of this book is as follows. The first two chapters provide a concise survey of the existing philosophical debate about scientific models, first from an ontological angle, by tackling the question 'What are Scientific Models?' (Chap. 1), and then by addressing the problem of scientific representation in relation to scientific models and theories (Chap. 2). While the main focus is on systematic questions, both chapters also retrace some of the historical trajectory of the debate, for example by showing how our current notion of 'scientific model' is indebted to the nineteenth-century notion of 'mechanical analogy' (Chap. 1, Sect. 1.2), or how philosophers in the twentieth century—especially in the wake of Nelson Goodman's philosophy of art—have reconsidered the notion of (scientific) representation (Chap. 2, Sect. 2.2). Chapter 3 looks in detail at a number of case studies from across the natural sciences in order to identify recurring strategies of model building. Examples discussed range from population biology (Lotka–Volterra model) to condensed matter physics (BCS and Ginzburg–Landau models of superconductivity); special attention is given to the question of whether modeling necessarily involves trade-offs between different theoretical desiderata (such as generality and precision) and whether the existence of trade-offs can serve as a demarcation criterion between different scientific disciplines, notably biology and physics. The final two chapters advance the philosophical debate in distinct ways, by identifying a number of previously overlooked functions and uses of scientific models. Thus, Chap. 4 discusses exploratory uses of scientific models and seeks to establish exploration as one of the core functions of scientific modeling, alongside the more traditional goals of explanation and prediction. Chapter 5, finally, links the debate about scientific models to questions in the philosophy of technology, in particular the question of how artifacts simultaneously enable and constrain certain actions and how we, as users of such artifacts, engage with them at a phenomenological level. Models, I conclude, are not simply neutral tools that we use at will to represent aspects of the world around us; rather, they contribute new elements—which are neither to be found in the underlying 'fundamental theory' nor to be found in the empirical data—to the process of scientific inquiry and, by mediating between different types of user–model–world relations, enable the generation of new scientific knowledge.

My philosophical interest in scientific models began when, as a physics student studying quantum many-body models, I first realized that the very same models could be used to describe radically different target systems and were sometimes invoked by different researchers in support of incompatible research agendas. Yet, in spite of this diversity of uses and functions of models, there is also a palpable sense in which model-based science is marked by great cohesion and has vastly improved our scientific understanding of the world around us. After a dozen or so years of thinking and writing about scientific models, I am now more convinced than ever that the strength of models as tools of inquiry lies precisely in their diversity and flexibility. While the choice of examples in this book—notably, the prominence given to models from many-body physics—no doubt reflects the early origins of my interest in models, special care has been taken to also include examples from disciplines such as biology, chemistry, and sociodynamics. While all the material in this book has been thoroughly rewritten, several of the chapters draw on previously published (or, in some cases, forthcoming) work. Thus, Chaps. 1 and 2 draw on material from my chapter 'The Ontology of Scientific Models' in the forthcoming *Springer Handbook of Model-Based Science* (eds. Lorenzo Magnani and Tommaso Bertolotti). Chapter 3 (esp. Sects. 3.6 and 3.7) overlaps with my paper 'Strategies of Model-Building in Condensed Matter Physics: Trade-Offs as a Demarcation Criterion Between Physics and Biology?', *Synthese*, Vol. 190, No. 2, 2013, pp. 252–273. Section 5.2 of Chap. 5 is based on my discussion note 'Symbol Systems as Collective Representational Resources: Mary Hesse, Nelson Goodman, and the Problem of Scientific Representation,' *Social Epistemology Review and Reply Collective*, Vol. 4, No. 6, 2015, pp. 52–61, while Sect. 5.3 of the same chapter (along with Chap. 3, Sect. 3.3) is heavily indebted to my chapter 'Between Rigor and Reality: Many-Body Models in Condensed Matter Physics' in Brigitte Falkenburg's and Margaret Morrison's jointly edited volume *Why More is Different: Philosophical Issues in Condensed Matter Physics and Complex Systems*, Heidelberg: Springer 2015, pp. 201–226.

Over the years, I have had the good fortune to encounter many sympathetic and supportive colleagues and scholars from whom I have learnt a great deal about the philosophy of scientific models. While it would be impossible to name all of them and acknowledge every single influence on my thinking about models, I do wish to acknowledge the following individuals, all of whom in one way or another have personally left their mark on the work presented here—whether by sending me written comments, by participating in joint workshops, or by simply making time to discuss my work on models during a coffee break at a conference: Anna Alexandrova, Sorin Bangu, Ann-Sophie Barwich, Robert Batterman, Justin Biddle, Agnes Bolinska, Marcel Boumans, Alex Broadbent, Anjan Chakravartty, Hasok Chang, Chuanfei Chin, Jeremy Chong, Tamás Demeter, Paul Dicken, Steffen Ducheyne, Kevin Elliott, Brigitte Falkenburg, Uljana Feest, Stephan Hartmann, Michael Heidelberger, Mary Hesse, Paul Humphreys, Cyrille Imbert, Stephen John, Jaakko Kuorikoski, Martin Kusch, Sabina Leonelli, Lorenzo Magnani, Simone Mahrenholz, Uskali Mäki, John Matthewson, Cornelis Menke, Boaz Miller, Teru Miyake, Jacob Mok, Mary Morgan, Robert Nola, Alfred Nordmann, Wendy Parker, Chris Pincock, Demetris Portides, Hans-Jörg Rheinberger, Mauricio Suárez, Adam Toon, Marion Vorms, and Jeff White. I am especially grateful to Gabriele

Gramelsberger, Tarja Knuuttila, and Margaret Morrison for their long-standing support and continued collaboration.

I have profited enormously from a reading group which convened four times over the course of two weeks in July 2015 and during which an almost complete draft of the book was discussed. I am especially grateful to Grant Fisher, Lina Jansson, Eric Kerr, Joel Chow, Bernadette Chin, and Will Zhang for taking the time to engage closely with my arguments and for offering a host of useful suggestions (along with several important corrections). Thanks are due to the Office of the Deputy President (Research and Technology), Humanities and Social Sciences Division, National University of Singapore, for providing funding for this reading group ('The Epistemology of Technology and Technoscience,' WBS: R-106-000-040-646). Chapter 4 ('Exploratory Uses of Scientific Models') was conceived and written while on sabbatical leave in Berlin during the first half of 2015; early versions of the argument were presented at the philosophy departments of Leibniz Universität Hannover, Technische Universität Darmstadt, and Freie Universität Berlin. I am grateful to audiences on all three occasions for their valuable feedback and to Sybille Krämer for hosting me in her research group during the Berlin portion of my sabbatical. Funding for research assistance and artwork was provided by the Faculty of Arts and Social Sciences, National University of Singapore, in the form of a book grant (WBS: R-106-000-048-133), for which I am grateful. This has allowed me to hire a research assistant, Bernadette Chin, who has done a superb job proofreading and indexing the manuscript, and also to commission illustrations from Jerry Teo with whom it was a pleasure to work. I am grateful to Anatomisches Museum Basel for granting me permission to use an image of one of their wax models and to Nick Hopwood for providing me with a high-quality reproduction of that image.

I hope that this book will prove useful to various audiences. While its main target audience is professional philosophers of science, it should also be accessible enough for classroom use at the graduate and advanced undergraduate levels. Working scientists, too, I hope, will find fresh insights in the following five chapters; while this book will not teach them how to construct models for specific scientific problems, it may alert them to some of the broader desiderata of model building as a scientific practice. So, with a bit of luck, readers may not only learn about how science is done with models, but may also develop an appreciation of why models are essential to good science.

Singapore
October 2015

Axel Gelfert

Contents

Chapter 1
Between Theory and Phenomena: What are Scientific Models?

1.1 Introduction

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 (a 200-acre hydraulic model of the waterways in the entire Mississippi River basin!), and general circulation models (GCM) which allow scientists to run simulations of the 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. Indeed, the deployment of models has not only become central to the scientific enterprise at large, but also to the very image scientists have of themselves. As John von Neumann put it, with some hyperbole: 'The sciences do not try to explain, they hardly even try to interpret, they mainly make models' [1, p. 492]. Whatever shape and form the scientific enterprise might have taken in the absence of models, given their de facto pervasiveness across many disciplines and subdisciplines, it seems safe to say that science without models would not look anything like science as we presently know it.

Philosophical discussions of scientific models likewise distinguish between a bewildering array of different kinds of models. The *Stanford Encyclopedia of Philosophy* gives the following list of model-types that have been discussed by philosophers of science: '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' [2]. As early as 1968, the proliferation of models and model-types, in the sciences as well as in the philosophical literature,

A. Gelfert, *How to Do Science with Models*, SpringerBriefs in Philosophy,
DOI 10.1007/978-3-319-27954-1_1

led Nelson Goodman to lament in his book *Languages of Art:* 'Few terms are used in popular and scientific discourse more promiscuously than "model"' [3, p. 171]. If this was true of science and popular discourse in the late 1960s, it is all the more true of twenty-first century philosophy of science.

As an example of a mathematical model in physics, 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 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 neighbouring 'elementary magnets' interact in such a way as to favour 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, 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 reproduce—yet they are used as representations either of another organism (for example 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 Rachel Ankeny and Sabina Leonelli put it, in their ideal form 'model organisms are thought to be a relatively simplified form of the class of organism of interest' [4, 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—e.g., 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 of models that occupy the middle ground between theoretical description and experimental system is no less perplexing. 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, Max Black [5] distinguishes between four such types: scale models, analogue models, mathematical models, and theoretical models. The basic idea of scale and analogue 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 analogue model also involves a change of medium (as in the Phillips machine, a once-popular hydraulic model 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. Peter Achinstein [6] 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.

When a scientist constructs a theoretical model, she may help herself to certain established principles of a more fundamental theory to which she is committed. These may then be adapted or modified, notably by introducing various new assumptions specific to the case at hand. Often, an inner structure or mechanism is posited which is thought to explain features of the target system. The great variety of models that may thus be generated makes vivid just how central the use of models is to the scientific enterprise. At the same time, it might make one wonder whether it is at all reasonable to look for a unitary philosophical account of models.

This has led some commentators to abandon the search for an account of the nature of models and further to the conclusion that, as Bernd Mahr puts it, modeling can only 'be understood if one stops looking for an answer to the question *of the nature of the model* and starts asking instead *what justifies conceiving of something as a model*' [7, p. 305]. In the absence of any widely accepted unified account of 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*' [8, p. 194]. Furthermore, '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' [9, p. 120]. 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 Ronald Giere [10] sees it, on the *instantial view*, models instantiate the axioms of a theory, where the latter is understood as being composed of linguistic statements, including mathematical statements and equations. 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 designated parts of the world' [10, p. 56]. 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. 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. Anjan Chakravartty nicely characterizes the informational variety of the representational view as follows: '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' [11, p. 198]. 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. 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.

These examples and classifications are necessarily rough sketches, and much of the rest of this book is devoted to giving more depth to our philosophical picture of scientific models. This will involve giving detailed discussions of various cases from across the sciences and exploring their implications for how we should best understand scientific models. The central assumption of this approach is that the key to answering any of the more fundamental questions about scientific models lies in the diversity of their varied uses and functions. While this will require careful attention to the actual uses and applications of scientific models, it would be quite misguided to think that a descriptive approach to scientific modeling could, by itself, tell us what makes something a model, let alone a *good* model. For this, we will need to look beyond the level of case studies and identify possible vantage points from which to judge the success or fruitfulness of a model. A number of philosophical theories have, of course, attempted just that, for example by thinking of models in the same terms as scientific theories. However, one should remain open to the possibility that careful attention to scientific modeling as a practice may itself turn up a 'middle range' of factors which, though strictly speaking not universal, nonetheless help us explain both the success of model-based science and certain recurring patterns of how models are deployed across different disciplines. As I shall argue in later parts of the book, some of the uses and functions of scientific models—e.g., their exploratory role in inquiry—are more akin to certain types of experimentation than to the traditional goals of scientific theories. Furthermore, models often contribute new elements to the theoretical description and empirical investigation of their target systems—elements which are neither part of the fundamental theory nor can be easily 'read off' from the data. Before turning to these questions in more depth, however, it will be instructive to first look more closely at the history of the term 'model' in science, so as to gain a better understanding of what motivates the use of models in scientific inquiry in the first place.

1.2 Models, Analogies, and Metaphor

Given their centrality to contemporary science, it should come as no surprise that scientific models have enjoyed a long and varied history. With our current concepts in hand, it may seem easy to identify past instances of models being employed in science. However, the term 'model' has itself undergone a number of changes throughout the history of science. Indeed, it was not until the nineteenth century that scientists began to engage in systematic self-reflection on the uses and limitations of models. Philosophers 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 onwards did models begin to attract significant philosophical interest in their own right. Yet in both science and philosophy, the term 'model' underwent important transformations. In this section, one such transformation—from a narrow focus on mechanical models to our much broader contemporary understanding of the term 'scientific model'—will be traced.

Take, for example, Pierre 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:

> 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. [12, p. 7]

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, i.e. *A* is related to *B* like *C* is related to *D*' [13, p. 110]. An analogy may be considered merely *formal*, when only the relations (but not the relata) resemble one 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:

> 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. [12, p. 97]

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 behaviour of light.

There is, however, a risk of ambiguity in that last sentence—an ambiguity which, as D.H. Mellor [14, 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 behaviour 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"' [15, 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 'analogue' *of the theory*. Furthermore, we need to distinguish between the analogue (e.g., the theory of sound waves, in Huygens's case) and the set of entities *of which the analogue is true* (e.g., the sound waves themselves). (On this point, see [14, p. 283].) What Duhem resents about the naive 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 Robert Campbell (1880–1949). Thus, Mary Hesse, in her seminal 1963 essay *Models and Analogies in Science* [16], 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 analogues 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' [16, 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 favour 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 part in science. This apparent tension can be resolved by being more precise about the target of his 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' [14, 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 overreliance 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' [17, p. 154]. Campbell is clearly more sympathetic than Duhem towards our 'craving for mechanical theories', which he takes to be firmly rooted in our psychology. But he insists that '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 mould' [17, p. 156]. 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' [17, 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 stance that makes room for scientific models as a way of confirming and modifying a scientific theory; see [14, pp. 286–287].)

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 [16]. 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 colour—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: '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 behaviour of gases' [16, p. 10].

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 analogy and the neutral analogy—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 analogue. (On Black's and Achinstein's terminology, mentioned in Sect. 1.1 above, 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 analogue 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: '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' [16, pp. 12–13].

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' [13, 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 Janet Soskice and Rom Harré note, there is a special affinity between models and metaphor:

> 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. [18, p. 302]

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 deal 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.

1.3 The Syntactic View of Theories

The syntactic view of theories originated from combining the fundamental tenets of two research programmes: the philosophical programme, going back to Pierre Duhem (1861–1961) and Henri Poincaré (1854–1912), of treating (physical) theories as systems of hypotheses designed to 'save the phenomena', and the mathematical programme, pioneered by David Hilbert (1862–1943), which sought to formalize (mathematical) theories as axiomatic systems. By linking 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.

The label 'syntactic' derives from the distinction, typically drawn in the study of formal languages and their interpretations, between the syntax and the semantics of a language *L*. 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 (for example via rules of composition) and can be assessed in terms of their truth or falsity in *R*. Though this distinction is sharpest in logic, the restriction to formal languages is often dropped, such that it provides a framework for thinking also about scientific theories (which are often formulated in ways that are closer to natural language than to logic). The contrast between the syntax and the semantics of a language allows for two different perspectives on 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 (that is, 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. The syntactic view adopts the former perspective and seeks to fashion scientific theories in the image of fully axiomatized systems of statements, perhaps the best example of which would be Euclidean geometry, which consists of five axioms and all the theorems derivable from them using geometrical rules.

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 the meaning of such higher-level terms of a theory could be explicated in observational sentences. In order for this approach to work, a clear distinction between theoretical and observational terms needs to be maintained. As the growing recognition of the theory-ladenness of observation began to call this distinction into question, so correspondence rules began to lose much of their appeal.

As an illustration of what the syntactic view looks like in practice, 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 (1891–1970) called 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, i.e. 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, '[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' [19, p. 57]. 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' [20, p. 125].

For proponents of the syntactic view of theories, models played a marginal or at best auxiliary role. Carnap famously urged his readers 'to realize that the discovery of a model has no more than an aesthetic or didactic or at best a heuristic value, but is not at all essential for a successful application of the physical theory' [21, p. 210]. Others were more forgiving, but the role they attributed to models looks rather unlike how models are, in fact, employed in science. For example, Richard Braithwaite (1900–1990) made room for models as a way of—'hypothetically', as it were—addressing an epistemological challenge we face when we confront scientific theories. As he sees it, theoretical principles, though meant to explain observable facts and in this sense '*logically prior* to the lower-level hypotheses', are '*epistemologically posterior*' to them in the sense that the meaning of theoretical terms (and, by extension, the development of theory) is itself dependent on empirical findings [22, p. 89]. Models are a way of bringing these opposite movements into alignment, if only hypothetically: in a model, 'the logically prior premisses determine the meaning of the terms occurring in the representation in the calculus of the conclusion' [22, p. 90]. While Braithwaite acknowledges that this is 'frequently the most convenient way of thinking about the structure of the theory' [22, p. 91], he emphasizes that models lend themselves to abuse. In particular, he identifies two dangers: the hasty identification of a theory with a model for it, and the projection of characteristics of the model—notably, the logical necessity of some of its features—onto the theory. Though more accommodating than other syntactic theorists, Braithwaite's discussion still ends with a stern warning: 'The price of the employment of models is eternal vigilance' [22, p. 93].

A further criticism, which is directed at both the syntactic and the semantic view (see next section), argues that underlying both views is a misguided general picture of how science works. As Nancy Cartwright has pointedly argued, there is a shared —mistaken—assumption that theories are a bit like vending machines: '[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: firstly, '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' [23, p. 247]. Cartwright rejects the assumed automatism implicit in the 'vending machine view', which she sees as fueling false hopes of a shortcut from evidence to theory, as though an assessment of the evidential significance of observations and their relation to our hypothesized models and theories could ever be divorced from the specific empirical circumstances of the case at hand. By contrast, real science, including model construction, 'is an incredibly difficult and creative activity'—in need of a 'much more textured, and [...] much more laborious, account' [23, p. 247/248] than the 'vending machine view'. Even if this stark contrast may seem a little exaggerated, 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 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 (non-linguistic) 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—centre-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 Frederick Suppe as follows:

> 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. [24, pp. 82–83]

According to a much-quoted remark by one of the main early proponents of the semantic view, Patrick 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' [25, 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 Bas van Fraassen, '[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' [26, p. 64]. 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 constraints (including laws)' [27, p. 154]—that is, they engage in the construction of theoretical models.

Unlike the syntactic view, then, the semantic view appears to devote more attention to the defining features of models in general. 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 Martin Thomson-Jones [28] 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 section follows [28, pp. 530–531].) Suppes [29] begins with the following definition (in slightly modified form):

Definition 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

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

DYNAMICAL AXIOMS

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

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

For p and q in P and t in T,

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

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' [28, 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).

What this suggests is that the semantic view's project of identifying scientific models with truth-making structures in the model-theoretic sense may, at least for the sciences, be an unfulfilled promise. While it may be possible, in principle, to proceed in this way, if applying the semantic approach to specific cases turns out to be rather unwieldy, it may be preferable to adopt a more pragmatic stance—or so the criticism goes. Perhaps, then, we should settle for a less ambitious, though nonetheless informative, definition—for example by thinking of a model as 'a mathematical structure used to represent a (type of) system under study' [28, p. 525]. Characterizing models in this way draws attention to their formats and uses, and also coheres well with analyses that are closer to scientific practice. For example, as we shall see in Chaps. 4 and 5, mathematical models may often be thought of as the output of a 'mature mathematical formalism' [30], and may be used for explanatory and predictive as well as exploratory uses.

1.5 'Folk Ontology' and Models as Fictions

What one finds across a range of scientific disciplines is the practice of taking models as 'stand-ins' for systems that are not, in fact, instantiated. As Peter Godfrey-Smith puts it, 'modelers often *take* themselves to be describing imaginary biological populations, imaginary neural networks, or imaginary economies' [31, 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'. The corresponding research practice of describing and characterizing model systems *as though* they were real instantiated systems (even though they are not) is sometimes referred to as the 'face-value practice' of scientific modeling [32, pp. 285–286]. Not surprisingly, this face-value practice has shaped a fairly widespread view of *what models are*:

> […] 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. [31, p. 735]

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 difficult 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' (ibid.)—such as Sherlock Holmes, J.R.R. Tolkien's Middle-earth etc. 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.

Exactly what it means to say that a scientific model is fictional 'in just the same way' as a literary character, is a thorny question. As early as 1965, Marx Wartofsky noted that, of the various frameworks that scientists help themselves to, scientific models bear an especially close affinity to fictions:

> But models, like characters in fiction, are not transparent as appearances are; they are deliberate constructions, artifacts, and however much we consider them 'ways' in which something is understood to exist, as 'crutches for the imagination', as 'computation devices' or 'inferences machines', they are themselves intermediate entitities. [33, p. 26]

Cartwright echoes this when she writes that a 'model is a work of fiction', in that it combines, in hybrid fashion, 'genuine properties of the object modelled' and mere 'properties of convenience' [34, p. 153]. More recently, Mauricio Suárez has argued that what sets apart *scientific* fictions from their (literary and non-literary) counterparts is their superior 'expediency in inference' [35, p. 169].[1] According to another recent fictionalist proposal, model systems are to be regarded as 'imagined physical systems, i.e. as hypothetical entities that, as a matter of fact, do not exist spatio-temporally but are nevertheless not purely mathematical or structural in that they would be physical things if they were real' [37, p. 253]. Plausible though this may sound, the devil lies 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

[1]Suárez's edited volume *Fictions in Science* [36] has recently sparked renewed interest in fictionalism about scientific models in particular.

cognitive sciences, such as economics and psychology, it is best to drop this restriction to physical systems. 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 Arthur Conan Doyle's *The Adventure of the Three Garridebs* (1924): '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 non-smoker', 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 anti-realist—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 Alexius Meinong (1904), 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, anti-realists 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 Kendall 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' [38, p. 69]. In games of make-believe, participants engage in behaviour 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’ [39, p. 38]—in other words, even though the model itself is only a metre long, it *represents* the bridge *as* a thousand metres 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 metres 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 ‘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’ [40, pp. 308, fn. 14]. Recently, more thoroughgoing *direct* views of models as fictions have been put forward, including by Roman Frigg and Adam Toon. As an illustration of what such a direct view amounts to, Toon considers the following sentence from H.G. 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’ [41, p. 229]. As Toon argues:

> 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. [40, p. 307]

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. (See [39, 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' [42, p. 745]. In response, proponents of the direct view have disputed the need 'to regard theoretical principles formulated in modelling as genuine statements'; instead, as Toon puts it, 'they are prescriptions to imagine' [39, p. 44]. One attraction of the direct approach, then, is its parsimonious metaphysics. As Frigg asserts, rather bluntly: 'What metaphysical commitments do we incur by understanding models in this way? The answer is: none' [37, p. 264].

One potential criticism of the models as fictions view derives from the worry that, by focusing on the user's imaginings, what a model is becomes an entirely subjective matter: if a model is merely a place-holder for whatever is needed to sustain the activity of imagining on the part of an agent, how can one be certain that the same kinds of props reliably give rise to the same kinds of mental imaginings? Yet, defenders of the models as fictions view can respond to this criticism as follows: 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' [37, p. 264]—though, it should be added, not necessarily the same *mental images*.

1.6 The Challenge from Scientific Practice

From the 1960s onwards, in a 'historical turn' that is often attributed to the publication, in 1962, of Thomas Kuhn's *The Structure of Scientific Revolutions* [43], philosophers of science have increasingly shifted attention from questions of how science can be formalized using logic and mathematics to questions of scientific practice. Unsurprisingly, this move has also affected their view of scientific models. While the semantic view, discussed above in Sect. 1.4, was initially part of a broader philosophical project that aimed at making sense of how it is that we can interpret representations as being *about* their targets, over time this gave way to a narrower focus on the ways in which scientists use models in inquiry. What helped this gradual transition was the fact that, from early on, the semantic view was perceived to be better able than its precursors 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.[2] Not all critics were outright opponents of the semantic view: thus, Stephen

[2]For a discussion of inconsistent modeling assumptions, not only in the application but also in the construction of models, see [44].

Downes argued that, as a result of the aforementioned ‘practice turn’ among semantic theorists, the term ‘model’ had simply become too vague, misleading philosophers into erroneously thinking that the logical sense of ‘model’ could also do the job of unifying our understanding of scientific practice. Instead, Downes argued, what was needed was a ‘deflationary view’, according to which philosophers of science should ‘form a loose confederacy for studying scientific theorizing, gathered around the common insight that model-building is one of the most important components of [scientific] theorizing’ [45, p. 151]. This challenge from the reality of scientific practice has given rise to two rather distinct responses. The first response consists in *modifying* the semantic view and finding ways of retaining its overall outlook while making adjustments that are able to accommodate those aspects of scientific practice that do not fit with the original view. An alternative way of responding consists in abandoning the attempt to provide a unitary theoretical account of scientific models and acknowledging, at a more fundamental level, the radical heterogeneity of what, in scientific practice, is considered a ‘scientific model’. In the remainder of this section, we will discuss prominent examples of each of these two responses.

As an example of an attempt to reconcile the spirit of the semantic view with scientific practice, let us consider what has been 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 [46], and references therein.) Like proponents 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 $\mathcal{A} = \langle D, R_i \rangle_{i \in I}$ where D is a non-empty 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’ [46, 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 imposed on models, 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’ [47, 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' (ibid.). Others have taken issue with the general idea that preserving the spirit of the semantic view is a worthwhile enterprise at all, and with the partial structures approach in particular. One frequent criticism directed at the latter 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? Christopher Pincock puts this nicely when he voices his worry that such a view is in 'danger of trivializing our representational relationships' [48, 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: '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' [49, p. 72]. The debate about whether the semantic view, or one of its immediate descendants, is able to account for the variety of models and their uses remains a lively one, with defenders of the semantic view arguing that it has the resources to account for, amongst others, iconic and material models, which were thought by its critics to pose insurmountable difficulties to the semantic view.[3]

As an alternative response to the challenge from scientific practice, a number of authors have proposed accounts that give pride of place to the model user's cognitive interests, intentions, and beliefs. Thus understood, models are no longer treated merely as abstract structures that stand in a relation of isomorphism, or partial isomorphism, to a target system, but as tools of inquiry *for a model user*. Attention is shifted from two-place relations that might obtain between a model and its target to three-place relations involving model, target, and user. Further contextual factors, ranging from facts about the user (e.g., her goals and interests) to the context of use (e.g., in instruction), may then be added. For example, Uskali Mäki proposes that an object M counts as a model if and only if an '[a]gent A uses object M as a representative of some target system R for purpose P, addressing audience E, prompting genuine issues of resemblance to arise, and applies commentary C to identify and align these components' [51, p. 32]. As this example indicates, the turn towards the pragmatics of modeling goes hand in hand with a focus on the variety of uses and functions of models. Indeed, it is not only the *function* of models—e.g., their capacity to represent target systems—which is seen as dependent on the beliefs, intentions, and cognitive interests of their users, but also their very nature: *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' [49, 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;

[3]For one such defence, see [50].

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 modellers piece together models from a variety of ingredients.

A good example of this second type of approach, which places scientific practice at the heart of its analysis of models, is the 'model-as-mediators' view (to be discussed in greater detail in Sect. 5.1). 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 Margaret Morrison and Mary 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' [52, 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)' [52, p. 23]. 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' [52, p. 16].

Other authors have taken up the idea of models as heterogeneous entities, but have emphasized that, over and above acknowledging that models are partially autonomous from theory and data, what is needed is an account of *how* models come to enjoy such partial autonomy. As Tarja Knuuttila argues, *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' [53, p. 48]. Materiality is also seen as explaining various of the 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 a variety of formats and representations, in a way that *enables* certain specific uses, but at the same time *constrains* them. On this view, models are seen as *epistemic tools*: 'concrete artefacts, 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' [54, p. 267]. This hints at an interesting link between the philosophical debate about models and questions in the philosophy of technology, for example the question of how artefacts function and how we, as users of such artefacts, engage with them at a phenomenological level. If it is indeed the case that specific encounters with models always require some concrete format or representation—be it a set of diagrams scribbled on a piece of paper, or an elaborate three-dimensional model that mimics the 'look and feel' of a target system—what we can learn from a model will fundamentally depend on how we encounter the world through it or in it.[4]

References

1. J. von Neumann, Method in the physical sciences, in *Collected Works. Theory of Games, Astrophysics, Hydrodynamics and Meteorology*, vol. VI, ed. by A.H. Taub (Pergamon Press, Oxford, 1961), pp. 491–498
2. R. Frigg, Models in science, Stanford encyclopedia of philosophy (2012). plato.stanford.edu/entries/models-science/. Accessed 10 Feb 2015
3. N. Goodman, *Languages of Art* (Bobbs-Merrill, Indianapolis, 1968)
4. R. Ankeny, S. Leonelli, What's so special about model organisms? Stud. Hist. Philos. Sci. **42** (2), 313–323 (2011)
5. M. Black, *Models and Metaphors: Studies in Language and Philosophy* (Cornell University Press, Ithaca, 1962)
6. P. Achinstein, *Concepts of Science: A Philosophical Analysis* (Johns Hopkins Press, Baltimore, 1968)
7. B. Mahr, On the Epistemology of Models, in *Rethinking Epistemology*, vol. 1, ed. by G. Abel, J. Conant (de Gruyter, Berlin, 2012), pp. 301–352
8. G. Contessa, Editorial introduction to special issue. Synthese **172**(2), 193–195 (2010)
9. S. Ducheyne, Towards an Ontology of Scientific Models. Metaphysica **9**(1), 119–127 (2008)
10. R. Giere, Using Models to Represent Reality, in *Model-based Reasoning in Scientific Discovery*, ed. by L. Magnani, N. Nersessian, P. Thagard (Plenum Publishers, New York, 1999), pp. 41–57
11. A. Chakravartty, Informational versus functional theories of scientific representation. Synthese **172**(2), 197–213 (2010)
12. P. Duhem, *The Aim and Structure of Physical Theory*. Transl. P.P. Wiener (Princeton University Press, Princeton, 1914/1954)
13. 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
14. D.H. Mellor, Models and analogies in science: Duhem versus Campbell? Isis **59**(3), 282–290 (1968)
15. D. Bailer-Jones, *Scientific Models in Philosophy of Science* (University of Pittsburgh Press, Pittsburgh, 2009)
16. M. Hesse, *Models and Analogies in Science* (Sheed and Ward, London, 1963)
17. N.R. Campbell, *Physics: The Elements* (Cambridge University Press, Cambridge, 1920/2013)
18. 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
19. R. Carnap, *Foundations of Logic and Mathematics* (The University of Chicago Press, Chicago, 1939)

[4]In Sect. 5.5, I shall discuss in detail how some scientific models enable us to gain knowledge by functioning as mediators between different types of user–model–target relations, specifically between 'embodied' ways of relating to the world and those that require more specialized interpretive activities (such as 'reading' an instrument or manipulating a set of mathematical equations).

20. R.F. 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. Lastowski, K. Paprzycka, P. Przybyzs (Rodopi, Amsterdam, 2007), pp. 123–158
21. R. Carnap, Foundations of Logic and Mathematics, in *Foundations of the Unity of Science*, vol. 1, ed. by O. Neurath, R. Carnap, C. Morris (The University of Chicago Press, Chicago, 1969), pp. 139–214
22. R.B. Braithwaite, *Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science* (Cambridge University Press, Cambridge, 1968)
23. N. Cartwright, Models and the Limits of Theory: Quantum Hamiltonians and the BCS Model of Superconductivity, in *Models as Mediators: Perspectives on Natural and Social Science*, ed. by M.S. Morgan, M. Morrison (Cambridge University Press, Cambridge, 1999), pp. 241–281
24. F. Suppe, *The Semantic Conception of Theories and Scientific Realism* (University of Illinois Press, Urbana, 1989)
25. P. Suppes, A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences. Synthese **12**(2–3), 287–301 (1960)
26. B. van Fraassen, *The Scientific Image* (Oxford University Press, Oxford, 1980)
27. C. Liu, Models and theories I: the semantic view revisited. Int. Stud. Philos. Sci. **11**(2), 147–164 (1997)
28. M. Thomson-Jones, Models and the semantic view. Philos. Sci. **73**(5), 524–535 (2006)
29. P. Suppes, *Introduction to Logic* (Van Nostrand, Princeton, 1957)
30. A. Gelfert, Mathematical formalisms in scientific practice: from denotation to model-based representation. Stud. Hist. Philos. Sci. **42**(2), 272–286 (2011)
31. P. Godfrey-Smith, The strategy of model-based science. Biol. Philos. **21**(5), 725–740 (2006)
32. M. Thomson-Jones, Missing systems and the face value practice. Synthese **172**(2), 283–299 (2010)
33. M. Wartofsky, Models, Metaphysics and the Vagaries of Empiricism, in *Models: Representation and the Scientific Understanding* (Reidel, Dordrecht, 1979), pp. 24–39
34. N. Cartwright, *How the Laws of Physics Lie* (Oxford University Press, Oxford, 1983)
35. M. Suárez, Scientific Fictions as Rules of Inference, in *Fictions in Science: Philosophical Essays on Modeling and Idealization*, ed. by M. Suárez (Routledge, London, 2009), pp. 158–178
36. M. Suárez, *Fictions in Science: Philosophical Essays on Modeling and Idealization*, ed. by M. Suárez (Routledge, London, 2009)
37. R. Frigg, Models and fiction. Synthese **172**(2), 251–268 (2010)
38. K. Walton, *Mimesis as Make-Believe: On the Foundations of the Representational Arts* (Harvard University Press, Cambridge, Mass., 1990)
39. A. Toon, *Models as Make-Believe: Imagination, Fiction, and Scientific Representation* (Palgrave-Macmillan, Basingstoke, 2012)
40. A. Toon, The ontology of theoretical modelling: models as make-believe. Synthese **172**(2), 301–315 (2010)
41. H.G. Wells, *War of the Worlds* (Penguin, London, 1897/1978)
42. R. Giere, How models are used to represent reality. Philos. Sci. **71**(5), S742–S752 (2004)
43. T.S. Kuhn, *The Structure of Scientific Revolutions* (The University of Chicago Press, Chicago, 1962)
44. M. Frisch, Models and scientific representations or: who is afraid of inconsistency? Synthese **191**(13), 3027–3040 (2014)
45. S.M. Downes, *The Importance of Models in Theorizing: A Deflationary Semantic View*. Proceedings of the PSA1992, vol. 1, Chicago, 1992
46. N. da Costa, S. French, *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning* (Oxford University Press, New York, 2003)
47. 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

48. C. Pincock, Overextending partial structures: idealization and abstraction. Philos. Sci. **72**(5), 1248–1259 (2005)
49. M. Suárez, N. Cartwright, Theories: tools versus models. Stud. Hist. Philos. Mod. Phys. **39**(1), 62–81 (2008)
50. S. French, J. Ladyman, Reinflating the semantic approach. Int. Stud. Philos. Sci. **13**(2), 103–121 (1999)
51. U. Mäki, MISSing the world. Models as isolations and credible surrogate systems. Erkenntnis **70**(1), 29–43 (2009)
52. M. Morrison, M. Morgan, Models as Mediating Instruments, in *Models as Mediators: Perspectives on Natural and Social Science*, ed. by M.S. Morgan, M. Morrison (Cambridge University Press, Cambridge, 1999), pp. 10–37
53. T. Knuuttila, *Models as Epistemic Artefacts: Toward a Non-Representationalist Account of Scientific Representation* (University of Helsinki, Helsinki, 2005)
54. T. Knuuttila, Modelling and representing: an artefactual approach to model-based representation. Stud. Hist. Philos. Sci. **42**(2), 262–271 (2011)

Chapter 2
Scientific Representation and the Uses of Scientific Models

2.1 Models and Their Functions

The great variety of models in scientific practice, which is reflected by the somewhat sprawling taxonomies of model-types we encountered in Chap. 1, has led some philosophers to propose quietism as the only viable attitude towards the ontological question of *what a model is*. As Steven French puts it, 'whereas positing the reality of quarks or genes may contribute to the explanation of certain features of the physical world, adopting a similar approach towards theories and models—i.e., reifying them as entities for which a single unificatory account can be given—does nothing to explain the features of scientific practice' [1, p. 245]. (Quietism is usually a position of last resort in philosophy and, perhaps not surprisingly, French's professed quietism has not stopped him and others from developing an elaborate, if controversial, unified framework for thinking about models and theories.) In the present chapter, we will instead follow the alternative suggestion to treat models as 'functional entities' [2, p. 120]; on this view, the various functions of models in scientific inquiry are our best—and perhaps only—guide when it comes to finding answers to any of the more fundamental questions about scientific models, including those about their ontology, epistemic status, confirmation, and so forth.

In exploring the different functions of scientific models, it will be useful to keep in mind some of the distinctions drawn in the previous chapter (Sect. 1.1), notably between *instantial* and *representational* views of models, and among the latter, between *informational* and *pragmatic* versions of the view. The instantial view regards models as instantiations of the axioms of a theory and considers the relationship between models and theories to be the primary locus of philosophical significance. By contrast, the representational view takes models to be a way of accessing the world, thereby shifting attention to the way models represent the world. Various locutions are typically employed in order to characterize what is involved in treating a model as a representation: thus, a model may be said to 'stand

A. Gelfert, *How to Do Science with Models*, SpringerBriefs in Philosophy,
DOI 10.1007/978-3-319-27954-1_2

in for' its target system, or its relation to the target system may be described as analogous to that between a map and the territory it shows. In its most general form, the representational view thus considers models to be 'tools for *representing the world*' [3, p. 44]. '[T]he crux of the problem of representation', as Margaret Morrison puts it, then becomes the following question: 'in virtue of *what* do models represent and how do we identify what constitutes a correct representation?' [4, p. 70] Informational views take representation to be an objective relation between the model and its target, which imbues the former with information about the latter, irrespective of a model user's beliefs or intentions, and regardless of the cognitive uses to which he or she might put the model. This contrasts with more pragmatic versions of the representational view, according to which one cannot 'reduce the essentially intentional judgments of representation-users to facts about the source and target object or systems and their properties' [5, p. 768]. Similarly, in his defence of similarity as the basis of model-based representation, Ronald Giere, although committed to the idea that 'what is said to be similar to what, in what ways, and to what degrees' can always be specified, insists that there are 'many possible specifications depending on the particular interests of those doing the modeling' [3, p. 46]. Model-based representation, thus understood, is essentially a three-place relation between the model, its target, and the model user. The value and function of models then derives from their role in inquiry, i.e. from the way they enable users to draw inferences on their basis about the target system, make predictions, or facilitate other courses of action. In its widest sense, the term 'model' may then be understood, as Günter Abel puts it, 'as a reconstruction of central features of a concrete object, process, or system, which itself becomes a matter of further investigation' [6, p. 33].

Before discussing two specific proposals of how model-based representation comes about—R.I.G. Hughes's DDI account (Sect. 2.3) and Mauricio Suárez's inferential account (Sect. 2.4)—it will be instructive to consider, in the next section (Sect. 2.2), whether there is anything distinctive about *scientific* representation in particular, or whether representation in science is of a piece with representation more generally. The final section of this chapter (Sect. 2.5) will comment on the issue of realism and anti-realism in the context of model-based representation and will broaden the perspective further to also consider non-representational uses of models in science.

2.2 Scientific Representation

Human beings, through biological and cultural evolution, have developed elaborate ways of representing the world around them: via the mental representations that feature in cognition, through language and its ever-expanding vocabulary, and through the deliberate creation of artefacts in art, technology, and science. 'Representation', conceived of as an umbrella term for these various activities and relations, may seem like a nebulous philosophical concept, which might explain why many philosophers of representation restrict their theories to rather more specific domains. Thus, Bas van Fraassen states rather categorically that he 'will have no truck with mental representation, in any sense', claiming that the way philosophers of mind have historically discussed the notion 'has nothing to contribute to our understanding of scientific representation' [7, p. 2]. With respect to scientific representation, van Fraassen's own position has changed considerably over the years. Whereas in *The Scientific Image* [8] van Fraassen argued that '[t]o present a theory is to specify a family of structures, its *models*', and that a theory 'is empirically adequate if it has some model such that all appearances [as described in experimental and measurement reports] are isomorphic to empirical substructures of that model' [8, p. 64], almost thirty years later his position has moved to a much more pragmatic account of scientific representation: '*There is no representation except in the sense that some things are used, made, or taken, to represent some things as thus or so.*' [7, p. 23; italics original] Hence, the notion of 'use' and pragmatic considerations are to be 'give[n] pride of place in the understanding of scientific representation' [7, p. 25], even if a unified 'theory of representation' may turn out to be elusive. In a similar spirit, Morrison argues that even '[w]ithout articulating a specific theory of representation we can nevertheless appeal to some general ideas about what it means for models to represent phenomena or systems' [9, p. 125]. This deflationary approach seems exactly right to me: even in the absence of a widely accepted *general* theory of representation, it seems perfectly possible to characterize *scientific* representation in productive and insightful ways —so long as one's account does not hinge on overly controversial assumptions regarding the general problem.

In his seminal book *Languages of Art* [10], Nelson Goodman aimed to bring together previously separate debates about the nature of representation in language, art, and science, as indicated by the subtitle of the book which promises 'an approach to a theory of symbols'. Traditionally, philosophers have distinguished between *natural* and *non-natural* signs. Whereas non-natural signs acquire their meaning by way of—e.g., linguistic—convention, natural signs stand in certain non-arbitrary (typically: causal) relations that connect them to their target. Thus, the word 'fire' refers non-naturally, simply because as competent speakers of English we take it to be a symbol of fire, whereas smoke is a natural sign of fire since the latter is usually the cause of the former. One might ask where along the natural/non-natural spectrum we should locate representation. Surely it would be too strong to demand that a representation must be causally dependent on its target; indeed, in the example just given, one might find it more appropriate to consider smoke *evidence* of fire (rather than a *representation* of it). Goodman introduces the notion of 'denotation' in order to capture an important ingredient in any representational relationship, namely the fact that a representation 'stands in for' its target (where such a target, in fact, exists).[1] Denotation, thus understood, frees the

[1]Whereas in *Languages of Art* (1968/1976), Goodman offers no definition of the term 'denotation', in *Of Mind and Other Matters* (1984), he writes: 'This common relationship of applying to or

representational relationship from the constraints of causality or resemblance, in that it may be entirely stipulative: 'almost anything may stand for almost anything else.' [10, p. 5]

From this it does not follow that any attempted act of denotation will automatically succeed at representing a target. For one, in order for there to be an instance of denotation, what is being denoted must exist; if the purported target does not exist—for example, because we are dealing with non-existent entities such as unicorns or the ether—denotation must necessarily fail. Goodman hints at workarounds for this problem, by distinguishing between a) what a representation *denotes*, and b) what *kind* of representation it is—so that, say, a certain picture could be a unicorn-representation (i.e. would belong to the class of unicorn-images) without thereby denoting a unicorn (since unicorns do not exist): 'A picture that represents a man denotes him; a picture that represents a fictional man is a man-picture; and a picture that represents a man as a man is a man-picture denoting him' [10, pp. 27–28]. Furthermore, successful representation of an existing target requires more than mere denotation. While Goodman, in various places, gives the impression that anything may represent anything else, he also acknowledges that *successful* representation is subject to de facto constraints. Even before issues of faithfulness, accuracy, and truth arise, there is the question of whether a given representation makes relevant information salient and whether it can draw on entrenched denotative practices: 'Representation […] is apt, effective, illuminating, subtle, intriguing, to the extent that' its originator 'grasps fresh and significant relationships and devises means for making them manifest' [10, pp. 32–33].

Much of Goodman's philosophy is an attempt to negotiate the tension between, on the one hand, the arbitrariness that seems to follow from the radical contingency of thought and action and, on the other hand, the apparent stability of our conceptual frameworks and practices. Hence, while 'there are countless alternative systems of representation and description', these are themselves 'the products of stipulation and habituation in varying proportions' [10, p. 40]. To be sure, we often find that we inherit established notational systems and representational formalisms, which—though in principle arbitrary—for this very reason are no longer 'up to us'. Although we have some degree of choice among such systems, given a particular system, 'the question whether a newly encountered object is a desk or a unicorn-picture […] is a question of the propriety, under that system, of projecting the predicate "desk" or the predicate "unicorn-picture" […], and the decision both is guided by and guides usage for that system' [10, pp. 40–41]. Yet critics of Goodman have found this appeal to entrenchment and past usage a little too casual: if the only resistance we face when introducing new concepts and predicates is in terms of past usage that needs to be overcome, this would seem to leave little room

(Footnote 1 continued)

standing for, I call *denotation*—not to preclude but rather to introduce examination of various types of denotation in different symbol systems and also the relationships between denotation and other types of reference.' [11, p. 80].

for any independent contribution from the world 'out there'. As Joseph Margolis remarks, somewhat pointedly, Goodman has 'no theory of the actual behavior of scientific thinking', which makes his account of entrenchment, along with the goal of '*ultimate* acceptability' (which Goodman considers a legitimate substitute for truth, [11, p. 38]), ultimately untenable: 'under historicized and radically relativized circumstances, the governing notion of "*ultimate* acceptability" and its regularized bearing on these other distinctions are rendered completely meaningless or inoperable' [12, p. 121]. Whatever the merits of these more specific criticisms of Goodman's project, what matters for our purposes is the general realization that representation arises from the interplay of denotation (which affords considerable, though not unlimited latitude) and the various factors that determine whether one thing can successfully 'stand in for' another (i.e., make relevant relationships in the target system manifest to its user).

Similarity, Goodman argues, fails as a criterion of successful representation: though it may play an auxiliary role in certain contexts, it contributes nothing essential to the representational relationship.[2] Considering different versions of this—as he puts it: 'most naive'—view of representation ('*A* represents *B* if and only if *A* appreciably resembles *B*', or '*A* represents *B* to the extent that *A* resembles *B*'), Goodman claims, with some hyperbole, that 'more error could hardly be compressed into so short a formula' [10, pp. 3–4]. Why does Goodman think resemblance fails as a basis of representation? For one, resemblance is a reflexive and symmetric relation, whereas representation is neither. Nothing resembles a portrait of the Duke of Wellington more than the painting itself, but that does not mean that the portrait represents itself. Furthermore, the painting resembles the Duke of Wellington to exactly the same degree as the Duke resembles the painting, but it does not follow that the Duke represents the painting. And in any case, the painting arguably is more similar to other two-dimensional paintings than it is to the three-dimensional, flesh-and-blood Duke or his identical twin brother—yet, this neither prevents the painting from representing the Duke, nor renders the twins representations of one another. We must therefore already have resolved to treat one thing as a representation of the other (and *not* vice versa!) before questions of faithfulness or accuracy can be raised: this is precisely the function of denotation. Importantly, any account of representation should also make room for *misrepresentation*. As van Fraassen reminds us: '*Misrepresentation is a species of representation*' [7, p. 14]. Thus, a caricature may, as a political statement, purposely represent Tony Blair as George W. Bush's lapdog and, in doing so, may misrepresent him as considerably smaller in size than his American friend, yet it remains no less a representation of the two men (as opposed to, say, a hypothetical owner/dog pair).

Having sketched some of the complexities and constraints of a general theory of representation, let us focus more narrowly on scientific representation. As we shall see, the very idea that 'scientific representation' merits special treatment has been

[2]For a defence of resemblance as the basis of representation, at least for the case of depiction, see [26].

the subject of much contestation—not least since it has proved notoriously difficult to arrive at any general demarcation criterion that would allow us to tell science from non-science. For the moment, let us assume that we have a good enough grasp of what constitutes a scientific context to be able to recognize certain representational devices—e.g. scientific theories, models, hypotheses, data etc.—as instances of *scientific representation*. Focusing on scientific models as one class of scientific representations, we may then return to Morrison's earlier question: 'in virtue of *what* do models represent and how do we identify what constitutes a correct representation?' [4, p. 70]. As Craig Callender and Jonathan Cohen have noted, this question really addresses two distinct problems: the first part of the question concerns the problem of what *constitutes* the representational relation between a model and the world, whereas the second relates to 'the normative issue of what it is for a representation to be correct' [13, p. 69]. Let us call the first problem the *constitutive question* and the second the *evaluative question*. Callender and Cohen claim that Morrison and other contemporary philosophers of science, in their focus on scientific practice, have tended to run both questions together when, in fact, the two should be contrasted sharply. We will return to Callender's and Cohen's criticism shortly; before doing so, it will be instructive to draw a few more useful distinctions.

In the previous section, we already encountered a broad distinction, within the representational view of models, between *informational* and *pragmatic* approaches. An even more basic distinction derives from opposing stances concerning the prospects of analyzing representation in terms of more basic relations, such as similarity or isomorphism. If one holds that representation—whether in science or in general—can be fully analyzed in terms of such more fundamental relations, one would properly be called a *reductionist* about representation. By contrast, if one believes such a reduction to be impossible and instead holds that representation is a basic relation *sui generis*, one should be deemed a *non-reductionist* about representation.[3] A further distinction may be drawn between *substantive* and *deflationary* accounts of representation, with the latter settling for a broad characterization of the functional point of representation—e.g., the fact that it allows users of representational devices to gain new information about the target—and the former aiming for a deeper, more 'robust' explanation of the functional utility of a representation in terms of an underlying constituent relation between a representation and its target (see [14, p. 94].) Accounts that equate representation with similarity relations between a representational device and its target are a good example of reductionism, as are structuralist accounts that analyze representation purely in terms of relations of (partial) isomorphism between models and their targets. Both types of accounts have been the target of criticism, as exemplified by Goodman's attack on similarity-based accounts and as discussed in the previous chapter in connection

[3]Similar to 'non-reductionism' about the representational relation, Suárez describes as 'primitivism' any position that 'claims that the representational relation, if there is any, may not be further analysed' [14, p. 94].

with structuralist accounts (see Sect. 1.6). In Sects. 2.3 and 2.4 below, we will encounter two examples of non-reductionist accounts which, however, will differ in regard to their place along the substantive/deflationary spectrum.

On the issue of terminology, while it has become customary to refer to that which is being represented as the 'target' (or 'target system'), there is less agreement on what, in general, to call that which does the representing. When dealing exclusively with one type of representation—portraits, say, or mathematical models—this difficulty can be easily avoided: what represents the Duke of Wellington is simply the portrait that depicts him. Speaking in more general terms, it is certainly possible to refer to 'a representation' of a target—as indeed I already have on various occasions. However, this usage runs the risk of eliding the distinction between the general representation relation and specific realizations of it. Some authors prefer to speak of 'sources' and their targets; other locutions include 'representational device' or 'representational vehicle'. It is obvious that the constitutive question pertains to the nature of the representation relation in general, whereas the evaluative question—what it takes for a given realization to be a correct, faithful, or accurate representation of its target—will depend on the specifics of the case at hand. Suárez makes a useful distinction between the *constituents* of representation and its *means*, with the former being implicitly defined by whatever it takes to establish representation for *any* source–target pair and the latter referring to the variable, context-dependent resources a model user draws on when reasoning about a target by way of engaging a model of it [14, p. 93]. It is clear that the means of representation employed by a given representational vehicle will heavily determine its overall effectiveness and 'representational power' [15, p. 294]. As we shall see in Chap. 5, an important function of models is that they allow us to move back and forth between the representational means and aspects of the target system—sometimes effortlessly, but often in a way that requires explicit attention to the format and medium of representation.

What about those who criticize as incoherent the very idea that there is something which sets scientific representation apart from representation-at-large? On this view, there simply is no special problem of 'scientific representation', since representation in science is no different in character from representation in other domains. Such critics, to be sure, can point to various bits of evidence in support of their denial of the coherence of the notion of 'scientific representation'. For one, there is the absence, already mentioned, of a clear-cut (e.g. logical) demarcation criterion between science and non-science. Within science, too, there is considerable disagreement between different disciplines about what constitutes a viable representational target; this is reflected in a mind-boggling diversity and disunity concerning the representational vehicles employed across the various sciences. Last but not least, scientists often differ in their axiological commitments regarding the standards and criteria for what makes something a good representation. Apart from these intra-scientific considerations, there is also the further observation that 'scientists use entities other than models—language, pictures, mental states, and so on—to represent the very same targets that models represent'. This, Callender and Cohen argue, points to model-based (scientific) representation being derivative of representation outside science since 'it would be surprising that scientific, linguistic, pictorial, mental, and other sorts of representations should coincide in their representational targets were they not at all related' [13, p. 71]. More specifically, they propose that all representations, including 'the varied representational vehicles used in scientific settings (models, equations, toothpick constructions, drawings, etc.) represent their targets (the behavior of ideal gases, quantum state evolutions, bridges) by virtue of the mental states of their makers/users' [13, p. 75]. Thus, when a theoretical biologist writes down the Lotka-Volterra equation and stipulates that it should represent the population dynamics of a predator–prey system, he intends that his audience recognize his intention to activate *in them* the belief that the equations should be taken as a stand-in for the real-world system. The representational vehicle —in this case, the set of equations—is merely a useful prop for facilitating conversation about predator–prey systems and for expressing the modeler's beliefs about them.

By linking representation-at-large—including scientific representation—to the expression of intentions on the part of the modeler, Callender and Cohen emphasize the stipulative element in our representational practices. However, they are keen to point out that representation by stipulative fiat alone is not the norm, in science or elsewhere; as already noted by Goodman, our representational devices often depend on entrenched symbolic systems and the utility of our representational vehicles depends on them. Questions of utility, however, are simply 'questions about the pragmatics of things that are representational vehicles, not questions about their representational status *per se*' [13, p. 75], or so Callender and Cohen argue. In other words, 'virtually anything can be stipulated to be a representational vehicle for the representation of virtually anything' [13, p. 74], leaving the evaluative question of the suitability of a given representational vehicle entirely a matter of contingent, context-dependent factors. The basic idea that scientific representation is continuous with representation-at-large and is fully derivative of actions and intentions on the part of the modeler, is not new. Marx Wartofsky, in a paper first published in 1966, makes essentially the same point the other way around, by equating all representation with model-based representation of one sort or another:

> We begin by modelling, therefore, with our first mimetic acts, and with our first use of language. And we continue modelling by way of what, on various grounds, have been distinguished as analogies, models, metaphors, hypotheses and theories. [16, p. 10]

Models, Wartofsky argues, are 'used to communicate an intended factually true description' (ibid.); that is, they serve communicative purposes and depend on us as modelers: 'Our own cognitive activity enters here, to take one as representing the other', subject only to pragmatic 'constraints on what may or may not be made into a model' [16, p. 4]. Regarding the specific constraints that various substantive accounts impose as conditions of scientific representation, Callender and Cohen argue that these, too, are of merely pragmatic significance:

> Likewise, we suggest that, while resemblance, isomorphism, partial isomorphism, and the like are unnecessary for scientific representation, they have important pragmatic roles to

play; namely, they can (but need not) serve as pragmatic aids to communication about one's choice of representational vehicle [13, p. 76].

This attempted dissolution of the problem of scientific representation by reducing it to a matter of stipulation—subject only to pragmatic constraints, in order to facilitate the communication of the modeler's intention to represent a given target—has been criticized for assigning the model user rather too central a role in bringing about successful representation. As Morrison objects: 'There may be no representation without users, but that doesn't mean that users determine what's required for something to represent something else' [9, p. 128].

Callender's and Cohen's approach lends itself to an even more fundamental criticism. While models, along with other scientific representations, often serve the purpose of enabling communal inquiry, by functioning as means for the communication of one party's intentions and mental states to another, their role goes far beyond that of being a mere 'facilitator' of inquiry. Morrison hints at this when she argues that, often, 'scientific representation is about conceptualising something in a way that makes it amenable to a theoretical or mathematical formulation' [9, p. 129], and Callender and Cohen seem to acknowledge as much when they note that sometimes a modeler may include himself in the audience at which the model is aimed [13, p. 77]. What, one might ask, would be the point of directing a model *at oneself*, if a model is nothing but a mere prop for communicating one's beliefs and intentions? The answer, it seems to me, must be that the role of models in inquiry is not exhausted by their functioning as mere props for communicating mental states in the way suggested by Callender and Cohen. Models can surprise us, open up unforeseen lines of inquiry, and lead to novel insights about their targets, all of which suggests that they enjoy considerable autonomy. Mathematical models, in particular, are imbued with a considerable internal structure and dynamics, which renders them partially independent from the intentions of their users. None of this is easily captured by Callender's and Cohen's account, which accords them only an auxiliary role as vehicles of pre-existing intentions and beliefs on the part of their users. Rather than thinking of models as mere *facilitators*, I want to suggest—in a phrase that I will unpack in detail in Chap. 5—that we should think of them as *contributors* to inquiry.

2.3 The DDI Account of Model-Based Representation

R.I.G. Hughes [17] has proposed an account of scientific representation, according to which the representational capacity of theoretical models is due to the interplay between three components: denotation, demonstration, and interpretation. *Denotation*, following Goodman, is conceived of as the basic relation whereby certain elements of a model 'stand for', or are 'a symbol of', elements in the physical world; as such, it accounts not only for the fact that theoretical elements of a model purport to refer to elements in the physical world, but also for the asymmetry that exists between a representational device and its target system. The possibility of *demonstration*—either within a theoretical model, through the application of mathematical derivation techniques, or via experimental intervention in the case of material models—attests to the fact that models possess an internal dynamic and can lead to new results and insights. *Interpretation*, finally, relates what has been demonstrated back to the physical world. Though Hughes is careful to distance himself from the reductionist claim 'that denotation, demonstration, and interpretation constitute a set of acts individually necessary and jointly sufficient for an act of theoretical representation to take place' [18, p. 155], he considers all three components to be involved in scientific representation and takes the interplay between them to be distinctive of the way models represent reality—which is why he takes his *DDI account* to be 'a very general account of theoretical representation' [18, p. 153]. Thus, in the terminology introduced in the previous section, Hughes's account may be deemed a *substantive*, *non-reductionist* account of scientific representation.

Although Hughes's DDI account follows Goodman's advice that 'we must examine the characteristics of representation as a special kind of denotation' [10, p. 5], it does not simply equate denotation and representation; instead, it demands that denotation be put to the test by successful demonstration and interpretation. Consider the case of a mathematical model of a physical phenomenon, e.g. a set of partial differential equations intended to represent the flow of heat in a solid. The theoretical activity of modeling heat flow using the calculus of partial differential equations involves the interplay between what Chris Pincock has called the *physical attitude*—'which insists that throughout we are talking about physical systems and physical magnitudes'—and the *mathematical attitude*, which considers such steps as taking the 'unphysical' limit $\Delta x \rightarrow 0$ (e.g. in order to mathematically define the temperature 'at a given point'—even though temperature, in the physical sense, only applies to spatially extended ensembles of particles) as 'involving only mathematical objects' [19, p. 88]. When we resolve to treat certain variables as denoting physical quantities, we clearly do so by taking a physical attitude towards the model, and we again need to adopt this stance when interpreting results—e.g. concerning the final distribution of temperature—as predictions the model makes about the target system. In between, however—that is, during the phase of *derivation*—we can rely on a host of tried and tested mathematical derivation techniques. While such techniques may have their own practical and conceptual problems, they do not directly touch upon the question of how theoretical models represent a reality external to themselves. Mathematical demonstration, thus, may take place entirely from within the mathematical attitude, yet it is no less essential to the process of modeling as a whole, both insofar as it allows for the derivation of results and predictions from a model, and because it makes salient that a mathematical model, by virtue of its being a mathematical object, has an internal dynamic. For Hughes, this insight is fundamental to the use of models in physics in general:

> To be predictive, a science must provide representations that have a dynamic of this kind built into them. That is one reason why mathematical models are the norm in physics. Their internal dynamic is supplied, at least in part, by the deductive resources of the mathematics they employ [17, p. 332].

Once again, this points to models as being more than a vehicle for a user's intentions or beliefs: by tapping into the rich resources of mathematics, mathematical models are imbued with considerable deductive resources and an internal dynamic that may go well beyond what an individual user may intend or be able to survey.

Interpretation, like the other two components of the DDI account, is an important ingredient in the way theoretical models represent. Without it, demonstrated results would remain merely formal results within a deductive mathematical structure, lacking empirical meaning. What is needed is 'a function that takes us from what we have demonstrated […] back into the world of things' [17, p. S333], and interpretation plays this role. Whereas denotation picks out features in the world, which are then referred to by elements within the model, interpretation projects internally-derived results back onto the world, where they must be assessed in terms of their empirical adequacy. This may require considerable ingenuity and imagination. In the case of a mathematical model, even when a result has been successfully derived within the formalism of the model equations, its empirical interpretation may not always be self-evident. As an example, consider the case of mathematical divergences: if one or more of a model's variables diverge for certain parameter values, the user may be faced with the choice of either dismissing it as an 'unphysical' result—for example because the corresponding physical magnitude is recognised as necessarily finite for any finite physical system under consideration—or interpreting it as an indicator of a real feature in the world (e.g., a phase transition), which the model may simply be unable to capture in its entirety. Neither denotation nor interpretation comes with a guarantee of success, but *when* they succeed—that is, when a model picks out the right features in the world, and interpretation assigns empirically adequate meanings to demonstrated results—denotation and interpretation may indeed be said to be the inverse of each other, and the model as a whole may be deemed a successful representation.

2.4 Representation and Surrogate Reasoning: Suárez's Inferential Account

By relying on denotation for the requisite representational asymmetry between model and target, the DDI account of scientific representation is open to the earlier criticism that, for denotation to be successful, its intended target must exist. Yet scientific models sometimes deals with systems that do not—or perhaps could not—exist, such as higher-dimensional or infinitely extended systems. While there may be workarounds for the problems arising from the non-existence of intended targets, some modifications of the DDI account appear to be necessary. The second key ingredient of the DDI account—*demonstration*—may also be less straightforward than appears at first sight. As Mauricio Suárez has noted, 'for Hughes, representation involves demonstration essentially, and hence requires the actual carrying out of inferences about the target on the part of some agent' [5, p. 770]: that is, it requires the actual performance of *demonstrating* a result, whether mathematically or, in the case of material models, via physical manipulation and reasoning on its basis. Even more fundamentally, although Hughes distances himself from reductionism (see previous section), his account remains committed to giving a *substantive* account of scientific representation, in that scientific representation is thought to be characterized by the tight integration of the three theoretical ingredients of denotation, demonstration, and interpretation—that is, by more than just its functional role in inquiry.

Partially in response to these shortcomings, Suárez has proposed an alternative, *inferential account* of scientific representation—one that is unabashedly 'deflationary' in character, in that it seeks 'no deeper features to representation other than its surface features' [5, p. 771]. Giving up on the possibility of a substantive account, however, should not be misunderstood as a lack of ambition; instead, it reflects the need to refocus on the core question of what makes certain types of representation instances of *scientific* representation. In this regard, Suárez's inferential account explicitly commits itself to a demarcation between scientific and non-scientific forms of representation. While both types share certain general features of the representation relation—its asymmetry, non-reflexivity, and non-transitivity—what distinguishes scientific representations is their 'characteristic form of objectivity', which renders them of 'cognitive value because they aim to provide us with specific information regarding their targets' [5, pp. 771–772]. More specifically, on the inferential account, a representational vehicle A and its target B are related in such a way that

> A represents B only if (i) the representational force of A points towards B, and (ii) A allows competent and informed agents to draw specific inferences regarding B. [5, p. 773]

The expression 'representational force' here refers to what, on the DDI account and following Goodman, is achieved by denotation, namely the stipulated asymmetry whereby A is to be treated as a representation of B (but not vice versa). Though denotation may often be involved in generating representational force, the inferential account allows for the possibility of other sources of representational force, thereby sidestepping the problems associated with non-existent targets of representation.

It is worth comparing the second part of the above formulation of the inferential account with its correlative element in the DDI account: demonstration. Recall that one criticism of the DDI account was that it requires the actual carrying out of steps amounting to a demonstration of results, either by mathematical derivation or by physical manipulation. The inferential account's demands, by contrast, are substantially weaker, in that it merely requires that A have 'the internal structure that allows informed agents to correctly draw inferences about the B, but […] does not require that there be any agents who actually do so' [5, pp. 774–775]. Potential suitability for the purpose of enabling inferences about the target system may thus take the place of actual derivation of results. By separating the issue of representational force from the question of whether a given model supports the drawing of inferences about its target, the inferential account is able to account for various ways in which our use of models can go awry: either because a model misses its target, or because an agent lacks the requisite competence to draw valid inferences on the basis of the model. Furthermore, the account recognizes the importance of choosing formats and media of representation that enable the drawing of inferences, for example by making relevant information in the model salient to its user. As Suárez puts it, models must be 'inferentially suited to their targets' [5, p. 778]. Whether a given model is suited to its target in such an inference-enabling way is thought to be an objective fact and not reducible to the intentions or mental states on the part of a specific user.

There exists an unresolved tension within the inferential account, which, though falling far short of inconsistency, calls for further refinement. On the one hand, the account recognizes that, in spelling out the necessary conditions for scientific representation, 'the reference to the presence of agents and the purposes of inquiry is essential' [5, p. 773]. On the other hand, the account insists that we need not attribute any properties to those agents—not even, as we saw in the previous paragraph, their actual existence. This raises the question of how we are to think about such hypothetical agents, especially given that the absence of competent users need not invalidate a model's status as a scientific representation. If all that matters is that *some* hypothetical agent with sufficient inferential prowess and access to relevant information could, in principle, use the model to correctly draw inferences from it about the target system, one may wonder just how much in terms of inferential and epistemic ability it is reasonable to demand. Presumably, an omniscient (or nigh-omniscient) agent would be able to achieve a great deal more by way of inference than we ever could, and she would be able to do so on the basis of models that are far too complex for us mere mortals to comprehend. Yet we would rightly hesitate to speak of 'scientific representation' in such a case. At the very least, then, the degree of competence and inferential prowess required for a model to serve as a representation must be within human reach.

Whereas Suárez is explicit about his deflationism concerning scientific representation, others have built on his inferential account in an attempt to reinstate a full-fledged substantive account of scientific representation. Thus, Gabriele Contessa has argued for what he calls an 'interpretational conception' of scientific representation, according to which 'a vehicle is an epistemic representation of a certain target (for a certain user) if and only if the user adopts an interpretation of the vehicle in terms of the target' [20, p. 57]. On this view, an agent employing a model of the atom—say, Thomson's plum pudding model, which conceives of the electrons as embedded in an evenly distributed positive charge the size of the atom, like plums in a pudding—resolves to treat (i) the representational vehicle as a whole as standing for the target system (the atom), (ii) some elements of the vehicle as standing for some component parts of the target system, and (iii) some of the properties and relations that obtain in the vehicle as corresponding to properties and relations holding between component parts of the target system [20, p. 59]. This interpretation of the representational vehicle *in terms of* the target system, Contessa argues, allows us to explain 'why, if a vehicle is an epistemic representation of a certain target, users are able to perform valid surrogative inferences from the vehicle to the target and allows us to tell which inferences from a vehicle to a target are valid' [20, p. 61]. As he sees it, this renders the (substantive) interpretational account superior to the (deflationary) inferential account proposed by Suárez, since the latter simply posits the agent's ability to perform valid inferences from a vehicle to a target as a 'brute fact' [ibid.]. However, it seems to me that Contessa is moving too quickly here. For, as mentioned earlier in this section, Suárez is well aware of the fact that representational vehicles—in virtue of the different formats and media they employ—have different constraining and enabling effects on their users: whether a vehicle is inferentially suited to its target depends not only on factors intrinsic to the vehicle itself, but also on how we conceive of the epistemic capacities of the prospective model users—including their inferential prowess and interpretative abilities. Recognizing that the interaction between the model user and the representational vehicle is mediated by a variety of representational means renders the ability to perform valid inferences from a vehicle to a target far less mysterious than it might seem at first sight.

2.5 Realism, Instrumentalism, and the Varied Uses of Models

One of the core debates in the philosophy of science concerns the issue of *scientific realism*. Even setting aside sceptical worries about the existence of the external world or regarding the possibility of knowledge in general, one might harbour doubts about the status of scientific knowledge. Is the world really as science describes it? Are scientific claims to be taken 'at face value', and are they (by and large) true, or at least approximately true? And is science as a collective enterprise getting ever closer to a true and complete account of the world? These are some of the staple questions in the debate about scientific realism, and while they were traditionally directed at scientific theories, it is easy to see why they may also be raised—perhaps with even more urgency—in relation to scientific models. Anti-realists who are doubtful either about the historical thesis that science is moving closer to the truth or about the existence of unobservable entities posited by scientific theories, may be aghast at the casualness with which scientists readily help themselves to inconsistent models and employ idealizations and false assumptions in their model-building practices. Scientific realists, in turn, need to explain how science as a whole can be 'on the right track', when so much of it relies on models, many of which are false 'by design', as it were.

In the past, it was not uncommon to assume a stance of instrumentalism towards models. As discussed in the previous chapter (Sect. 1.3), during the time the syntactic view of theories held sway, philosophers of science tended to accord models at best a marginal role in scientific inquiry. Models were largely seen as limiting cases or approximations, or as mere heuristic tools to be used in the derivation of explanations or predictions from fundamental theories, which in turn were regarded as the proper object of realist evaluation. Models were at best thought 'to serve an auxiliary function in leading theories to the test' [21, p. 31] by generating testable predictions. As Wartofsky puts it, on this view

> the burden of commitment is passed on to the theory of which some […] model may be constructed. The postulates of the theory may make existential claims, therefore, but the model serves merely to channel these to some confrontation with experimentally testable consequences [21, p. 31].

While the view reported (but not endorsed) by Wartofsky may simply reflect an overly narrow understanding of the role of models in scientific inquiry, a more thoroughgoing instrumentalism would extend similar considerations to the underlying theory itself, with the latter

> being itself no more than an instrument for coherent organization and testing, and the question remains—*of what*? The reference beyond such theory-model 'instruments' remains forever delayed; or it is defined in terms of practical purposes, decisions concerning which lie outside the theory, but are vaguely defined as 'successful prediction' or 'control of the environment'. [ibid.]

Though instrumentalism at first sight may appear to be more modest, in that it foregoes a commitment to the truth, or approximate truth, of models and theories, this may be seen as simply pushing the crucial question one step further back, since the instrumentalist must now lay out criteria for what constitutes an instance of successful prediction or control.

Given that the focus in philosophical discussions of scientific models, as in the present chapter, is often on their representational function, one might expect the issue of scientific realism in connection with models to be largely decided by the question of how faithfully scientific models represent their targets. Against this expectation, William Wimsatt, in a paper with the programmatic title 'False Models as Means to Truer Theories', has argued that philosophers should not ignore 'the role that false models can have in improving our descriptions and explanations of the world' [22, p. 94]. Taking his lead from evolutionary biology, Wimsatt considers the case of so-called *neutral models*, i.e. models of species and populations which do not include selection pressures. Absence of selection does not entail that there is no change in the form of speciation or extinction events; rather, it might mean that such events, when they occur, are simply random. For many—perhaps most—evolutionary processes that biologists are interested in, including all those that are the result of adaptation to environmental pressures, a neutral model would be false; yet even in those cases neutral models may be essential, insofar as they provide a 'baseline' for further inquiry, 'for the explicit purpose of evaluating the efficacy of variables that are not included in the model' [22, p. 100].[4] Other epistemically valuable uses of false models include situations where an incomplete model may be used 'as a template, which captures larger or otherwise more obvious effects that can then be "factored out" to detect phenomena that would otherwise be masked or be too small to be seen', or the consideration of two or more false models which 'may be used to define the extremes of a continuum of cases in which the real case is presumed to lie' [22, p. 100]. All in all, Wimsatt considers twelve distinct ways in which false models may facilitate, or even be essential to, the search for better theories and scientific inquiry more generally, and it seems plausible to assume that any such list is likely to be incomplete.[5]

Wimsatt's observation that models, even when false—*and sometimes deliberately false*—may make a positive contribution to our overall epistemic situation is significant, in that it productively blurs a number of distinctions, whether between realist and instrumentalist stances towards models or between representational and non-representational uses of models. Non-representational uses of models, in particular, have been treated only cursorily in philosophical discussions—as hinted at in the earlier quote by Wartofsky who notes that notions such as 'control of the environment' or 'success' are often left undefined. Yet non-representational uses of models abound, in pure science as well as in more applied contexts such as engineering. Of course, not every non-representational use is of interest: someone might find a three-dimensional material model beautiful and use it as a decorative sculpture in his living room, but this would be of little relevance to us. What matters for the present argument are non-representational uses that nonetheless facilitate learning about the world—where, following Till Grüne-Yanoff's analysis of such cases, model-based learning can be defined as occurring when a model justifies 'changing one's confidence in some hypothesis about the world' [23, p. 852]. As an example, Grüne-Yanoff discusses Thomas Schelling's *checkerboard model*, which consists of two types of tokens distributed randomly across a checkerboard, with tokens being moved in each iteration according to a fixed rule, until no further movements occur. The rule is simple: if more than half of the neighbouring fields are occupied by tokens of the opposite type, a given token will move to any vacant field where this is not the case. In other words, tokens of a given type may be interpreted as having a preference for being in a neighbourhood where they are not in a minority. Over time, this reliably gives rise to patterns in which the two types of tokens are spatially segregated. Schelling did not claim that the rule reflected actual behavioural patterns or that the geometry of the checkerboard, the initial distribution of tokens, or their relative proportion represented aspects of the actual world. Yet, as Grüne-Yanoff rightly notes, we learn from Schelling's model: what

[4]Similarly, Uskali Mäki [28, pp. 12–13] notes that, in many cases, apparent falsehoods included in models are best interpreted as (true) claims about the neglibility of certain empirical factors.

[5]For example, Alisa Bokulich has argued that models that are false in virtue of being 'fictionalized'—because they involve 'fictional entities or processes that are not related to the true ones in the world by what might be thought of as a distortion or series of successive cases' [27]—can nonetheless offer genuine scientific explanations.

the model shows is that spatial segregation, of the sort found in racialized urban geographies of American cities, can occur simply due to individuals not wanting to be in a minority (rather than due to overtly racist preferences): 'The model result thus justified changing one's confidence in hypotheses about racist preferences being a necessary cause of segregation.' [23, p. 856] Another non-representational aspect of models is their *performative* use. A case in point is the discipline of economics which, as Michel Callon [24] has argued, not only *studies*, but at the same time *performs* the economy. If this sounds too abstract, consider the example of the Black-Scholes equation in finance, which purports to be a model of the efficient pricing of stock options. As Donald MacKenzie [25] has shown, the model gained empirical adequacy largely because traders adopted it as a method for identifying, say, overvalued stocks and based their selling decisions on it, thereby effectively bringing stocks in line with what the model 'demanded'—at least until the next stock market crash.[6]

Finally, it is worth noting the case of *exploratory models*, discussed in detail in Chap. 4. Much exploratory modeling aims only indirectly at the representation of actual target systems or empirical phenomena, and instead concerns itself more immediately with models that lack specific intended targets. For example, mathematical physicists might study certain model equations in higher ($d > 3$) spatial dimensions, so as to get a qualitative understanding of certain limiting cases or of the range of behaviours their model may be expected to display. Given that such models do not purport to represent, however imperfectly, any real target system, it does not seem quite right to consider them 'false' in the way that idealized representations of actual system may be deemed false. Often, the exploratory use of models aims at greater mastery and understanding of the repertoire of modeling techniques as a whole. All else being equal—with some obvious caveats, to be discussed in Chap. 4—the exploratory use of models is entirely legitimate, yet it, too, requires moving beyond the traditional narrow focus on the representational functions and uses of scientific models.

References

1. S. French, Keeping quiet on the ontology of models. Synthese **172**(2), 231–249 (2010)
2. S. Ducheyne, Towards an ontology of scientific models. Metaphysica **9**(1), 119–127 (2008)
3. R. Giere, Using models to represent reality, in *Model-based reasoning in scientific discovery*, ed. by L. Magnani, N. Nersessian, P. Thagard (Plenum Publishers, New York, 1999), pp. 41–57

[6] Against this conclusion, Mäki has argued that one should resist such talk of 'performativity': 'If it happens that certain practices and arrangements and patterns in real world finance are in line with the Black-Scholes-Merton formula, this naturally does not mean that the theoretical formula or its uttering by […] academic scholars—or by practitioners in the world of finance—"performs" those practices', because as he sees it, 'there is no constitutive relationship here between the theoretical model and some empirical practices and patterns' [29, p. 448].

4. M. Morrison, Models as representational structures, in *Nancy Cartwright's Philosophy of Science*, ed. by S. Hartmann, C. Hoefer, L. Bovens (Routledge, Abingdon, 2008), pp. 67–88
5. M. Suárez, An inferential conception of scientific representation. Philos. Sci. **71**(5), 767–779 (2004)
6. G. Abel, Modell und Wirklichkeit, in *Modelle*, ed. by U. Dirks, E. Knobloch (Peter Lang, Frankfurt am Main, 2008), pp. 31–45
7. B.V. Fraassen, *Scientific Representation: Paradoxes of Perspective* (Oxford University Press, Oxford, 2008)
8. B.V. Fraassen, *The Scientific Image* (Oxford: Oxford University Press, 1980)
9. M. Morrison, *Reconstructing Reality: Models, Mathematics, and Simulations* (Oxford University Press, New York, 2015)
10. N. Goodman, *Languages of Art: An Approach to a Theory of Symbols* (Hackett, Indianapolis, 1976)
11. N. Goodman, *Of Mind and Other Matters,* (Harvard University Press, Cambridge, Mass., 1984)
12. J. Margolis, Pragmatism, praxis, and the technological, in *Philosophy of Technology*, ed. by P.T. Durbin (Kluwer, Dordrecht, 1989), pp. 113–130
13. C. Callender, J. Cohen, There is no special problem about scientific representation. Theoria **55** (1), 67–85 (2006)
14. M. Suárez, Scientific representation. Philos. Compass **5**(1), 91–101 (2010)
15. M. Vorms, Representing with imaginary models: Formats matter. Stud. Hist. Philos. Sci. **42** (2), 287–295 (2011)
16. M. Wartofsky, The model muddle: Proposals for an immodest realism, in *Models: Representation and the Scientific Understanding*, ed. by R.S. Cohen, M.W. Wartofsky (Reidel, Dordrecht, 1979), pp. 1–11
17. R. Hughes, Models and representation. Philos. Sci. **64**, S325–S226 (1997). (Proceedings of the PSA1996, Pt. II)
18. R. Hughes, *The Theoretical Practices of Physics: Philosophical Essays* (Oxford University Press, Oxford, 2010)
19. C. Pincock, Conditions on the use of the one-dimensional heat equation, in *Essays on the Foundations of Mathematics and Logic*, ed. by G. Sica (Polimetrica, Monza, 2005), pp. 85–98
20. G. Contessa, Scientific representation, interpretation, and surrogative reasoning. Philos. Sci. **74** (1), 48–68 (2007)
21. M. Wartofsky, Models, metaphysics and the vagaries of empiricism, in *Models: Representation and the Scientific Understanding*, ed. by R. Cohen, M. Wartofsky (Reidel, Dordrecht, 1979), pp. 24–39
22. W. Wimsatt, False Models as Means to Truer Theories, in *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality* (Harvard University Press, Cambridge, Mass., 2007), pp. 94–132
23. T. Grüne-Yanoff, Appraising models nonrepresentationally. Philos. Sci. **80**(5), 850–861 (2013)
24. M. Callon, What does it mean to say that economics is performative?, in *Do Economists Make Markets? On the Performativity of Economics*, ed. by D. MacKenzie, F. Muniesa, L. Siu (Princeton University Press, Princeton, 2007), pp. 311–357
25. D. MacKenzie, Is economics performative? Option theory and the construction of derivatives markets, in *Do Economists Make Markets? On the Performativity of Economics*, ed. by D. MacKenzie, F. Muniesa, L. Siu (Princeton University Press, Princeton, 2007), pp. 54–86
26. B. Blumson, *Resemblance and Representation: An Essay in the Philosophy of Pictures* (Open Book Publishers, Cambridge, 2014)
27. A. Bokulich, How scientific models can explain. Synthese **180**(1), 33–45 (2011)
28. U. Mäki, Remarks on models and their truth. Storia del Pensiero Economico **3**(1), 7–19 (2006)
29. U. Mäki, Performativity: Saving Austin from MacKenzie, in *EPSA 11: Perspectives and Foundational Problems in Philosophy of Science*, ed. by V. Karakostas, D. Dieks (Springer, Dordrecht, 2013), pp. 443–453

Chapter 3
Strategies and Trade-Offs in Model-Building

3.1 Strategies of Model-Building

In the previous two chapters, we have come across a number of accounts of how scientific models may be thought of as functioning in general. This chapter will look in some detail at a number of case studies from across the natural sciences in order to identify recurring strategies of model-building. Speaking of 'strategies' in the context of scientific modeling may call out for explanation. Richard Levins, in the title of an influential paper published in 1966, refers to 'The Strategy of Model-Building in Population Biology' in the singular. Yet in the main part of his argument, he notes that 'several alternative strategies have evolved' [1, p. 422], each of which models biological populations in a way that sacrifices one of several theoretical desiderata—generality, realism, and precision—to the others. Models in population biology, Levins argues, thus are subject to inevitable trade-offs. Where Levins distinguishes between alternative strategies of model-building *within* a specific discipline, Peter Godfrey-Smith, more recently, has suggested that we should 'treat models and model-building as characteristic of one particular approach to theorizing, a *strategy of model-based science*' [2, p. 725]. The present chapter aims to steer a middle path between conceiving of model-based science as a unitary strategy of scientific theorizing and distinguishing between different discipline-specific strategies of model-building. The underlying methodological assumption in what follows is that it is possible to identify a 'middle range' of recurring strategies that cut across different scientific disciplines. In this sense, the plurality of strategies of model-building to be discussed in this chapter is located somewhere between the overarching general accounts in Chaps. 1 and 2, and the intra-disciplinary concerns that might dominate discussions in individual disciplines.

Among the general types of scientific models to be discussed in what follows are *phenomenological* models, *causal-microscopic* models, and instances of *target-directed* modeling. This is by no means an exhaustive or mutually exclusive

A. Gelfert, *How to Do Science with Models*, SpringerBriefs in Philosophy,
DOI 10.1007/978-3-319-27954-1_3

taxonomy, yet discussing an example of each type will help bring out recurring features of model-building strategies that can be found across a range of different branches of science. Thus, whereas phenomenological models start from questions such as ‘What are the sets of phenomena that naturally occur?’ or, more generally, ‘What are the world’s possible manifestations?’, causal-microscopic models attempt to answer the question ‘How does it work?’.[1] It has been claimed, notably by Nancy Cartwright, that phenomenological models, in virtue of making an honest attempt to relate to the world of phenomena directly (rather than through theoretical intermediaries, such as positing unobservable causal mechanisms or allegedly fundamental laws of nature), are better able to capture reality in its full empirical detail. However, as we shall see in Sect. 3.2.3, this leads to rather counterintuitive restrictions on what types of models are permissible, especially in the realm of quantum physics. As it turns out, phenomenological models—though purportedly less ‘theoretical’ in character—have trouble accounting for the sense of intelligibility that comes with building causal-microscopic models. Target-directed modeling, finally, involves a ‘match-making’ process between an existing or novel structure and a target system and then analyzing to what extent the model indeed represents its target accurately.

Each type of model-building strategy has its advantages and disadvantages, and different strategies are suited to different purposes. While this may sound like a truism, it raises the question of what determines the suitability of a given strategy to a particular purpose. A philosophical account that gives centre-stage to the varied functions of models, I submit, is best able to give some unity to the bewildering array of model-building strategies. Thus, target-directed modeling may be appropriate when one wants to get a tentative grasp of a phenomenon for which one lacks causal or theoretical understanding. Phenomenological models may be tailored more precisely to fit specific cases, while causal-microscopic models may allow a researcher to pick out a class of target systems that broadly share the same underlying causal structure. Different intended functions of models come with different sets of desiderata, and it will often be impossible to simultaneously maximize all desiderata. This leads to the question of trade-offs, mentioned above (and discussed in detail in Sect. 3.5), which first attracted attention amongst population biologists, but which is increasingly becoming relevant to such areas as nanophysics, or so I shall argue (Sect. 3.6). Trade-offs are also a common occurrence in the context of application (Sect. 3.7), not least when results are needed before one has the chance to assemble a full representation of the target system; sometimes, instead of aiming for an empirically adequate model of the total target system, it may be necessary to make do with partial representations that are adequate for specific (e.g. practical) purposes.

[1] I am borrowing this way of contrasting phenomenological and mechanism-based models from [29, p. 427].

3.2 The Case of Superconductivity: Ginzburg-Landau Approach and the BCS Model

An important characteristic of solids is their electric behaviour, which allows us to group them into, amongst others, insulators, conductors, and semiconductors. In the absence of an external electric field, most solids—including conductors—are neutral. How a solid responds once an external electric field is applied depends on a number of factors, including the type of chemical bond that dominates: when electrons are tightly bound to particular sites, as in an ionic crystal lattice, few electrons are available for sustaining an electric current, and the solid will behave as an insulator. By contrast, in metals, where the valence electrons are free to move through the crystal lattice and are not tightly bound to specific lattice sites, even a small external field will give rise to a macroscopic electric current. When travelling through the solid, electrons experience resistance in the form of thermal vibrations (phonon scattering) and lattice defects, due to grain boundaries, inclusions, or other impurities; these are the reason why, under normal circumstances, a conductor maintains its resistance even at very low temperatures. Yet, in 1911, Heike Kamerlingh Onnes (1853–1926), who was studying the electrical resistance of pure metals at temperatures only a few degrees above absolute zero, found that, in mercury, the resistance drops abruptly to zero when the sample is cooled below a critical temperature T_c of –269 °C (4.2 K). Since then, such transitions to a *superconducting state* have been observed in hundreds of substances, mostly at critical temperatures just above absolute zero, but in some cases—the so-called *high-T_c superconductors*—at temperatures as high as 138 K.

Superconductivity is associated with other striking phenomena. Perhaps the most iconic laboratory demonstration is a small magnet levitating above a superconductor (see Fig. 3.1).

This phenomenon is caused by the Meissner effect, which consists in the expulsion of magnetic flux from a metal when it undergoes a phase transition to a superconducting state. The expulsion is due to electric surface currents, which in

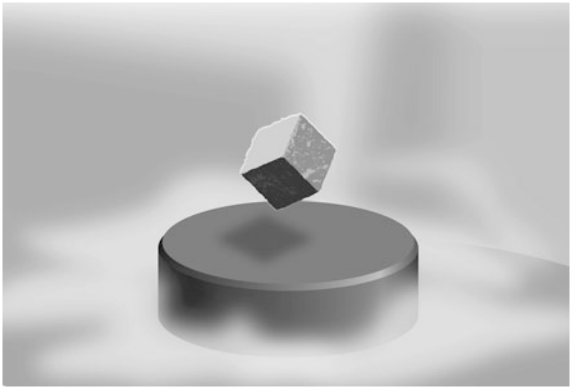

Fig. 3.1 A magnet levitating above a superconducting disk as the result of the Meissner effect

turn induce a magnetic field that cancels out the applied magnetic field within the superconductor's interior. Due to the virtually infinite conductivity of the superconductor, the currents producing this effect do not decay over time and the magnetic field they produce can counteract that of a small magnet, which then floats above the superconductor. Other changes associated with the superconducting state involve the solid's thermal conductivity and specific heat. The salience of these various phenomena prompted intense theoretical speculation and led researchers to devise a number of models that attempted to describe superconducting behaviour. Among the first were the brothers Fritz and Heinz London who, in 1935, proposed a two-fluid model of the superconductor, which posited that a fraction of the conduction electrons behaved abnormally, in that they were immune to scattering by impurities or the quantized vibrations of the crystal lattice (phonons); the relative proportion of such superconducting electrons defines an order parameter for the phase transition (and vanishes above the critical temperature). This idea required a modification of the usual electrodynamic equations—not by calling into question Maxwell's fundamental equations, but by modifying Ohm's law. The London brothers' macroscopic model happily accommodated measurable regularities such as the Meissner effect and predicted the (small) penetration depth of an external magnetic field inside the superconductor.[2] Further refinements of this approach, notably a non-local generalization of the Londons' equations by Brian Pippard, eventually led Vitaly Ginzburg and Lev Landau to propose, in 1950, their phenomenological model, to be discussed below; a microscopic explanation, based on a theoretical model of a purported causal mechanism, was not forthcoming until 1957, when John Bardeen, Leon Cooper, and John Robert Schrieffer developed what has come to be known as the *BCS model* of superconductivity (see Sect. 3.2.2).

3.2.1 Ginzburg and Landau's Phenomenological Approach

Phenomenological models take as their starting point known phenomena, remaining largely neutral with respect to the purported causal mechanisms or underlying fundamental theory. This is not to suggest that phenomenological models do not rely on theory at all or never include theoretical assumptions that go well beyond what is given in the form of empirical observations. Indeed, as the discussion of the Ginzburg-Landau and BCS models of superconductivity will show, the extent to which a model may be deemed 'phenomenological' is a matter of degree. Thus, while the BCS model is more theoretically ambitious than the Ginzburg-Landau model, in that it develops a theoretical model of an underlying microscopic mechanism, it still does not amount to a derivation from 'first principles', i.e. from the fundamental laws of quantum mechanics. In scientific usage, too, one finds

[2]See [30] for further discussion of the London model.

considerable fluidity of the terms 'phenomenological' and 'theoretical'—which is why Ginzburg and Landau's phenomenological approach is often referred to as the Ginzburg-Landau *theory*.

The starting point of Ginzburg and Landau's approach to the phenomenon of superconductivity was a more general theory, developed by Landau in 1937, which explained second-order phase transitions in fluids in terms of the minimization of the Helmholtz free energy. Following the London brothers, Ginzburg and Landau conceived of the conducting electrons as constituting a fluid that could appear in two phases, a superconducting phase and a normal (non-superconducting) phase.[3] However, they extended the phenomenological London model in a number of ways: first, by taking into account spatial variations and, second, by adding on certain quantum-mechanical considerations, in order to account for the observation that the motion of the 'electron fluid' is affected by the presence of magnetic fields. Specifically, the 'strength' of the superconducting state was to be defined by an order parameter ψ, a complex-valued quantity with some of the characteristics of a quantum mechanical wave function ('pseudo-wave function'):

$$\alpha\psi + \beta|\psi|^2\,\psi + \frac{1}{2}\left(-i\hbar\nabla - \frac{2e\boldsymbol{A}}{c}\right)^2\psi = 0$$

where α and β are experimental constants and $\boldsymbol{A}$ is the local vector potential. Spatial variations in the concentration of superconducting electrons and the kinetic energy of the supercurrent contribute to the free energy; minimizing the free energy in order to calculate the equilibrium state leads to a complicated nonlinear differential equation for ψ which, together with a standard formula for the electric current density (Gordon's formula, which had already been employed by the London brothers, but now needed to be applied to the pseudo-wave function ψ), constitutes the Ginzburg-Landau equations. Several aspects of the model are worth noting: the spatial gradient of ψ favours long-range order and accounts for why spatial changes in a superconductor occur at the scale of a characteristic coherence length. Another characteristic length, the penetration depth, which determines the quick exponential decay of an external magnetic field inside the superconductor, also emerges naturally from within the Ginzburg-Landau model.

Insofar as the Ginzburg-Landau approach conjoins disparate theoretical insights, building on earlier descriptions of the phenomenon (the classical London equations) and uniting them with new theoretical concepts (the quantum wave function), it cannot be 'derived' from any one of the theories involved: models such as Ginzburg and Landau's 'are not models of any of the theories that contribute to their construction' [3, p. 244].[4] Noting that physicists describe Ginzburg and Landau's

---

[3]This discussion follows [31, p. 248f.].

[4]There has been considerable debate about whether the case of superconductivity supports Cartwright's claims, or whether it can be accommodated by theory-driven accounts of modeling. (For a defence of the latter claim, see [33].) At the same time, as Cartwright points out in a joint

approach as proceeding from broadly inductive generalization on the basis of experimental findings, Towfic Shomar gives an apt characterization of the goal of such phenomenological modeling: namely, 'to present a mathematical structure that can be consistent with a representation of the phenomenon, trying to relate different bits and pieces from the shattered information provided through years of experimentation' [4, p. 1262]. Furthermore, in line with the textbook characteristics of a phenomenological model, the Ginzburg-Landau model does not purport to give a fundamental microphysical explanation of the phenomenon of superconductivity. To be sure, by drawing on theoretical elements such as Gordon's formula, Ginzburg and Landau's model 'also depended partly on microscopic factors [… and] employed their knowledge about fundamental theories; yet nobody considered their model as fundamental' [4, p. 1262].

Instead of attempting to either deduce a model from fundamental theory or identify a fundamental causal mechanism, the Ginzburg-Landau model draws liberally on general thermodynamic results and relations, combining them with an eclectic mix of theoretical tools in an attempt to reproduce a number of macroscopic properties, ranging from the general shape of the phenomena (e.g. the Meissner effect) to specific features such as the characteristic penetration depth and coherence length. Just how much the phenomenological approach of Ginzburg and Landau's model was held against it, can be seen from the initial disapproval it received from more theory-minded physicists:

> Why in 1950 did it take such a long time for the Ginsburg-Landau theory to be recognized? […] Well, there is quite a good reason for this. […] It was that in the Ginsburg-Landau theory the parameter κ […] is determined by the penetration depth and by certain other parameters such as the transition temperature. The penetration depth in London theory, which the Ginsburg-Landau theory incorporates, is fixed by the number of superconducting electrons and their mass. In other words, the penetration depth is a fundamental parameter according to London. […] But Ginsburg and Landau implied that when you alloy a superconductor, making the mean-free-path shorter, the penetration depth increases and κ changes because *the fundamental parameters which go into the theory change*. I found that quite unacceptable. [5, p. 8]

Even though the Ginzburg-Landau model offered a 'prepared description' of the phenomenon of superconductivity—that is, it succeeded in 'presenting the phenomenon in a way that will bring it to the theory' [6, p. 133]—and even today is seen by some as 'a more fruitful theoretical representation to understand and to predict the features of superconductivity and superconductive materials' [4, p. 1256], it lacks the grounding in fundamental theory that many physicists continue to regard as key to a full theoretical explanation. As Pippard recounts, the successes of the phenomenological Ginzburg-Landau model came to be accepted

(Footnote 4 continued)

paper with Mauricio Suárez, the position she and her collaborators defend has sometimes been misinterpreted as an outright rejection of any constraining role of theory, when in fact it only asserts 'that theories function as tools, not as sets of models already adequate to account for the startling phenomena that reveal their power' [32, p. 66].

only once they had been reproduced using the microscopic BCS model: 'It was only, I think, when Gorkov produced from the Green's function treatment of the B.C.S. theory an explicit demonstration of how the microscopic parameters could be interpreted, that the Ginsburg-Landau theory fell into place.' [5, p. 9] By contrast, in the case of high-T_c superconductivity, physicists 'are still very far from having a generally agreed microscopic model', yet 'substantial progress has been achieved' in using Ginzburg-Landau-type models 'to calculate the observable electromagnetic properties' [7, p. 134], attesting to the power of the phenomenological approach.

3.2.2 Bardeen, Cooper, and Schrieffer's Microscopic Model

What prompted theoretical physicists, notably John Bardeen, to renew their efforts to develop an alternative to the phenomenological Ginzburg-Landau model by pursuing a theoretically grounded microscopic model, was the experimental confirmation, in 1950, of the so-called *isotope effect*. The phenomenological models so far, from the Londons' two-fluid model onwards, had all assumed that superconductivity was a purely electronic phenomenon. The isotope effect, however, showed that the critical temperature at which electrical resistance vanishes depends strongly on the isotopic mass of the substance, which is a characteristic of the atoms in the crystal lattice. This suggested that the crystal lattice somehow had to be involved in bringing about the superconducting state.

Two theoretical ideas preceded, and drove, the formulation of the microscopic BCS model of superconductivity. First, it was shown that electrons, which would normally repel each other, may under certain conditions experience an attractive force when in a crystal; second, as Cooper demonstrated, electrons with opposite momentum can form correlated pairs (now known as 'Cooper pairs'), allowing them to interact with each other via phonons, thereby changing their individual momenta without varying the total (zero) momentum of the electron pair.[5] This mode of interaction, in the presence of the lattice potential, may lead to an overall *decrease* of total potential energy that is greater than the increase in *kinetic* energy associated with the electrons' moving about in the crystal (thus carrying an electric current). In other words, the ground state of a system—that is, the state in which the system is most energy-efficient—may correspond to a situation in which some electrons move about freely in Cooper pairs, rather than each being bound to individual atoms in the crystal lattice. If this is the case, the substance will display superconducting behaviour.

The BCS model of superconductivity reflects this theoretical picture of how electrons behave at the microscopic level in a superconductor, by modeling the

[5]For an insightful discussion of how the notion of 'electron pairing' developed over time, see [28, pp. 140–145].

behaviour of the system as a whole as the collective effect of a small number of mechanisms, each of which is represented by a separate theoretical model. These are then to be added up to form the overall Hamiltonian (after [8, p. 1179]):

$$H = \left(\sum_{k > k_F} \epsilon_k n_{k\sigma} + \sum_{k < k_F} |\epsilon_k| (1 - n_{k\sigma}) \right) + H_{Coul}$$
$$+ \frac{1}{2} \sum_{\boldsymbol{k},\boldsymbol{k}',\sigma,\sigma'\boldsymbol{\kappa}} \frac{2\hbar\omega_{\boldsymbol{\kappa}} |M_{\boldsymbol{\kappa}}|^2 c^*(\boldsymbol{k}' - \boldsymbol{\kappa}, \sigma') c(\boldsymbol{k}', \sigma') c^*(\boldsymbol{k}' + \boldsymbol{\kappa}, \sigma) c(\boldsymbol{k}, \sigma)}{(\epsilon_{\boldsymbol{k}} - \epsilon_{\boldsymbol{k}+\boldsymbol{\kappa}})^2 - (\hbar\omega_{\boldsymbol{\kappa}})^2}$$

Thus, one finds contributions to the Hamiltonian that represent the movement of all electrons through the crystal potential field, the Coulomb repulsion between electrons (partially 'screened off' by the positive lattice ions), as well as the phonon-mediated electron–electron interaction. The first two contributions are not specific to the BCS model of superconductivity: all conduction electrons in a metal ('Bloch electrons') have a certain kinetic energy, associated with their movement, and experience a periodic lattice potential. Likewise, all electrons in close proximity to one another will experience some degree of mutual Coulomb repulsion; how much of it is screened off depends on the geometry of the lattice. The geometry of the crystal lattice is itself an important, though sometimes overlooked, ingredient of the model: after all, each of the additive terms of a quantum many-body Hamiltonian is itself the sum over all entities involved in the process in question; however, which particles we need to sum over depends partly on the stoichiometry of the crystal lattice.

It is in the last term of the Hamiltonian, which specifies an effective (indirect) electron–electron interaction, that genuinely new content enters the BCS model. Unlike the screened Coulomb potential, this electron–electron interaction does not arise from any properties the electrons have either intrinsically or because of immersion into a uniform crystal; rather, it arises from dynamic interactions between electrons and phonons. The fundamental idea is that an electron passing through a crystal deforms the lattice in its immediate neighbourhood. In the formalism of quantized lattice phenomena, deformation is represented microscopically as the absorption or emission of phonons (corresponding to the intensity of particular normal modes of vibration). A second electron passing by may then 'register' this lattice deformation and react to it. This results in an effective—indirect, phonon-mediated—electron–electron interaction, which is independent of the usual Coulomb interaction and, therefore, need not be repulsive. Indeed, it is the emergence of an *attractive* indirect electron–electron interaction that is credited with bringing about the formation of Cooper pairs which, on the BCS model, are the microscopic basis of the phenomenon of superconductivity.

Taking a theoretical picture of a possible causal mechanism as its starting point, the BCS model thus proceeds *constructively*, drawing on theoretical resources to model the imagined processes separately, rather than aiming for a 'prepared description' of the phenomenon per se.

3.2.3 How Phenomenological is the BCS Model?

Earlier, in Sect. 3.2.1, I mentioned that contrasting the Ginzburg-Landau and the BCS models would lead to the conclusion that the extent to which a model may be deemed 'phenomenological' is a matter of degree. The discussion so far has made clear that, both in terms of their aims and strategies, the two approaches differ considerably. Where the Ginzburg-Landau model aims at capturing the macroscopic phenomena in a way that makes them amenable to mathematical analysis, the BCS model conjectures microscopic mechanisms, which are then modeled using the complete apparatus of many-particle theory, in the hope of finding behaviour *in the model* that will map onto the observed phenomenon. Whereas the Ginzburg-Landau model is closely tied to empirical phenomena and is able to reproduce, and give meaning to, two characteristic lengths—the correlation length and the penetration length—the BCS model, like other quantum many-body models, offers little to go on in terms of easily accessible empirical content. Its component parts are theoretical representations of posited fundamental mechanisms, which may or may not be related to macroscopically observable quantities.

One might think that, in light of this prima facie dissociation between the phenomenon—superconductivity, with all its attendant empirical aspects—and the piecemeal nature of the BCS Hamiltonian, whose components are determined by theoretical fiat, adherents of the phenomenological approach might turn up their noses at the BCS model. Not so. Its saving grace, from the viewpoint of proponents of phenomenological models, is the fact that the BCS model is not derived from 'first principles', i.e. from what purports to be fundamental theory. Talk of 'first principles' (or, in the context of simulations, '*ab initio*') is widespread among condensed matter physicists, who use the expression in explicit contradistinction to phenomenological approaches:

> The *first principles* approach to condensed matter theory is entirely different from this. It starts from what we know about all condensed matter systems—that they are made of atoms, which in turn are made of a positively charged nucleus, and a number of negatively charged electrons. The interactions between atoms, such as chemical and molecular bonding, are determined by the interactions of their constituent electrons and nuclei. All of the physics of condensed matter systems arises ultimately from these basic interactions. If we can model these interactions accurately, then all of the complex physical phenomena that arise from them should emerge naturally in our calculations. [9]

While the BCS model's strategy may superficially resemble the first-principles approach, in that it likewise assumes that all the relevant physics arises ultimately from a small number of basic interactions—which therefore, in contrast to the strategy of phenomenological modeling, need to be identified first, before one attempts to reproduce the empirical phenomena—it is far more selective and constructive in the way it proceeds. In particular, there is no attempt to derive the BCS Hamiltonians from 'fundamental theory', if by that one means the complete set of coupled $\sim 10^{23}$ equations that a quantum mechanical representation of the solid would involve. Instead, the constituent parts of the model's Hamiltonians are conjectured 'bottom up', by an exercise of the modeler's theoretical imagination.[6]

Furthermore, the components of the BCS Hamiltonian happen to correspond, by and large, to a set of 'basic interpretative models' that have been studied independently and are well-understood, both on theoretical grounds and from empirical contexts [3, p. 264]. For adherents of the phenomenological approach such as Cartwright, this renders them inoffensive, as they have proved their mettle in the past by successfully describing empirical phenomena. Hence, basic interpretative model Hamiltonians, such as the textbook examples of the central potential, scattering, the Coulomb interaction, the harmonic oscillator, and kinetic energy, may be considered an innocuous part of the background theory—in much the same way that the Ginzburg-Landau model relies on background theories such as thermodynamics and statistical physics. Indeed, Cartwright turns this concession into a broader thesis about quantum theory itself which, she argues, 'extends to all and only those situations that can be represented as composed of central potentials, scattering events, Coulomb interactions and harmonic oscillators' (and possibly a small number of others that may in due course be added to our 'catalogue of interpretative models'; [3, p. 265]). Only the use of those five or so stock examples, Cartwright argues, is licensed by 'bridge principles' which help 'make the predictions about what happens intelligible to us' [3, p. 246]. Representative models that tell us what happens in specific situations—for example, when superconductivity occurs—need to be built up from these basic interpretative models. As Margaret Morrison puts it, on Cartwright's view interpretative models are a way of 'fitting out' abstract theoretical principles 'in more concrete form before representative models can be built in a principled or systematic way' [10, p. 68].

While I concur with Cartwright that theoretical 'first principles' can rarely be applied straightforwardly to concrete situations, I disagree with her restrictive view that only a handful of interpretative models (all of which have counterparts in classical physics) are able to make quantum phenomena 'intelligible' to us. Intelligibility is, of course, an important function of modeling; as Mieke Boon and Tarja Knuuttila put it rather succinctly, '[m]odels are typically constructed in such a way that they constrain the problem at hand [...] thereby rendering the situation more intelligible and workable' [11, p. 695]. But intelligibility can be achieved in a variety of different ways, and prior familiarity with classical counterparts is neither necessary nor sufficient. For example, as we shall see in the next section (and in more detail in Chap. 5, Sect. 5.2), mastery of a formalism—especially when the latter is set up in a way that allows one to 'build up' more complex representations from simpler ones—can equally contribute to the intelligibility of representational models and to our understanding of how they relate to the specific situations at

[6]Because of the presence of parameters in the model that have not been derived from 'first principles', the BCS model is sometimes classified as 'phenomenological' by physicists: 'However, [the BCS theory] must be considered as a phenomenological theory with respect to the use of an "effective potential" which describes the Coulomb and phonon-induced interactions between the electrons in a model.' [34, p. 79].

hand. In other words, intelligibility arises not just from being able to link a representative model to an empirically and predictively successful interpretative model, but also from theoretical integration and from 'the presence of model resources that can be manipulated' by the model user [12, p. 29]. Yet, Cartwright is adamant that 'with each new case it is an empirical question whether these models, or models from some other theory, or no models from any theory at all will fit' [3, p. 266]. Hence, even as proponents of phenomenological models on this occasion defend the more theoretically-oriented approach of the BCS model, they insist that specific knowledge of the target phenomenon, which the model is meant to render intelligible, always trumps whatever theoretical reasons one might have for positing specific mechanisms and modeling them separately.

3.3 The Hubbard Model: Constructing Many-Body Models

The case of the Ginzburg-Landau approach and the BCS model, discussed in the previous section, illustrates nicely how, even where models are part of the same research programme, significant differences persist. Where the BCS model aimed to specify theoretical models of the purported causal mechanisms operating at a microscopic level in the superconductor, Ginzburg and Landau approached the problem of superconductivity from a more global perspective, attempting to account for features of the phenomenon itself rather than searching for causal mechanisms at the quantum level. The BCS model, in this sense, is more *theory-driven* than Ginzburg and Landau's phenomenological model. Yet even within the theory-driven paradigm of modeling, different types of approaches can be distinguished. The most ambitious such approach—deriving a model from 'first principles', i.e. by applying the underlying theory to the complete problem situation (e.g. a solid consisting of $\sim 10^{23}$ interacting particles)—has already been mentioned, but in most cases is simply not feasible. The BCS model is more typical: it stipulates a limited number of causal mechanisms and gives a theoretical description of this more restricted model system. The example to be discussed in this section, the Hubbard model of strongly correlated many-body systems with itinerant electrons, illustrates two further theory-driven approaches, which will be found to be complementary. The first such approach *simulates* a 'first-principles' derivation, not for the complete system (an extended solid), but only for its building-blocks (a cell consisting of an atom and its nearest neighbours). The second approach proceeds entirely *constructively*, in that it draws on the mature mathematical formalism of many-particle physics to create—'from scratch', as it were—the relevant components of the Hubbard Hamiltonian. As we shall see, this second approach casts doubt on Cartwright's insistence, already mentioned in the previous section, that only Hamiltonians that are licensed by specific bridge principles linking them to empirical features of the problem situation, are legitimate candidates for quantum many-body models.

The first approach takes as its starting point not the full theory of all $\sim 10^{23}$ particles in the solid, but instead begins from the smallest building-block of the extended crystal, by considering the minimal theory of two atoms that are gradually moved together to form a pair of neighbouring atoms in the crystal. One can think of this way of constructing models as involving a thought experiment regarding how a many-body system condenses from a collection of isolated particles. Such an approach remains firmly rooted in 'first principles', in that the thought experiment involving the two neighbouring atoms approaching one another is being calculated using the full theoretical apparatus (in this case, the theoretical framework of non-relativistic quantum mechanics). Needless to say, numerous background assumptions and approximations need to be made in deriving the final model equations, not least in order to capture aspects of the intended target system. For example, it is assumed that the lattice potential is strong and the mobility of the electrons small (though not zero, as otherwise no itinerant behaviour of the electrons could be expected). With these assumptions in place, the many-body system's overall Hamiltonian can be approximated as the sum of the atomic (single-particle) Hamiltonians, and the wave function as the atomic wave functions. This significantly eases calculations, as the wave functions will then satisfy Bloch's theorem, which states that in a system with periodic lattice potential, the wave function should be invariant with respect to translation (except for a phase factor). Assuming further that the wave functions are highly localized, the effect that two neighbouring particles have on each other as they are being moved closer together can be approximated by neglecting higher-order terms, leaving as contributions only the non-interacting part of the Hamiltonian and—in keeping with the idea that only the basic building-block of the crystal should play a role—the most important interactions between nearest neighbours.

In order to convey a sense of what the result of such a derivation from 'first principles', restricted to neighbouring atoms in a unit cell, looks like, consider the interacting part of the Hamiltonian H_{ee}:

$$H_{ee} = \frac{1}{2}\sum_{ijkl} v(ij;kl)\hat{a}^{\dagger}_{i\sigma}\hat{a}^{\dagger}_{j\sigma'}\hat{a}_{l\sigma'}\hat{a}_{k\sigma}$$

where the sum runs over neighbouring atoms. Much of the physics is contained in the matrix element $v(ij;kl)$, which is constructed from atomic wave functions φ:

$$v(ij;kl) = \frac{e^2}{4\pi\varepsilon_0}\iint d^3r_1 d^3r_2 \frac{\varphi^*(\vec{r}_1-\vec{R}_i)\varphi^*(\vec{r}_2-\vec{R}_j)\varphi(\vec{r}_2-\vec{R}_l)\varphi(\vec{r}_1-\vec{R}_k)}{|\vec{r}_1-\vec{r}_2|}$$

The matrix element bears a clear resemblance to the classical Coulomb potential, $F \sim q/(4\pi\varepsilon_0|\overrightarrow{r_1}-\overrightarrow{r_2}|)$, but it also takes into account the quantum effects between the two particles, as indicated by the 'mixed' integral. As a final step of approximation, it can be assumed that, because of the small overlap between atomic wave functions centred around different lattice sites, the *intra-atomic* matrix element $U = v(ii;ii)$ will dominate. This also allows one to replace the mixed creation and annihilation operators, $\hat{a}_i^\dagger$ and $\hat{a}_j$, with the simple number operator $\hat{n}_i = \hat{a}_i^\dagger \hat{a}_i$,. Adding the result to the non-interacting part of the Hamiltonian, which accounts for the movement of electrons from one lattice site i to another j (with the so-called 'hopping integrals' T_{ij} indicating the likelihood of such movement), one arrives at the final *Hubbard Hamiltonian:*

$$H = \sum_{ij\sigma} T_{ij} \hat{a}_{i\sigma}^\dagger \hat{a}_{j\sigma} + \frac{1}{2} U \sum_{i\sigma} \hat{n}_{i\sigma} \hat{n}_{i,-\sigma}$$

Such 'derivations' of many-body models from 'first principles'—with restrictions and approximations as noted above, and often with considerable degree of hindsight—are usually found in textbooks of many-body theory (e.g. [13]). However, while such a derivation makes vivid which kinds of effects—e.g., single-particle kinetic energy, particle–particle Coulomb repulsion, and genuine quantum exchange interactions between correlated particles—may be expected to become relevant, they are not the only way one can go about constructing a many-body Hamiltonian. This is where the second kind of procedure in model construction—what I shall call the *formalism-driven* approach—needs to be highlighted. Far more ubiquitous than is commonly acknowledged, this approach helps itself to (physically interpreted) mathematical formalisms as tools for the construction and interpretation of models.[7] Considering mathematical models as the *output* of well-established formalisms, rather than as either derived from theoretical first principles or tailored to specific empirical situations, drives home the point that many models enjoy a considerable degree of independence from specific experimental contexts, and even from quantitative standards of accuracy.

On the account I am proposing, *formalism-driven* model construction relies on the availability of a 'mature mathematical formalism', i.e. of 'a system of rules and conventions that deploys (and often adds to) the symbolic language of mathematics; it typically encompasses locally applicable rules for the manipulation of its notation, where these rules are derived from, or otherwise systematically connected to, certain theoretical or methodological commitments' [14, p. 272]. In order to understand how the formalism-driven strategy in model construction works, let us return to the case under consideration, namely strongly correlated systems with itinerant electrons. How is one to model the itinerant nature of conduction electrons in such metals as cobalt, nickel, and iron? In order to answer this question, the formalism-driven approach helps itself to an existing theoretical resource: the quantum mechanical formalism of so-called creation and annihilation operators,

[7] Regarding the notion of 'mathematical formalisms', and their ubiquity across the sciences, see [37].

$\hat{a}_i^{\dagger}$ and $\hat{a}_j$, which has its natural place in elementary particle physics, where particles and anti-particles can be created and annihilated in particle collisions at extremely high speeds. In many-body physics, creation and annihilation operators may refer to the addition, or removal, of a single particle from the many-body quantum state. Neither of these processes—total annihilation or creation *ex nihilo* of particles at high energies, or the removal of an electron from a solid—can, however, be assumed to take place in a metal, where electrons *flow*, rather than being destroyed or removed. From the viewpoint of 'fundamental theory', therefore, turning to the formalism of creation and annihilation operators is by no means an obvious choice. Yet the formalism can be seamlessly adapted to model the behaviour of moving electrons in a solid at room temperature (i.e. low energies): we simply need to posit, as a further rule for applying the formalism to the case of electrons in a solid, that creation and annihilation operators must never appear in isolation. Instead, an annihilation operator acting at one lattice site must always be matched by a creation operator acting at another lattice site. For, it is this sequence—disappearance of an electron from one lattice site and reappearance at another—which reflects precisely what itinerant behaviour of electrons is about in the first place: the unrestricted movement of electrons from one place to another.

As this example shows, the formalism of creation and annihilation operators, in conjunction with the basic assumption of preservation of particle number, already suggests how to model the kinetic behaviour of itinerant electrons, namely through the following contribution to the Hamiltonian:

$$H_{kin} = \sum_{ij\sigma} T_{ij} \hat{a}_{i\sigma}^{\dagger} \hat{a}_{j\sigma}$$

When the operator product $\hat{a}_{i,\sigma}^{\dagger} \hat{a}_{j,\sigma}$ acts on a quantum state, it first annihilates an electron of spin σ at lattice site *j* (provided such an electron happens to be associated with that lattice site) and then creates an electron of spin σ at another site *i*: because electrons are indistinguishable, it appears, from within the formalism, as if an electron of spin σ had simply moved from *j* to *i*. The value of the hopping integrals T_{ij} then simply determines the probability of occurrence of such electron 'hopping' from one place to another. In cobalt, nickel, and iron, the electrons are still comparatively tightly bound to their associated ions, so hopping to distant lattice sites will be rare. This is incorporated into the model for the kinetic behaviour of the electrons by including in the model the assumption that hopping only occurs between nearest neighbours. The interacting part of the Hubbard Hamiltonian can be derived even more straightforwardly: since the Coulomb force will be greatest for electrons at the same lattice site (which must then have opposite spins, σ and –σ, due to the Pauli exclusion principle), one can simply count—using the number operator $\hat{n}_{i\sigma} = \hat{a}_{i\sigma}^{\dagger} \hat{a}_{i\sigma}$ —whether this is the case for a particular lattice site *i*, in which case the electron pair contributes *U*/2 to the overall energy:

$$H_{Coulomb} = \sum_{i\sigma} \frac{U}{2} \hat{n}_{i\sigma} \hat{n}_{i,-\sigma}$$

Once again, the formalism of creation and annihilation operators itself suggests a straightforward way to account for the Coulomb contribution to the Hamiltonian, proving itself to be a powerful resource for model construction.

Formalism-driven model construction differs considerably from both 'first-principles' and phenomenological approaches. For one, it models the presumed microscopic processes such as hopping and Coulomb interaction separately, adding up the resulting components and, in doing so, constructing a many-body model 'from scratch', without any implied suggestion that the Hamiltonian so derived is the result of approximating the full situations as described by the underlying fundamental theory. Interestingly, Cartwright, whom we earlier found to be tolerant towards the use of certain 'basic interpretative models' in the BCS Hamiltonian, argues against what she calls a 'mistaken reification of the separate terms which compose the Hamiltonians we use in modelling real systems'. Although Cartwright grants that, on occasion, such terms 'represent separately what it might be reasonable to think of as distinct physical mechanisms', she insists that 'the break into separable pieces is purely conceptual' [3, p. 261] and that what is needed are 'independent ways of identifying the representation as correct' [3, p. 262]. For Cartwright, when modeling many-body systems, we face a stark choice: *either* construct phenomenological models from the empirical ground up, in a way that incorporates empirically observable regularities, *or* rely on the very small number of independently licensed 'stock interpretative models' as building blocks for more complex models. Only then do we have any assurance at all that our attempts at model construction will be true to the world of phenomena: 'When the Hamiltonians do not piggyback on [its] specific concrete features […] then their introduction is ad hoc and the power of the derived prediction to confirm the theory is much reduced' [3, p. 264].

To be sure, the formalism-driven approach often proceeds in disregard of specific empirical phenomena and in this respect might be considered as remote from Cartwright's preferred level of description—the world of physical phenomena—as the more 'first-principles'-based approaches. It is also clear that reliance on a formalism allows for greater arbitrariness. In the example at hand, a modeler can simply decree, through the choice of operators, whether (and when) electrons should flip their spins, which lattice sites they can move to, and so forth. But one should not prematurely reject the formalism-driven approach for this reason alone, just as one should not jump to the conclusion that it is simply an extension of 'fundamental theory'. It is certainly true that the formalism-driven approach is not theory-free. But much of the fundamental theory is hidden in the formalism: the formalism, as we shall see at greater length in Chap. 5 (Sect. 5.2), may be said to 'enshrine' various theoretical, ontological, and methodological commitments and assumptions. Furthermore, it seems hasty of Cartwright to dismiss as 'a mistaken reification' the tendency of many-body physicists to interpret different components of their models as 'picturing individually isolatable physical mechanisms' [3, p. 261]; for, while such interpretations are necessarily tentative, they need not be naïve: in many cases, it is *because* Hamiltonian parts can be interpreted literally, drawing on the resources furnished by fundamental theory as well as by (interpreted) domain-specific mathematical formalisms, that they generate *understanding*.

Rather than thinking of the formalism-based approach as drawing a veil over the world of physical phenomena, shrouding them in a cocoon of symbolic systems, one should think of formalisms such as the many-body operators discussed above as playing a liberating role, by allowing for the exploration of a greater number of *potential* processes and scenarios. (On the issue of model-based exploration, see also Chap. 4.) While the formalism-based approach is not unique in its ability to model selected aspects of complex systems (in particular, different co-existing 'elementary' processes), it does so with an especially high degree of economy, thereby allowing the well-versed user of a many-body model to develop a 'feel' for the model and to probe its properties with little explicit theoretical mediation.

3.4 Modeling Dynamic Populations: The Lotka-Volterra Model

The models discussed so far were drawn from physics, more specifically: the physics of quantum many-body systems. It would, of course, be misleading to think that all forms of model constructions can be adequately illustrated by examples drawn from just one branch of physics. However, my choice of examples was not so much driven by a belief in the primacy of physics, let alone quantum many-body physics, but rather by a desire to identify models and phenomena that occupy a middle ground between fundamental theory and mere descriptions of phenomena, either because of the complexity of the target system or because of uncertainty about what the general shape of the phenomenon is in the first place, and what a fundamental theory might look like for the domain in question. Solid-state phenomena and models in condensed matter physics, by the lights of fundamental physics, are already far removed from a direct application of fundamental theory—if by that one means our current best account of, say, elementary particle physics. Yet, obviously, this does not render the examples above any less heavily geared towards physics. Not least in order to address this imbalance, the present section will discuss an example from a rather different scientific discipline: population biology.[8]

The model to be discussed is the *Lotka-Volterra model* of predator–prey dynamics, which has its origins in empirical observations: in the aftermath of World

[8]Further examples from other disciplines, including theoretical chemistry and traffic flow theory, will be discussed in Chap. 4.

War I, Italian fishermen experienced lean years, even though most had anticipated abundant catches, given that there had been comparatively little fishing during the war years—which, people expected, should have given the fish populations time to recover. When the biologist Umberto d'Ancona (1896–1964) examined the statistical data—the numbers of different species sold in the fish markets of Trieste, Venice, and his hometown Fiume—he found that, during the war years, the percentage of predator species (such as sharks and other selachians) had trebled, at the expense of traditional food fish species. In other words, the lack of fishing activity during the war years seemed to have selectively benefited the predator species. D'Ancona consulted his father-in-law, the mathematical physicist Vito Volterra (1860–1940), who devised a set of equations from which it followed that, contrary to what one might expect, some degree of fishing needs to be maintained in the presence of predators, if one is to get the highest sustainable catches of desirable food fish species.[9]

Mathematically, the Lotka-Volterra model consists of a pair of first-order, non-linear, differential equations which are intended to mimic the population dynamics of a two-species systems, with one species feeding on the other. More specifically, it models the rate of change in each population as dependent on the other, though not in the same way: For the prey—typically a fast-reproducing species—the dominant effects are reproduction (proportionate to the existing population size) and mortality due to predation (proportionate to its own population size and to that of the predator species); for the predator species, mortality is due to a constant death rate, so the total number of deaths is proportionate to its population size, while the total number of births is assumed to be proportionate to both its own population size and to that of the prey which, after all, sustains the predators. This gives rise to the following equations (with x standing for the size of the prey population, y for the number of predators, t for time, a for the prey's birth rate, b for the predator's death rate, and α, β positive coefficients representing the effect each population has on the other):

$$\frac{dx}{dt} = x(a - \alpha y)$$

$$\frac{dy}{dt} = y(\beta x - b)$$

Because of the way the two populations are coupled, plotting the size of the predator and prey populations over time leads to a remarkable finding: both populations will oscillate indefinitely, with the predator population lagging slightly behind in time and the prey population overshooting more dramatically (see

[9]A few years before, in 1920, Alfred Lotka (1880–1949) had published essentially the same set of equations, though Volterra apparently had no knowledge of Lotka's work. For the original articles, see [38, 39].

Fig. 3.2 Predator-prey cycles as modeled by the Lotka-Volterra equations

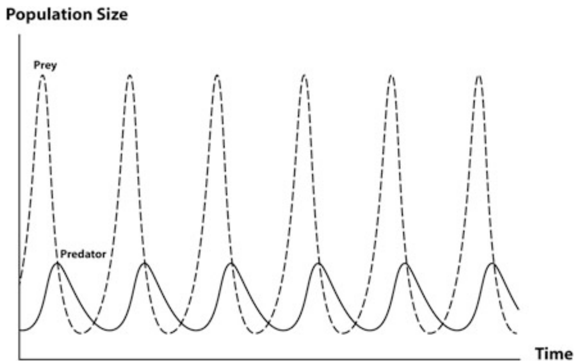

Fig. 3.2). There is no stable equilibrium that would withstand even slight perturbations (which, of course, would inevitably occur in the wild).

Each 'cycle', in the world of the Lotka-Volterra model, follows the same pattern: When there are few predators, the prey population will increase rapidly, even as the predator population begins to recover, which in turn will grow until it begins to bring down the total number of prey, even below the number that would be required to sustain the (now increased) predator population.

Michael Weisberg aptly characterizes the Lotka-Volterra model as an example of *target-directed modeling*. This strategy of modeling involves three distinct elements: 'development of the model, analysis of the model, and targeting the model to a real-world system' [15, p. 74]. Though conceptually distinct, these elements need not always be separated in practice and, in particular, should not be thought of as always coming in temporally distinct stages. The Lotka-Volterra model is a good illustration of how an initial concern for a specific target phenomenon guided (but did not, in any strong sense, determine) the development of the model equations. As Weisberg notes, 'when Volterra first constructed the mathematical structure for the Lotka-Volterra model, he had no previous biological models from which to work' [15, p. 75]; in this sense, he proceeded 'from scratch', guided only by some basic assumptions about the macroscopic dependencies between predator and prey species. In other cases of target-directed modeling, scientists may borrow existing structures and equations—not 'any old structure', but 'a structure that has an adequate representational capacity for their chosen target' [15, p. 75]. In the Lotka-Volterra case, the three components—model construction, analysis, and targeting—were deeply intertwined: the target phenomenon, the unexpected increase in the relative number of predators due to the lack of fishing, both led to the search for a suitable mathematical model (thereby 'fixing the target') and was subsequently recognized, through close analysis of the mathematical equations, as a more general consequence of the dynamics of two-species predator–prey systems. This is why, beyond Volterra's initial target phenomenon, biologists now speak more generally of *Volterra's principle*, according to which a uniform reduction of both populations, proportional to their total number, will lead to an increase of the average prey population and a decrease of the average predator population. While model construction, analysis, and targeting proceeded largely simultaneously in this case, the three elements can, on occasion, come apart. As we shall see in Chap. 4, this may give rise to certain types of *exploratory modeling*, such as when a model's mathematical (or other) characteristics imbue it with superior *potential* representational capacity. In such cases, a model may be 'in search of a target phenomenon', so to speak, and the same model equations may find unexpected applications across a range of different domains and disciplines.

3.5 The Question of Trade-Offs: Origins of the Debate

As mentioned in the introduction to this chapter, Levins argued that the models used in population biology were subject to inescapable constraints, insofar as certain theoretical desiderata—notably generality, precision, and realism—cannot simultaneously be maximized. This places significant restrictions on what individual models in population biology can achieve. For example, if one tailors a model exactly to a particular ecosystem, by including in detail all operative causal mechanisms (e.g., the various predator-prey relationships) as well as precise measurements of significant parameters (e.g., of reproduction rates and the nutritional needs of each species), this will inevitably restrict the generality of the model—if successful, it will pick out one, and only one, real target system in the world. This suggests that the theoretical desiderata of generality, precision, and realism 'trade off' against one another.

Levins's claim that not all three desiderata can be optimized simultaneously leads to a natural classification of modeling strategies into three types, depending on which desideratum 'loses out' in the process of optimizing the other two. In type I cases, the ecologist tailors her model to the specific empirical detail and causal mechanisms of a particular system, thereby sacrificing generality for realism and precision. In type II modeling, by contrast, realism is sacrificed for precision and generality; models of this type are characterized by 'general equations that give precise outputs, but involve unrealistic idealisations and assumptions' [16, p. 325]. Type III models, finally, sacrifice precision for generality and realism; while such models do not lend themselves to making quantitatively precise predictions, they are thought to be true to the dominant causal relationships that exist in the general class of systems whose behaviour they are meant to explain. Though precision, realism, and generality are all equally considered to be desiderata of modeling, it is the latter—type III models—that Levins is often thought to have promoted (cf. [17, p. 1273]). This may have been more than just a personal preference, in that Levins found it important to rehabilitate generality and realism against a perceived overemphasis on precision as the ultimate goal of model-building in his own discipline, population biology. Indeed, as Peter Taylor puts it:

> Levins' strategy is, in effect, an advocacy of exploratory modeling as a means of theory generation. I take this to be the meaning of his favoring generality and realism at the expense of precision. [18, p. 202]

As we shall see in Chap. 4, exploration is an often overlooked function of scientific modeling, and exploratory modeling, in spite of the diversity of forms it can take, is an activity that is quite distinct from target-oriented modeling.

Why might one expect the simultaneous maximization of precision, realism, and generality to be unattainable in a domain like population biology? There are two main types of reasons. The first is the result of practical and technical limitations, which constrain our ability to maximize all three desiderata; the second is due to features of the target system itself, which may, for example, make it impossible to arrive at suitable generalizations. Consider first what it would take to construct, and evaluate, type I models that aim at representing ecosystems in their full complexity. Such models would involve 'perhaps 100 simultaneous partial differential equations' [1, p. 421], each with numerous parameters, to be obtained from lengthy field studies. Even if it were possible to obtain accurate measurements of the relevant parameters, Levins argues, the resulting equations would be 'insoluble analytically and exceed even the capacities of good computers' [1, p. 421]. Furthermore, in those rare cases where solutions might be within reach, interpreting the results might still be beyond the cognitive capacity of finite human reasoners. As Jay Odenbaugh has argued, the fact that we have difficulty making sense of comparatively simple models does not bode well for the interpretation of complex, 'photographically perfect' models [19, pp. 1498–1499].

A second set of considerations stems from general characteristics of the target system itself. In order to better appreciate which features of a system may be relevant to the question of trade-offs, let us briefly look in particular at the trade-off between precision and generality.[10] In certain cases it is immediately obvious that this trade-off cannot be blamed on issues of feasibility and lack of information alone: for example, if one specifies the parameter values that go into a mathematical model more precisely, then it will trivially pick out fewer possible target systems. Conversely, by relaxing one's standards of precision, a larger set of possible target systems may be accommodated by the model. However, as Weisberg [20] has argued, this inverse relation between precision and generality, strictly speaking, only holds if generality is measured by how many logically possible target systems a model picks out. Such '*p*-generality' is conceptually distinct from '*a*-generality', which is measured by how many *actual* target systems a model applies to. It is obvious that the two can come apart: many *logically* possible target systems can be excluded on the basis of background knowledge about what the world is like, and any loss of *p*-generality that is due to the exclusion of such 'unphysical' (or otherwise uninstantiated) possibilities is not going to make any difference in actual contexts of empirical inquiry. The extent to which the intuitive trade-off between *p*-generality and precision translates into an *actual* trade-off between *a*-generality

[10] My presentation in this paragraph mainly follows [20, pp. 1075–1079].

and precision depends, Weisberg argues, on the homogeneity of the set of target systems the model is intended to apply to, as well as on the scope of inquiry—i.e., which aspects of the target system(s) are deemed relevant.

In contexts of actual scientific inquiry, the costliness of the—otherwise largely abstract—trade-off between precision and generality is thus determined by the degree of heterogeneity within the set of intended target systems: the more heterogeneous a class of target systems, the more difficult it will be to simultaneously increase precision and generality, for example by subsuming a range of target systems under one and the same model-based account. Levins likewise notes that, for population biologists today, who work under the evolutionary paradigm of selective pressures being exerted on organisms by the environment, 'environmental heterogeneity is an essential ingredient of the problems and therefore of our mathematical models' [1, p. 422]. Conversely, when dealing with highly homogeneous sets of target systems, increases in precision need not greatly affect generality, since the systems are similar in all relevant respects. Nothing in science, of course, is more similar than identical elementary particles. As John Matthewson argues:

> It is possible to model the behaviour of electrons very precisely and generally, because they all have the same properties. But it is not possible to model the behaviour of any particular type of ecosystem both precisely and generally, because ecosystems vary with respect to many of their important properties. [16, p. 331]

The lack of homogeneity in the case of biological entities, especially complex entities such as ecosystems, is reflected by the relative dearth of law-like generalizations in biology. While physics and, to a large extent, chemistry rely heavily on purported laws of nature, most biological 'laws' describe overall empirical patterns that typically allow for exceptions. To be sure, there are some biological regularities such as the Hardy-Weinberg principle, which states that, absent specific disturbances, the allele and genotype frequencies in a population remain constant. However, such biological laws typically either supervene on factors, such as evolved genetic mechanisms, that are themselves contingent (in ways that relevantly contrast with, say, physics), or apply at the systems level (e.g., ecoystems or idealized populations) rather than, with nomic force, at the object level of individual organisms.[11] Whereas an electron always responds to an external magnetic field in precisely the same way, organisms are complex adaptive systems that often exhibit a range of possible reactions to external stimuli. In addition, how a biological entity behaves often depends on its past exposure and responses; its trajectory is path-dependent. Examples include learned behaviour at the individual level and genetic drift at the species level. As we shall see in the next section, it is this *evolvability* of biological systems which is sometimes thought to render the existence of trade-offs a distinctive feature of biological models.

[11]Spelling out exactly how biological and physical 'laws' contrast with respect to universality, nomic force, or scope lies beyond the scope of this chapter. For a review of the debate about biological laws, see [35].

3.6 Trade-Offs as a Demarcation Criterion?

Talk of ‘trade-offs’, at least in relation to desiderata of theoretical models, is not as widespread in physics as it is, for example, in population biology. A quick search in physics databases reveals that the term is mostly used to refer to trade-offs between accuracy (of computer simulations and other calculations) and ease of computation, due to limited computational resources, not to trade-offs at the level of abstract desiderata of models as such. The relative sparsity of references to theoretical trade-offs in other disciplines has not gone unnoticed by philosophers of biology. Thus, Steven Orzack and Elliott Sober note:

> It is of relevance that claims about trade-offs similar to Levins’s have not, to our knowledge, arisen in physics and chemistry. [21, p. 544]

As discussed in the previous section, there are good reasons for expecting biological models to be especially prone to trade-offs. These have to do with the status of biological entities as *evolved* objects which exhibit variation, path-dependency, and adaptability. As Matthewson argues, this is what sets biology, and population biology with its emphasis on relations between (evolved) species in particular, apart from other branches of science:

> The requirement of ‘variation that leads to important downstream effects within a population’ does not arise in the other natural sciences. So population biology specifically deals with ensembles of entities that must be heterogeneous, in a way that does not arise in chemistry or physics. [16, p. 332]

It might seem, then, that the presence or absence of trade-offs in model-building might be considered a demarcation criterion of sorts between the physical sciences, which (to borrow a phrase from Orzack and Sober; [21, p. 544]) have the ‘potential for generality’, and disciplines such as ecology, population biology, and evolutionary theory, which cannot ignore the evolved, heterogeneous nature of their basic objects of investigation.

It would be wrong, however, to assume that the idea of trade-offs is wholly absent from physics and chemistry. Indeed, scientists in both disciplines are well aware of the theoretical choices that are forced upon them by the existence of unavoidable trade-offs. Daniela Bailer-Jones, in a series of interviews, has attempted to document how scientists think of models. While the sample size is too small to allow for wholesale generalizations, it is nonetheless striking that those of Bailer-Jones’s interviewees who hint at ‘trade-off’-like characteristics in scientific modeling all have a background in condensed matter physics, broadly construed as comprising both its ‘hard’ (solid-state physics) and ‘soft’ (granular media, surface physics) variety. John Bolton, one of the solid state theorists among her interviewees, is described as relating ‘the missing predictive accuracy of models with the insights provided by a model—insight compensates for lack of detail’ [22, p. 286]. In Bolton’s own words:

> [S]ometimes getting, I suppose, a possible match to reality is not everything. What you are looking for is an understanding of what's happening in nature, and sometimes a simple model can give you that, whereas a very large computer program can't. (Quoted after [22, p. 286].)

The sentiment that Bolton expresses in this quote stems, of course, from precisely the trade-off identified in the previous section: between the empirical success of one's models (as measured, amongst others, by their predictive accuracy) and the sense of understanding that comes with generality. The idea that accuracy—especially in situations where access to knowledge and (computational) resources is limited—may trade off against the explanatory goal of identifying the fundamental mechanisms that drive the system under investigation, is echoed by other interviewees. Thus, Nancy Dise observes that 'because you are limited by time and money and by your knowledge of the system you take what you believe are the most important drivers of that process', which are then included in the model.[12]

Notwithstanding the very real differences between biological entities and the fundamental entities of physics, in terms of both evolvability and heterogeneity, it may nonetheless be worth exploring alternative interpretations that do not start from presumed ontological differences between the research objects of various scientific disciplines, but instead look at differences in scientific practice in order to explain why trade-offs have been more salient in one discipline rather than another. For example, the relatively 'benign' nature of trade-offs in physics, as opposed to their salience and limiting consequences in ecology and population biology, might also be seen as a by-product of the general tendency of physicists to focus on comparatively homogeneous systems which can be characterized by the same small number of parameters across a wide range of situations. Arguably, most systems traditionally studied by physics do not exhibit the sort of 'path-dependence' and evolvability of biological systems. Yet it is not immediately obvious to what extent this preference is necessitated by fundamental facts about physical reality *in general* or simply by a preference of physicists to direct their attention at systems that display just the kind of homogeneity that we have come to expect from physics. This is not to say that homogeneity and heterogeneity are merely in the eye of the beholder: clearly, the world needs to cooperate in various ways for a researcher to be able to ignore issues of 'evolved uniqueness' when it comes to a target system or its constituents, thereby enabling her to treat the system as homogeneous. In some cases, homogeneity occurs naturally, as in the case of lattice systems such as crystals, where symmetry allows for the rare *macroscopic* expression of the underlying *microscopic* uniformity among constituent parts. In other cases, for example a population consisting of members of the same species, heterogeneity among its members is the norm, and homogeneity can at best be enforced artificially—e.g. via cloning—and only with considerable effort. These are the kinds of systems that biologists have long been accustomed to, whereas physicists in the past have tended to study systems of the former kind. But as experimental and

[12] Quoted after [22, p. 285].

computational abilities have advanced over time, there is no in-principle reason why physicsts should not also turn their attention to systems of the latter kind.

In the case of biology, it is the heterogeneity of biological entities which explains why they resist law-like generalizations; the heterogeneity itself is explained, in turn, by their status as evolved objects. Evolution and evolvability are the driving forces behind it: after all, one of the important realizations of the synthetic theory of evolution is not only that no species is quite like any other, but that no two subpopulations of the same species will typically behave in quite the same way. Ultimately, it is the evolved uniqueness of particular systems that gives rise to the overall heterogeneity among them. In contemporary condensed matter physics, the traditional focus on analyzing highly homogeneous systems—describing macroscopic phenomena basically as perturbations (of various sorts) of ordered systems that lend themselves to description in terms of law-like generalizations—has increasingly given way to a broader perspective that includes systems that are heterogeneous in ways that resemble the situation in biology. This includes such systems as granular media, quasi-crystals, or colloids, where the symmetries that assure homogeneity among target systems are broken. Granular media are known to exhibit phenomena such as *hysteresis*, whereby the behaviour of the system is path-dependent: it then becomes impossible to predict, at a macroscopic level, a system's future behaviour without knowledge of its past history. As a result, systems that appear to be in the same macroscopic state may well behave quite differently, depending on the trajectories by which each arrived in this state. While this is still a far cry from the evolutionary path-dependence of biological systems, it does introduce an 'historical' element into the study of physical systems, thus increasing their heterogeneity. As a result, increases in the ability to model and predict the specific behaviour of such systems may come at the expense of generality, given that models now need to be individuated by, and tailored to, their initial conditions and causal histories, not merely in terms of their macroscopic properties or a set of basic mechanisms that are thought to drive their behaviour.

Whereas path-dependence means that systems can no longer be classified on the basis of their macroscopic descriptions alone, the final example I wish to discuss appears to dispense with the idea of classes of target systems altogether—at least to the extent that it no longer regards it as the primary goal to explain the actual behaviour of a specific system *as an instance of* a general class of target systems. The example I have in mind concerns so-called 'fingerprint effects' in mesoscopic systems, which manifest themselves as 'time-independent stochastic magnetoresistance patterns', which 'vary between samples but are reproducible (at a given temperature) within a given sample' [23, p. 1039]. It is thought that such 'magneto-fingerprints' arise as the joint effect of, on the one hand, disorder and impurities and, on the other hand, the fact that quantum interference in mesoscopic systems acts over a characteristic length much larger than the size of an atom. As a result, each sample of a material will have its own—experimentally reproducible, yet theoretically unpredictable—"fingerprint-like" behaviour (see [24, p. 171]). It is important to emphasize the experimental reproducibility of these 'magnetic fingerprints', since—unlike in the case of thermal fluctuations—the seemingly chaotic behaviour does not ‘average out’ over time but is ‘frozen in time’, as it were. Magneto-fingerprints are unusual—and quite unlike traditional statistical features of complex systems—in that their shape and form is determined not, as it were, by a given system’s membership in a larger reference class of like systems, but instead by the brute atom-by-atom particularity of the specific sample in question.

Due to this novel combination of empirical replicability with sample-specificity, magnetic fingerprints are genuinely new characteristics that can also be exploited technologically. Thus, physicists have developed nano-scale transistors that bear ‘unique fingerprint-like device-to-device differences attributed to random single impurities’; the same group emphasizes the ‘critical need’ [25] to model such fingerprint-like behaviour. It is clear that no model of the unique ‘fingerprint-like’ behaviour of a *specific* target system can possibly generalize to another sample (unless the two are microscopically one and the same). The case of magneto-fingerprints is but one example of a broader trend in physics, as technological and computational advances make it increasingly possible to analyze matter at ever greater resolution, allowing researchers to identify highly localized, sample-specific—yet individually reproducible—regularities that previously would have been either dismissed as ‘noise’ or regarded as idiosyncracies of a given experimental setup, standing neither in need of explanation nor much chance of being accounted for by one’s models. Not just biological organisms, but matter itself—at least in its condensed form—may thus come to be recognized as ultimately consisting of individually unique and collectively heterogeneous assemblages, whose behaviour it may only be possible to predict accurately on a case-by-case basis, by sacrificing some of the generality of explanatory models for the ability to describe (and eventually exploit) the material constitution and sample-specific characteristics of systems at the nano-scale.

3.7 Models in the Context of Application

Thus far in this chapter, we have encountered a number of examples which—notwithstanding their specific idiosyncracies—are representative of more general approaches to scientific modeling: phenomenological models, causal-microscopic models, target-directed models, and so forth. In any sustained research programme, and especially in applied contexts, different approaches will typically coexist, so one should not expect particular instances of scientific modeling to fit one and only one category. This, once again, reflects the fact that scientific models may serve a number of different functions, each of which has its own set of theoretical desiderata. As the discussion of trade-offs in population biology (and, perhaps more controversially, in nanophysics) has shown, it is often not practical, and sometimes impossible, to simultaneously maximize all desiderata. For example, if the model user’s goal is to make precise quantitative predictions, this will trade off against the

—otherwise equally legitimate—goal of providing a general account that would subsume multiple targets under the same model. Given that the choice of theoretical desiderata depends on the interests and goals of the model user, the example of trade-offs already suggests that a full understanding of scientific models requires attention to the varying contexts of their application.

Yet the importance of applied contexts runs deeper than the question of trade-offs alone. For example, being aware of the relevant context of application is crucial to judging whether or not a model is sufficiently well-confirmed. This applies especially to situations that require time-sensitive decision-making. As Sandra Mitchell has argued, the traditional model of waiting until one has nearly complete information, and then basing one's future actions on predictions that reflect 'objective (or consensual) probability assignments' on its basis—what she calls the 'predict-and-act model' of decision-making—has its limitations, for,

> in cases of complex systems, it may very well be that waiting until there is agreement of confidence in the quantitative probability assigned to possible outcomes is unreasonable. For example, we may be waiting until it is too late to act to avoid seriously undesirable consequences. [26, p. 89]

Models in contemporary climate science are a case in point. Such models, no doubt, serve a representational function: they represent the Earth's climate system and its underlying processes. Yet, given the urgency of global climate change and the need for policy-relevant advice, it would be foolish to aim for a *complete* representation of all the possibly relevant processes and mechanisms—especially when this might mean postponing indefinitely the creation of a workable model. What is called for, in this particular instance of a 'context of application'—arguably, one of global significance—is not a perfect representation of the Earth's climate system in its full complexity, but a more pragmatic sense of 'adequacy-for-purpose' [27]. Yet, as Wendy Parker notes,

> adequacy-for-purpose does not work like truth and empirical adequacy [in regard to empirical fit]; from the assumption that a model is adequate for an explanatory or predictive purpose, information about how the model is likely to perform in various other respects, or information about what other properties the model is likely to possess, does not simply follow as a matter of course. [27, p. 238]

In other words, rather than aiming for a model that reflects every available detail of the target system, it may be preferable to have a model that makes adequate predictions primarily of those features that matter to us—say, changes in rainfall patterns in agriculturally productive parts of the world—even if it misrepresents other parts of the target system as a whole. Needless to say, such an approach raises difficult epistemological questions: for example, it will often be far from obvious how we can extrapolate from variable X (for which, let us assume, the model has been optimized from the start) to variable Y (which, although related to X by some causal process, was not initially the focus of achieving adequacy-for-purpose). This is why testing and cross-checking of complex models is of eminent importance, both at the process level and at the level of system-wide predictions.[13]

References

1. R. Levins, The strategy of model building in population biology. Am. Sci. **54**(4), 421–431 (1966)
2. P. Godfrey-Smith, The strategy of model-based science. Biol. Philos. **21**(5), 725–740 (2006)
3. N. Cartwright, Models and the limits of theory: quantum Hamiltonians and the BCS model of superconductivity, in *Models as Mediators: Perspectives on Natural and Social Science*, ed. by M.S. Morgan, M. Morrison (Cambridge University Press, Cambridge, 1999), pp. 241–281
4. T.L. Shomar, Phenomenologism vs fundamentalism: the case of superconductivity. Curr. Sci. **94**(10), 1256–1264 (2008)
5. B. Pippard, The historical context of Josephson's discovery. in *Superconductor Applications: SQUIDS and Machines (NATO Advanced Study Institute Series B: Physics*, vol. 21), ed. by B.B. Schwartz, S. Foner (Plenum Press, New York 1977), pp. 1–20
6. N. Cartwright, *How the Laws of Physics Lie* (Oxford University Press, Oxford, 1983)
7. A.J. Leggett, What do we know about high Tc? Nat. Phys. **2**(3), 134–136 (2006)
8. J. Bardeen, L.N. Cooper, J.R. Schrieffer, Theory of superconductivity. Phys. Rev. **108**(5), 1175–1204 (1957)
9. M.C. Gibson, Implementation and Application of Advanced Density Functions. (Ph.D. Dissertation, University of Durham), 2006. Available: http://cmt.dur.ac.uk/sjc/thesis_mcg/node6.html. Accessed 21 Sept 2015
10. M. Morrison, Models as representational structures, in *Nancy Cartwright's Philosophy of Science*, ed. by S. Hartmann, C. Hoefer, L. Bovens (Routledge, Abingdon, 2008), pp. 67–88
11. M. Boon, T. Knuuttila, Models as Epistemic Tools in Engineering Sciences: A Pragmatic Approach, in *Philosophy of Technology and Engineering Sciences (Handbook of the Philosophy of Science, Vol. 9)*, ed. by A. Meijers (Elsevier, Amsterdam 2008), pp. 693–726
12. M.S. Morgan, *The World in the Model: How Economists Work and Think* (Cambridge University Press, Cambridge, 2012)
13. P. Nozières, *Theory of Interacting Fermi Systems* (Benjamin, New York, 1963)
14. A. Gelfert, Mathematical formalisms in scientific practice: from denotation to model-based representation. Stud. Hist. Philos. Sci. **42**(2), 272–286 (2011)
15. M. Weisberg, *Simulation and Similarity: Using Models to Understand the World* (Oxford University Press, New York, 2013)
16. J. Matthewson, Trade-offs in model building: a more target-oriented approach. Stud. Hist. Philos. Sci. **42**(2), 324–333 (2011)
17. J. Justus, Qualitative scientific modeling and loop analysis. Philos. Sci. **72**(5), 1272–1286 (2005)
18. P. Taylor, Socio-ecological webs and sites of sociality: Levins' strategy of model building revisited. Biol. Philos. **15**(2), 197–210 (2000)

[13]Contemporary climate models do well on this score and in a variety of other respects, such as robustness, variety of independent sources of evidence, and fit between observations and predictions (as well as retrodictions); indeed, as Elizabeth Lloyd has emphasized, 'climate models are supported empirically in several ways that receive little explicit attention' [36, p. 228].

19. J. Odenbaugh, Complex systems, trade-offs, and theoretical population biology: Richard Levins's "Strategies of model building in population biology" Revisited. *Philos. Sci.,* **70**(5), Proceedings of the PSA2002, 2003, pp. 1496–1507
20. M. Weisberg, Qualitative theory and chemical explanation. *Philos. Sci.,* **71**(5), Proceedings of the PSA2002, 2004, pp. 1071–1081
21. S.H. Orzack, E. Sober, A critical assessment of Levins's "The Strategy of Model Building in Population Biology" (1966). Q. Rev. Biol. **68**(4), 533–546 (1993)
22. D. Bailer-Jones, Scientists' thoughts on scientific models. Perspect. Sci. **10**(3), 275–301 (2002)
23. P.A. Lee, A.D. Stone, H. Fukuyama, Universal conductance fluctuations in metals: effects of finite temperature, interactions, and magnetic field. Phys. Rev. B **35**(3), 1039–1070 (1987)
24. Y. Imry, *Introduction to Mesoscopic Physics*, 2nd edn. (Oxford University Press, Oxford, 2008)
25. G.P. Lansbergen, R. Rahman, C.J. Wellard, J. Caro, N. Collaert, S. Biesemans, G. Klimeck, L. C. L. Hollenberg, S. Rogge, Transport-based Dopant Metrology in Advanced FinFETs, 15 December 2008. Available: http://docs.lib.purdue.edu/nanodocs/153/. Accessed 21 Sept 2015
26. S.D. Mitchell, *Unsimple Truths: Science, Complexity, and Policy* (The University of Chicago Press, Chicago, 2009)
27. W. Parker, Confirmation and adequacy-for-purpose in climate modelling. Proc. Aristot. Soc. **83**(1), 233–249 (2009)
28. M. Morrison, *Reconstructing Reality: Models, Mathematics, and Simulations* (Oxford University Press, New York, 2015)
29. M.H. Krieger, Phenomenological and many-body models in natural science and social research. Fundamenta Scientiae **2**(3/4), 425–431 (1981)
30. M. Suárez, 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 University Press, Cambridge, 1999), pp. 168–195
31. P.F. Dahl, *Superconductivity: Its Historical Roots and Development from Mercury to the Ceramic Oxides* (American Institute of Physics, New York, 1992)
32. M. Suárez, N. Cartwright, Theories: tools versus models. Stud. Hist. Philos. Mod. Phys. **39**(1), 62–81 (2008)
33. S. French, J. Ladyman, Superconductivity and structures: revisiting the London account. Stud. Hist. Philos. Mod. Phys. **28**(3), 363–393 (1997)
34. M. Crisan, *Theory of Superconductivity* (World Scientific, Singapore, 1989)
35. A. Hamilton, Laws of biology, laws of nature: problems and (dis)solutions. Philos. Compass **2** (3), 592–610 (2007)
36. E.A. Lloyd, Varieties of support and confirmation of climate models. Proc. Aristot. Soc. **83**(1), 213–232 (2009)
37. A. Gelfert, Symbol systems as collective representational resources: Mary Hesse, Nelson Goodman, and the problem of scientific representation. Social Epistemology Review and Reply Collective **4**(6), 52–61 (2015)
38. A.J. Lotka, Analytical note on certain rhythmic relations in organic systems. Proc. Natl. Acad. Sci. **6**(7), 410–415 (1920)
39. V. Volterra, Variazioni e Fluttuazioni del Numero d'Individui in Specie Animali Conviventi. Memorie dell'Accademia Nazionale dei Lincei **2**, 31–113 (1926)

Chapter 4
Exploratory Uses of Scientific Models

4.1 Model-Based Understanding and the Tacit Dimension

In the popular imagination, the main goals of science include giving explanations, making predictions, applying its findings in instrumentally successful ways, and generally furthering our understanding of the world around us. Unlike explanation, prediction, and instrumental success, however, *scientific understanding* has tended to be sidelined in 20th-century philosophy of science. To a large extent, this was the result of the logical empiricist emphasis on objective criteria for assessing science, such as empirical testability and logical form. Carl Hempel's distinction between scientific explanation and scientific understanding is a good case in point. Whereas Hempel believed that explanation could be construed rigorously and objectively, namely along the lines of his famous deductive-nomological (D-N) model—according to which the explanandum is to be logically deduced from premisses containing at least one universal law—*understanding*, according to Hempel, necessarily involves a subjective element: the feeling of grasping a deeper connection between the explanandum and the explanans. On this view, 'understanding in the psychological sense of a feeling of empathic familiarity' and scientific understanding 'in the theoretical, or cognitive, sense of exhibiting the phenomenon to be explained as a special case of some general regularity' [1, p. 413] needed to be strictly separated; best, then, to dispense with the notion of understanding and its psychological overtones altogether.

Such wholesale dismissal of the notion of understanding on the part of the epistemic subject was by no means uncontroversial. Even at the time, it was met with significant criticism. Thus, Michael Scriven criticized Hempel's account for unjustifiably neglecting the importance of 'context, judgment and understanding', insisting instead that 'understanding is *not* a subjectively appraised state any more than knowing is; both are objectively testable and are, in fact, tested in examinations' [2, pp. 196/176]. Others have sought to develop positive accounts of scientific understanding that would both respect its distinctiveness as an aim of science

A. Gelfert, *How to Do Science with Models*, SpringerBriefs in Philosophy,
DOI 10.1007/978-3-319-27954-1_4

and satisfy the demands of objectivity. The main two approaches along these lines are *unificationism*, pioneered by Michael Friedman and Philip Kitcher, and Wesley Salmon's *causal-mechanistic approach*. According to the former, explanations increase our understanding if they succeed in providing a unified account of a greater range of different phenomena; according to the latter, 'to understand *why* certain things happen, we need to see how they are produced by these mechanisms' [3, p. 132]. In recent years, it has become more acceptable to take scientific practice at face value, along with judgments and pronouncements concerning scientific understanding as an aim of science.[1] This turn towards the pragmatics of scientific understanding takes as its starting point the way scientists typically invoke the term 'scientific understanding' when they evaluate scientific models and theories, and thus renders its philosophical analysis closer to actual usage. However, as critics have urged, it is not always clear that the role of 'understanding' in science is substantively different from the more traditional business of devising scientific explanations. Perhaps what has been called 'scientific understanding' is merely a psychologically salient way of relating to scientific explanations, in the manner of an 'aha experience'. Thus, Michael Strevens has defended what he calls 'the *simple view*', according to which an individual 'has scientific understanding of a phenomenon just in case they grasp a correct scientific explanation of that phenomenon' [4, p. 510]. Others have doubted whether we should set much store by the mere feeling of understanding, even if we grant that such feelings may well be 'the phenomenological mark of the fulfilment of an evolutionarily determined drive' [5, p. 300] for seeking explanations. As Anna Alexandrova and Robert Northcott put it rather memorably: 'We know better than to look for orgasm to make sure that reproduction happened. Similarly, we should know better than to look for "aha" feelings to make sure that actual explanation happened' [6, p. 266]. These are valid concerns, and we will address the limitations—and dangers—of equating scientific understanding with a merely professed *sense of* understanding in the final section of this chapter. For the moment, given that our main interest is in scientific models (rather than in scientific understanding per se), an informal understanding of the term 'scientific understanding' will suffice for bringing out the multiple functions and uses of models in scientific inquiry.

Peter Godfrey-Smith notes that, where scientific modeling is pursued self-consciously—that is, in a manner characterized by 'its own skill-set, subculture, and language'—what one tends to find is 'scientific elaboration and formalization of a more general and psychologically deep capacity for model-based understanding' [7, p. 729]. Indeed, the thought that understanding in science is deeply connected with scientific models has a long ancestry, going back at least to William Thomson, who declared: 'It seems to me that the test of "Do we or do we not understand a particular subject in physics?" is, "Can we make a mechanical

[1]For recent collections of papers on the problem of scientific understanding, see [57, 60].

model of it?"'[2] [8, p. 111] In the case of model-based understanding in science, one needs to further distinguish between our understanding *of the model* (considered qua representational vehicle, e.g. as a set of mathematical equations) and our model-based understanding *of the target system*. Sang Wook Yi puts this nicely when he writes that understanding the model 'involves, among other things, exploring its potential explanatory power using various mathematical techniques, figuring out various plausible physical mechanisms for it and cultivating our physical intuitions', and that only '[a]fter we understand a model, we may employ the model to understand its target phenomena in the world' [9, p. 89]. It is clear from this characterization that understanding a model is not purely a matter of theoretical knowledge, but also requires skill—for example, in the case of mathematical models, the ability to derive new results or identify interesting limiting cases. This coheres well with recent analyses of scientific understanding more generally, which have begun to acknowledge the role of skill and tacit knowledge in generating understanding. For example, de Regt maintains that understanding a scientific theory T requires the ability to 'recognize the qualitative consequences of T without performing exact calculations' [10, p. 33].

This tacit dimension of model-based understanding is sometimes loosely described as developing 'a feel for' the model (and, by extension, for the behaviour of its target system). Drawing on Michael Polanyi's notion of *personal knowledge*, Theodore Kisiel gives an apt characterization of this locution, which enjoys considerable currency among scientists in general:

> In the vernacular, it is a matter of "getting a feel for" nature in the way science currently comes in contact with it. This tacit knowledge can only be conveyed by practice and from practicing scientists, through whom the novice assimilates the subliminal premises of his science. These premises weave the framework within which all of his scientific assertions are made, and yet, for this very reason, they themselves cannot be asserted. [11, p. 270]

When it comes to models, this language is often employed in connection with instances of manipulation or simulation. As Manfred Stöckler puts it: '[S]imulations help to develop a "feeling" for the decisive features of higher-level description' [12, p. 365]. By simulating a model's behaviour in time, one may get a feeling for how sensitive a model is to changes in the initial conditions; by varying other parameters, one may be able to get a sense of what kinds of real-world scenarios or phenomena a model can represent. Simulation, however, is but one way to develop 'a feel for' a model. Manipulation, too, is often a good way of deepening one's understanding of a model. Such manipulation may be physical—for example, when we try out possible molecular configurations using a material ball-and-stick model

[2] It is worth keeping in mind that the term 'model', as discussed in Chap. 1 (Sect. 1.2), used to have a more restrictive meaning, referring primarily to mechanical models, with other models commonly referred to as 'analogies'.

—or it may be symbolic, as when we explore chemical configurations using structural formulas.[3] As Mary Morgan puts it succinctly, 'representations only become models when they have the resources for manipulation' [13, p. 27]. What this suggests is that, beyond their explanatory, predictive, and instrumental performance in specific empirical contexts, models often possess significant internal resources that allow for a more exploratory mode of interacting with them.

4.2 On the Notion of 'Exploration'

Exploration, in common parlance, is first and foremost an activity: an 'explorer' is someone who sets out to traverse as yet uncharted territory or—in an age that no longer knows any blank spots on the map—seeks to navigate difficult terrain, for example by climbing a mountain. Historians used to speak of the 'Age of Exploration' and the 'Age of Discovery', using the terms virtually synonymously to refer to the era of European global exploration from the 15th century onwards. One might conclude from this usage that exploration is simply an activity that aims at the discovery of new facts, with the term 'exploration' designating a *behavioural* pattern and 'discovery' referring to a new *epistemic* advance in our cognitive state. Unlike the concept of scientific discovery, however, the various kinds of exploration that precede such discoveries have not, so far, received systematic attention from philosophers of science.

Unlike philosophers, psychologists—especially developmental psychologists, and those studying motivation, attention, and interest—along with scholars in education have attempted to devise a taxonomy of different types of exploratory behaviour. In his seminal book *Conflict, Arousal and Curiosity* (1960), Daniel Berlyne distinguished between *specific* and *diversive* exploration [14]. Specific exploration is a set of behaviours in response to a novel or unexpected stimulus; that is, it is stimulus-oriented. Some such behaviours may, of course, be reflexive—such as blinking and turning one's head in response to a sudden puff of air across the side of one's face. Other behaviours may include focusing one's attention on an incongruity in what one perceives, trying to resolve the incongruity by rehearsing learned behaviours, physically manipulating a new object, or engaging in more theoretical strategies of focused investigation. Though behavioural psychologists have tended to study specific exploration in terms of their subjects' response to environmental stimuli—where relevant attributes include incongruity, novelty, and change in the presentation of external objects—similar considerations apply to other problem situations. For example, we may also focus our attention on a salient theoretical question, explore ways of completing a mathematical proof, or attempt to resolve an ambiguity in meaning by trying out different interpretations.

[3]For a discussion of mature symbol systems as a form of '"cognitive scaffolding", which allow their users to "offload" or "externalize" cognitive load', see [33, p. 59].

Understood in this way, specific exploration *converges* upon a specific question, fact, detail, or 'missing link'. This contrasts with a *divergent* sense of 'exploration'—one that is not directed at a specific object, question, or stimulus, but is response-oriented, in that the cognitive subject seeks novelty or surprise for its own sake. Such diversive exploration aims less at finding answers and more at relieving boredom; ranging over an open-ended (thematic or spatial) domain, a subject engaging in diversive exploration may hit upon surprising findings, perhaps accidentally, but may quickly lose interest and move on—or, alternatively, may latch on to a narrower set of questions and switch to a mode of *specific* exploration.

Interestingly, this dual character of *exploration* is reflected in the etymology of the term. According to the *Oxford English Dictionary* [15, p. 575], the verb 'to explore', in its earliest usage dating back to the late 16th century, has the meaning 'to investigate, seek to ascertain or find out (a fact, the condition of anything)', and relatedly, 'to look into closely, examine into, scrutinize'. This closely mirrors the convergent character of *specific exploration*. From the 17th century onwards, however, 'to explore' has taken on a second meaning, namely 'to search into or examine (a country, a place, etc.) by going through it: to go into or range over for the purpose of discovery'—that is, a meaning much closer to *diversive exploration*. To be sure, exploration in this sense need not be solely driven by a desire for diversion, or relief of boredom, but it tends to be less focused and less constrained. Although there is a tension between the two senses of exploration—inasmuch as it is hard to see how one could simultaneously be focusing on a specific set of particulars *and* traversing a potentially open-ended domain—they may also be thought of as complementary. As we saw earlier, it is via diversive exploration that a subject may latch on to a specific set of questions and switch to a more specific mode of investigation. In the case of science, diversive exploration may be an excellent general tool for the generation of hypotheses, the more promising of which may then be investigated in more detail. The researcher may then switch to a mode of specific exploration by investigating the consequences of particular hypotheses, either theoretically or by conducting an experiment.[4]

As we shall see in the remainder of this chapter, experimentation and modeling both have important exploratory uses. Just as an experiment does not always serve the function of *testing* a theory, neither does a model always have to render an empirical phenomenon amenable to subsumption under a pre-existing theory. While traditional analyses of modeling may give us a good enough grasp of the various functions of models in situations where the underlying theory cannot be applied *directly*, an analysis of its exploratory uses is needed to account for situations where an underlying theory is *unavailable*, or where—as James Clerk Maxwell put it—it is essential 'to avoid the dangers arising from a premature theory' [16, p. 159].

[4]This is one of the reasons why experimentation and exploration are not mutually exclusive; drawing a contrast between 'the experimental' and 'the explorative' gives rise to a number of false dichotomies, of the sort found in [59, p. 191].

4.3 Exploration and Experimentation

The importance of exploration to science has recently been emphasized by a number of historians and philosophers of science writing on scientific experimentation. Based on historical case studies from 19th century research on electromagnetism, Friedrich Steinle has described *exploratory experimentation* as a research activity that is 'driven by the elementary desire to obtain empirical regularities and to find out proper concepts and classifications by means of which those regularities can be formulated'; such activity typically occurs in periods or problem situations where 'no well-formed theory or even no conceptual framework is available or regarded as reliable' [17, p. S70]. Exploratory experimentation, as Steinle sees it, is marked by an elaborate intertwining of experimental intervention and concept formation—that is, of action and meaning. Unlike in the case of testing —which has traditionally been regarded by philosophers as the main function of carrying out an experiment—exploratory experimentation aims not just at bringing about a well-defined observable change in the world, but also serves as a testing ground for new, yet to be stabilized concepts. As a result, 'the process of forming and stabilizing the [new] conceptual framework' [17, p. S72] may take considerable time and effort, and exploratory experimentation is seen as playing a key role in this process.

While Steinle's examples are mainly drawn from the physical sciences, other authors have come to similar conclusions on the basis of examples from the biological and social sciences. Richard Burian, in a paper published the same year as Steinle's, studied the exploratory character of the histochemical techniques employed in the work of the Belgian biochemist Jean Brachet (1909-1988); like Steinle, Burian notes the existence of a mode of experimentation in situations which, at a theoretical level, are at best 'partially understood'. The main goal, in such cases, is once again *stabilization*—in this case: 'stabilization of the protocols for locating particular molecular species and for identifying, and reidentifying the molecules thus localized', and rendering them 'relevant to the experimental and theoretical analyses of such other investigative traditions' [18, p. 42] as may be available, without subsuming them under any one theoretical framework in particular. Writing from the perspective of the social sciences, Uljana Feest emphasizes that the link between experimentation and concept formation works both ways: in disciplines like psychology, proposing operational definitions of concepts, however tentatively, may be a precondition for meaningful experimentation. As Feest notes, 'operational definitions function as tools for the generation of empirical evidence in a given domain, but they are themselves gradually refined, stabilized, and validated' [19, p. 185]. Concepts may thus themselves play an exploratory role in enabling the experimental study of empirical phenomena. In spite of such differences in emphasis, most contributors to the debate agree that exploratory experimentation, even though it is constituted by 'a bundle of different experimental strategies' [17, p. S73], is nonetheless a distinct mode of experimental research in the absence of fully-formed scientific theories about the domain in question. C. Kenneth Waters summarizes this nicely, when he writes that '*the aim of exploratory experiments is to generate significant findings about phenomena without appealing to a theory about these phenomena for the purpose of focusing experimental attention on a limited range of possible findings*' [20, p. 279] (italics original).

In order to make vivid just how exploratory experimentation differs from theory-driven experimentation, it may be instructive to consider a historical example. For this purpose, I shall draw on Steinle's discussion of how various researchers in the 1820s and 1830s made early forays into the study of electromagnetic phenomena, but will limit myself to his example of the work of Michael Faraday (1791–1867). Unlike many of the more prominent scientists at the time, Faraday did not have an academic education in the sciences, but was largely self-educated. Following Hans Christian Ørsted's discovery, in 1820, that a compass needle would be deflected from its position by a nearby electric current, many established scientists had only grudgingly come around to the idea that electricity and magnetism were not separate classes of phenomena. The reason for this reluctance, as André-Marie Ampère put it in a letter to Jacques Roux-Bordier in February 1821, was 'Coulomb's hypothesis on the magnetic effect; everybody trusted this hypothesis as if it were a fact; it denied any possibility of interaction between electricity and magnetism'.[5] Faraday, in an act of what Steinle calls 'tentative, but well-directed, theoretical speculation' [17, p. S68], reckoned that Ørsted's findings might be reversed: if electricity had been found to act on magnetism, perhaps magnetism in turn could act on electricity. Faraday attempted various experimental set-ups, and in doing so 'tried to facilitate and to enhance the expected effect, for example by winding up the wire into coils or by using soft iron', which by then was known to be highly effective in electromagnets. Finally, in August 1831, Faraday settled on the design of an induction ring:

> Two coils are wound on different sides of a massive iron ring. At the moment when one of them is connected to a battery, one can detect a short pulse current in the other, which is connected to the galvanometer. Since there is no electric conduction between the two coils, Faraday inferred that the effect was due to [magnetic] induction. [17, p. S68]

Interestingly, Faraday's design appears to be driven as much by the desire to demonstrate a (positive) link between electricity and magnetism as by the (negative) need to rule out electric conduction as a possible cause of any effects that might be observed in the second coil. Though, at the time, no theoretical framework was available for how one should interpret the observed effect of magnetism on electricity, the experiment was designed to demonstrate an expected effect. In this regard, the induction ring experiment contrasted with another of Faraday's experiments, conducted two months later, which he deemed 'truly elementary':

> A wire was bent into rectangular form. A sensitive galvanometer was integrated in one of the longer sides. During the experiment, this side remained fixed. Every time the other side

[5]Quoted after [48, p. 351]. The French original is: '…elle est dans l'hypothèse de Coulomb sur la nature de l'action magnétique; on croyait à cette hypothèse comme à un fait; elle écartait absolument toute idée d'action entre l'électricité et les prétendus fils magnétiques' [49].

> was moved such that it crossed the direction of the earth's magnetic dip, a current was induced in the circuit. [17, p. S68]

Unlike in the earlier case of the induction ring, which was designed to demonstrate a predicted—though, from the vantage point of physical theory, still problematic—effect, the second experiment came 'after intense experimental activity', during which 'Faraday systematically varied a lot of parameters of the arrangement such as the direction of motion (relative to the magnetic dip), the mode of motion (e.g., various parts of the circuit or the circuit in its entirety), the form of the circuit, and so on' [17, p. S68]. Such simultaneous variation of different parameters, following the realization that there was a causal link between magnetism and electricity, contrasts with the former case, given that the induction ring is 'a rather fixed device which does not allow many variations, either in the arrangement, or in the experimental outcome' [17, p. S69]. In this sense, the induction ring experiment—although temporally prior—may be regarded as less exploratory in character than the subsequent investigation that led to Faraday's 'truly elementary' experiment.[6]

While the absence of a fully-developed underlying theory is often crucial to the activity of exploratory experimentation, it would be misleading to think of exploration as entirely *theory-free*. For one, in devising experiments, one typically needs to be able to rely on significant background knowledge, including background theories. While these need not be overtly about the specific phenomena one is exploring, they may inform one's interpretation of the results and one's experimental design. Second, as Kevin Elliott has noted, theory often 'plays the role of a starting point or a "foil" in the exploratory process' [21, p. 327]—as it did in Faraday's case, when he took the potential reversibility of Ørsted's findings as his starting point. One way to think of the way in which expectations—guided by theoretical background knowledge, but without thereby aiming at the testing of scientific theories—may shape experimental activity is in terms of David Gooding's notion of *construals*. The term 'construal', according to Gooding, is meant to draw attention to its dependence 'on the context of action (in a way the word "interpretation" does not)'. On this view, construals 'are to the experience of *processes* what ostensive definition is to naming entitities and properties'; they order 'phenomena into an intelligible form' and 'enable an ascent from the immediate and concrete world', thereby creating 'communicable representations of new experience and at the same time integrat[ing] these into an existing system of experimental and linguistic practices'. Importantly, 'a construal cannot be grasped independently of the exploratory behaviour that produces it' [22, p. 87]. Not only does this speak to the significance of exploration as a precursor to theorizing, it also broadens the

[6] Both experiments were exploratory, insofar as no firm theoretical foundation was available at the time; therefore, an alternative interpretation might consider the first experiment—which aimed at demonstrating the existence of a causal link—an instance of 'convergent' exploration, whereas the second experiment, with its simultaneous variation of different parameters, might be considered as a case of 'divergent' exploration.

notion of exploration itself to include various strategies that may be deemed theory-laden. Steinle includes the following in his list of typical methodological 'guidelines' for exploratory experimentation:

- varying a large number of different experimental parameters,
- determining which of the different experimental conditions are indispensable, which are only modifying,
- looking for stable empirical rules,
- finding appropriate representations by means of which those rules can be formulated [...] [17, p. S70].

However, beyond the demand that exploratory experimentation must not be *driven* by theory, there is no need to exclude theoretical considerations altogether from this list. Rather, as Elliott puts it, it may be best to think of exploratory experimentation as 'an attempt to study a phenomenon using as many tools and techniques as possible so as to understand it more fully and to gain more solid epistemic access to it' [21, p. 328].

4.4 Exploratory Models

Once it is acknowledged that exploratory strategies are not limited to experimentation, but are a feature of scientific practice more generally, one can try to apply the idea of exploration to the case of scientific models. It is worth emphasizing, however, that extending the above discussion to scientific models is not as simple and straightforward as one might think. For one, not all strategies of exploration are equally interesting and informative. Consider the first in Steinle's list of strategies for exploratory experimentation: simultaneously varying a large number of different parameters. For the experimenter who intervenes in nature and explores the dynamics of the target system through causal means, variation of experimental parameters requires skill and, when successful, constitutes a great achievement. For models, this need not be the case: variation of parameters, for example in the case of mathematical models, may come too cheaply. When dealing with a set of polynomial equations, one can independently vary the coefficients in a largely arbitrary way—yet such a scanning of the parameter space is *exploratory* at best in a generic sense. It may have its place in scientific inquiry, in that it may enable preliminary curve-fitting (e.g. as a first step towards model-building), but it may not be the best—or even the most typical—way of generating understanding or granting more solid epistemic access to a target phenomenon. Variation of parameters in a model is by no means always trivial: not all models are mathematical in character, and varying the parameters of a material model—e.g., its size, geometry, density, or material constitution—is more akin to carrying out an experiment than it is to changing the value of mathematical coefficients. It is important, however, to remain alert to the differences between experimentation and modeling, and to survey those exploratory strategies that afford specific insight into the practice of scientific modeling.

Scientists themselves have occasionally commented on the exploratory character of much of what goes under the heading of scientific modeling, though they have typically done so in passing and without much concern for conceptual rigour. Thus, while the following examples are intended to convey a sense of why scientists turn to models as exploratory tools, this descriptive approach can only be a first step towards a theory of exploratory modeling; a more systematic account of the exploratory uses and functions of models will be given in the next section. As a first example from scientific practice, consider how John H. Holland, a computer scientist and pioneer in research on genetic algorithms, describes how exploratory models shed light on 'complex phenomena in terms of a limited set of mechanisms and constraints'. As he sees it, such models often suggest '"places to look" for salient phenomena, regularities hidden in complex data, etc.' and, over time, may 'take on aspects of an existence proof or predictive model' [23, p. 25]. Holland refers to a group of theoretical ecologists around Joan Roughgarden who have discussed the exploratory character of modeling under the label of 'minimal models for ideas'. As they see it, 'a minimal model for ideas is intended to explore a concept without reference to a particular species or place' [24, p. 26]. Most early models in theoretical ecology, they claim, were of this type. These models were not intended to be applied to specific target systems in the real world, let alone to make testable predictions. Examples include the Lotka-Volterra models of predator–prey dynamics and the logistic equation, which mimics, in a qualitative way, the speed-up and slow-down of population growth in an environment with limited resources. Roughgarden et al. [24] contrast the notion of a 'minimal model' with that of a 'synthetic model for a system', which synthesizes a great deal of ecological knowledge, including descriptive detail about the system's components, into a model that aims to be an empirically adequate representation of a specific target system.[7]

Interestingly, the term 'minimal model' has also caught on in the theory of phase transitions, where it is used to describe a model that, as the theoretical physicist Nigel Goldenfeld puts it, 'most economically caricatures the essential physics' [25, p. 33]. As in theoretical ecology, minimal models in physics are not intended to be faithful representations of any target system in particular, but are meant to allow for the exploration of universal features of a large class of systems; indeed, as Robert Batterman notes, '[t]he adding of details with the goal of "improving" the minimal model is self-defeating—such improvement is illusory' [26, p. 22]. In the social sciences, too, exploring general features and relationships may take precedence over representing specific target systems or phenomena—even if this is not widely recognized. As Daniel Hausman laments, 'few writers on economic methodology

[7]Whereas early minimal models tend to derive their dynamics from system-level equations, modern synthetic models in contemporary theoretical ecology tend to be 'bottom-up, representing many small spatial units or individuals and their behavior', with their 'system dynamics emerg[ing] from the interaction of the components' [36, p. 367].

recognize that the activities of formulating economic models and investigating their implications are a sort of conceptual exploration'; instead, 'most mistakenly regard these activities as offering empirical hypotheses' [27, p. 115].[8]

As these examples show, scientists are ready to acknowledge the exploratory role of modeling, yet this acknowledgment is not usually followed up by a more detailed analysis of exploratory strategies and their specific functions. Among philosophers of science, the overall picture is similarly sketchy. While scientific models and the activity of modeling are rarely explicitly described as 'exploratory', a number of authors have recognized functions of models that deserve this label, or have proposed accounts that can accommodate exploratory uses of scientific models. As an example of the former, consider William Wimsatt's defence of 'the deliberate use of false models as tools to better assess the true state of nature' [28, p. 132]. Several of the potential uses that Wimsatt identifies for false models may, in fact, be best accounted for by acknowledging their exploratory character:

> An oversimplified model may act as a starting point in a series of models of increasing complexity and realism. [...] A false model may suggest the form of a phenomenological relationship between the variables [... Such a model may serve] to generate new predictive tests of or to give new significance to features of an alternative preferred model. [28, pp. 104–105; 127]

By framing his discussion in terms of the truth or falsity of models, Wimsatt risks obscuring what it is that makes *some* false models more successful than others: namely their suitability for exploratory purposes.[9] That idealization and abstraction render most models literally false as representations of a specific target system is, after all, a widely acknowledged feature of scientific modeling; what calls out for an explanation, then, is the continued success of some—by representational standards: egregiously—false models, while many other ('truer') models fall out of favour. As an early example of a general account that can accommodate exploratory uses of scientific models, we can turn to Mary Hesse's analogical view of models (see Sect. 1.2). Recall that Hesse distinguishes between negative, neutral, and positive analogy, depending on whether the model's properties differ from, may turn out to be in common with, or are in fact shared with the target system. While the negative part is typically disregarded, it is precisely the neutral part of the analogy that allows us to arrive at novel insights about the target system by exploring what

---

[8]Hausman's account of modeling as a form of conceptual exploration, however, is weaker than what I have in mind: as Uskali Mäki observes, on Hausman's account, 'a model as such contains no truth claims about the world, it is rather a definition of a predicate given by the assumptions of the model' [56, p. 15]; only theoretical hypotheses about the *applicability* of model to particular situation are truth-valued. I agree with Mäki that this insulates models too much from potential challenges arising from the realist concern with truth.

[9]Michael Redhead seems to entertain a similar idea when he suggests that, '[b]y exploring models [...] for a theory T we can probe how approximations A_M to the model M [...] misrepresent M and the true behaviour of M as opposed to A_M can now be used as a guide how to T behaves' [54, p. 153].

follows from the model that isn't already included from the outset in the positive analogy.[10]

An interesting, and unusually detailed, take on exploratory modeling was provided by the ecologist Peter J. Taylor in a short article that focused on the 'different ways we *revise* models and thereby contribute to theory generation' [29, p. 122]. Taking mathematical models in biology as his starting point, Taylor argues that some theoretical tools function merely as a *schema*, in that they focus our attention on certain relevant processes or constraints—e.g., in the case of the logistic equation, population growth and limitations on available resources. However,

> if the schema can be expressed in a mathematical formulation, the model becomes what I call an *exploratory tool*. It can be explored systematically as a *mathematical* system, e.g. how does the system's behaviour change as its parameters change or become variables, as time lags are added, and so on? Such mathematical investigation may help us derive new questions to ask, new terms to employ, or different models to construct. [29, p. 122]

While such a model may occasionally fit the observations associated with a specific empirical phenomenon, such observational fit—at this stage of 'theory generation'—is neither necessary nor sufficient. As Taylor insists, only if '*independently of that fit* there is evidence for its accessory [background] conditions, then we are justified in acting as if the model represented the biological relations that generated the observations, i.e. in accepting the model as a *generative representation*' [29, p. 122]. Echoing Batterman's point regarding the futility of adding more empirical details to minimal models in physics, Taylor notes that revision of exploratory models 'is not necessarily directed at tightening the [empirical] fit of a model; it may run a gradient from attempting to expand acceptance of the model to attempting to disturb acceptance' [29, p. 123]. The exploratory use of models may thus itself become a tool of persuasion—which, as we shall see in the final section of this chapter, is not without its own problems.

While mathematics is not the only representational format that enables exploration, it is perhaps significant that a number of authors have independently argued that mathematics—whether considered in a wholesale manner as mathematics-at-large, or in the form of locally applicable (e.g. discipline-specific) 'mature mathematical formalisms' [30]—is especially suited for this purpose. Contrasting the use of schematic flow charts with more explicit mathematical models in a 1933 essay by the economist Ragnar Frisch (1895–1973), Morgan writes that, '[i]n the [visual] schema, there are resources that can be reasoned with, but they can not be manipulated in such a way' as to generate deeper understanding. Though rich in content, the format of visual flow charts inhibits systematic exploration of the model's consequences. By contrast, when the model is cast in mathematical form,

[10]On this point, see [58, p. 119]. While Bailer-Jones is one of the few philosophers of science to explicitly state that models provide 'material for exploration' (ibid.), she appears to regard exploration as mainly associated with an understanding of models as metaphors; this, it seems to me, does not do justice to the specific character of model-based exploration discussed in the present chapter.

'[t]he equations have less content in the sense that there are fewer elements and causal links, but the form (or language) of that content (equations) enables the use of a deductive mode of manipulation so that Frisch can reason mathematically about the nature of the business cycle' [13, pp. 29–30]. Traditionally, one of the main advantages of mathematical reasoning in science has been thought to be the fact that it is truth-preserving; yet, in addition to this general desideratum, mathematics also affords exceptionally well-developed theoretical and symbolic means for manipulation and conceptual exploration. This allows for diverse ways of model-building, while at the same achieving 'integration of [the various] ingredients in such a way that the result—the model—meets certain a priori criteria of quality' [31, p. 94]. Mathematics, in this regard, is not merely an auxiliary tool for applying fundamental theories to specific empirical situations or phenomena. Rather, as Mary Hesse puts it, 'any particular piece of mathematics has its own ways of suggesting modification and generalisation' which, in turn, may take the place of the more traditional 'pointers towards further progress' (such as the easy visualizability of mechanical models). [32, p. 200] While mathematics thus affords considerable potential for exploration, whether or not its exploratory use in modeling will lead to overall scientific progress also depends on the presence, or absence, 'of a scientific community whose members are skilled in applying and modifying models and theories, and which, collectively, is able to arrive at determinations regarding the fruitfulness (or "progress") of new theoretical proposals' [33, p. 57].

4.5 The Uses and Functions of Exploratory Models

So far, our discussion of the exploratory potential of models has been entirely general, with examples cited mainly as precedents or for illustrative purposes. In the present section, I aim to give a more systematic account of the exploratory uses and functions of scientific models. In particular, I shall identify four distinct functions that models often serve in exploratory research: they may function as a *starting point* for future inquiry, feature in *proof-of-principle* demonstrations, generate *potential explanations* of observed (types of) phenomena, and may lead us to assessments of the *suitability of the target*. These functions are neither mutually exclusive, nor are they thought to exhaust the exploratory potential of models. They do, however, represent the spectrum of exploratory uses to which models may be put, which ranges from what one might call 'weak' uses—such as taking a model as a 'starting point', for want of a better alternative—to 'stronger' (e.g. explanatory) uses, which may result in greater understanding or lead to a reformulation of the initial research question. Let us discuss each of the four functions in turn.

4.5.1 Exploratory Models as Starting Points

In his list of potential uses of 'false models as means to truer theories', Wimsatt lists as the first example cases where a model 'may act as a starting point in a series of models of increasing complexity and realism' [28, p. 104]. Perhaps because he deems this the most basic among the twelve types of uses of false models he distinguishes, Wimsatt neither provides an example nor elaborates on how the falsity of a model affects its suitability as a starting point for a series of more realistic models. Perhaps the thought is simply that, in the absence of a firm grasp of the underlying theory or mechanism, model-based inquiry—if it is to get off the ground—has to begin *somewhere* and that, at such an early stage, questions concerning the model's truth or empirical adequacy would be premature. If this is the case, however, then it would perhaps be more accurate to say that a model's suitability for *exploratory purposes* may, in the early stages of inquiry, outweigh its truth or empirical adequacy. In other words, whether a model is true or false—which, in the early stages of inquiry, may be impossible to judge, given the lack of a good theoretical measure—it may nonetheless serve as a tool for exploration *by providing a starting point for future inquiry*, which may then lead to increasingly realistic and sophisticated models. One might worry that this renders too many models 'exploratory', simply in virtue of their being used as a starting point of inquiry. But recall that, as in the case of experimentation, exploratory modeling is a distinct research activity only inasmuch as it takes place in the absence of a fully-formed theoretical framework, even while aiming at gaining more solid epistemic access to a phenomenon (or class of phenomena). Hence, a physicist who sets out to construct a model of the mathematical pendulum, based on the well-known theoretical principles of Newtonian physics, would not (normally) be considered as engaging in exploratory modeling. Likewise, merely fitting a parametric equation or curve to observed data, or sifting through measurements by means of sophisticated mathematical techniques such as nonlinear system identification, need not be considered exploratory—unless it is matched by an equal effort to improve our understanding of the system in question, i.e. by an attempt to peer into the 'black box', as it were. Exploratory modeling neither simply applies fundamental theory nor limits itself to aggregating (or accommodating) observational data, but typically involves a constructive effort at model-building.

As an example of exploratory modeling that serves as a starting point in a series of increasingly realistic models, consider car-following models of traffic flow.[11] Whereas early models of traffic flow were sometimes inspired by analogies with fluid dynamics in physics, there exists no readily available underlying fundamental theory of how human-operated vehicles behave at a collective level. While some basic constraints, such as the definition of a vehicle n's velocity as $v_n(t) = dx_n(t)/dt$, carry over from physics, other assumptions familiar from physics, such as uniform acceleration, do not apply. Acceleration, in particular, will

[11]For a detailed account of the evolution and timeline of traffic flow models, see [50].

depend on a variety of factors, such as roadway conditions, the drivers' preferred maximum speed, their tendency to 'close the gap' with the car in front, their reaction time, and so forth. This gives rise to a great latitude in potential ways of modeling traffic flow, both in terms of relevant factors and in terms of mathematical methods. For example, should the model be time-continuous or should it involve discrete time steps? This might depend on whether one sees traffic flow primarily as vehicular motion or as a series of decisions by the driver to brake or accelerate. It is no surprise, then, that the history of early car-following models is marked by a great deal of exploration of relevant factors and relationships. Louis Pipes, in 1953, was the first to model the position x_{n-1} of the leader in terms of the position of its follower n and his safe following distance: $x_{n-1} = x_n + d + Tv_n + l_{n-1}^{veh}$, with d the distance between the two vehicles at rest, l_{n-1}^{veh} the length of the car in front, and Tv_n the (velocity-dependent) 'legal distance'. The model captures the idea that traffic flow is determined by the dynamic between each vehicle and its follower but, apart from a few obvious geometrical constraints (e.g. length of the vehicles), makes minimal theoretical assumptions. Subsequent models went on to elaborate, and modify, various components of Pipes's elementary model. Thus, Eiji Kometani and Tsuna Sasaki (1961) introduced a time lag τ to reflect the reaction time needed for the follower to adjust his speed in response to the vehicle in front. Finally, following a series of modifications to render the basic model more realistic, Peter Gipps refined car-following models by assuming that each driver 'travels as fast as safety and the limitations of the vehicle permit' [34, p. 108] and defining various coefficients accordingly. While there has been much theoretical refinement in the study of traffic flows in the meantime, along with a proliferation of mathematical and numerical approaches, the early car-following models are a good example of how exploratory models can, in a non-trivial way, serve as a starting point in a series of models marked by increasing realism and sophistication.

4.5.2 Exploratory Models and Proof-of-Principle Demonstrations

A second important exploratory use of models concerns their deployment in proof-of-principle demonstrations. The expressions 'proof of principle' or 'proof of concept' originate from the engineering sciences, where they refer to prototypes or designs that showcase the feasibility, in principle, of a manufacturing process, mechanism, or method, according to certain stated criteria of success. In an analogous way, *exploratory models often feature in proof-of-principle demonstrations*. There are at least two ways in which the notion of 'proof of principle' may be understood in this context. First, a model may establish, by way of its own example, that a certain type of approach or methodology is able to generate potential representations of a target phenomenon. Second, a model may propose a specific mechanism or process that, once it has been modelled and its consequences have been explored, is found—'within the world of the model' [35, p. 33], to borrow a phrase from Margaret Morrison and Mary Morgan—to exhibit the kind of behaviour associated with the phenomenon, or class of phenomena, to be explained.

A good illustration of both of these aspects of exploratory models as 'proofs of principle' is the Lotka-Volterra model, already discussed in Chap. 3 (Sect. 3.4), which, as mentioned in the previous section, has on occasion been labelled 'exploratory' by ecologists themselves. Recall that the Lotka-Volterra model of predator–prey dynamics grew out of Umberto d'Ancona's empirical observation that reduced fishing activity led to an increase in the population of predator species, at the expense of food fish. The model developed by Vito Volterra consists of a pair of first-order, non-linear, differential equations which are interpreted as describing the mutual dependencies in a two-species system, with one species feeding on the other. Whereas the birth rate of the predator species depends on an ample supply of prey, any increase in the number of predators will increase the prey's death rate. The dynamic behaviour of the model resembles that found in actual predator–prey systems: both populations oscillate, with the number of prey overshooting first, leading to an increase in the number of predators, before both populations decline again. While the model affords qualitative insight into how the mutual dependencies between different species can affect the dynamics of their populations, neither Volterra nor d'Ancona claimed to be able to predict the future percentages of predator and prey species on its basis. The significance of the Lotka-Volterra equations and similar exploratory models does not lie in their presumed ability to make empirically adequate predictions. In the words of two contemporary ecologists, '[t]hese models are not intended to make testable predictions, or even to be applied to specific real systems' [36, p. 367]. Instead, what makes the Lotka-Volterra model significant is its character as a two-fold proof of principle: first, it demonstrates that it is possible to generate insight about *discrete* populations —whose size is measured in integer values—using *continuous* differential equations. Second, and more importantly, it demonstrates that periodic oscillations in the size of predator–prey populations may arise *purely internally*, without any external forcings.[12] The lasting value of the Lotka-Volterra model, then, consists not in its suitability as a tool for the numerical prediction of fish stocks, but in its character as a *proof of principle*, opening up new ways of mathematically modeling the dynamics of populations.

[12]The Lotka-Volterra equations do, in fact, permit for an equilibrium solution; however, the equilibrium point is unstable, so that any perturbation—however small—suffices to trigger the oscillatory behaviour.

4.5.3 Exploratory Models and Potential Explanations

Much of the value of scientific models, in general, comes from their explanatory role: many models, especially those that explicitly represent mechanisms and causes, purport to tell us *why*, and *how*, things happen. Nancy Cartwright has proposed an account that places models at the heart of what constitutes a scientific explanation. On her *simulacrum account* of explanation, 'to explain a phenomenon is to construct a model which fits the phenomenon into a theory' [37, p. 17]. While the laws of the underlying fundamental theory are taken to be true of the objects in the model and are used to derive specific accounts of their behaviour, the question of whether or not the various elements in the model have counterparts in reality is left open; at most, we are committed to saying that they have 'the right sort of appearance'. Yet, if to explain a phenomenon is to construct a model by which it can be subsumed under a theory, the question remains what to do when no fundamental theory is available. This is where exploratory models have an important role to play: they help us devise *potential explanations*, for example by envisaging scenarios that, if true, would give rise to the kinds of phenomena that constitute the explanandum. While this use of exploratory models is closely related to the proof-of-principle demonstrations discussed above—especially when it concerns 'how-possibly' questions—it can also be more theoretically ambitious, either by drawing more explicitly on existing theories in the vicinity of the domain of investigation, or by aiming for the development of new theoretical frameworks. In what follows, I shall briefly discuss two such examples, one from 19th century physics, the other from early 20th century physical organic chemistry.

In Sect. 4.3, we encountered Faraday's experimental studies of electromagnetism as paradigmatic cases of what Steinle calls *exploratory experimentation*. Interestingly, some of the earliest attempts to develop theoretical frameworks for the newly discovered electromagnetic phenomena proceeded via *exploratory modeling*. This is especially evident in the work of James Clerk Maxwell (1831–1879). In his 1856 paper 'On Faraday's Lines of Force', Maxwell had deployed the mechanical analogy of flow of an incompressible fluid in order to explore the physical geometry of *lines of force*, which were imagined to be pervading space, even in the absence of any objects on which such a force could act. In pursuing the analogy with mechanics, Maxwell did not wish to suggest that electromagnetic phenomena were mechanical in nature—he insisted that the imagined substance was 'not even a hypothetical fluid' [16, p. 160]—but instead attempted 'to bring before the mind, in a convenient and manageable form, those mathematical ideas which are necessary to the study of the phenomena of electricity' [16, p. 157]. Why give the mathematical equations a physical interpretation in the first place? At the beginning of his paper, Maxwell lays out the shortcomings of both a *purely* mathematical and a theory-driven approach:

> In the first case we entirely lose sight of the phenomena to be explained; [in the second case] we see the phenomena only through a medium, and are liable to that blindness to facts and rashness in assumption which a partial explanation encourages. [16, pp. 155–156]

Given that no existing theory could satisfactorily explain the new phenomena, what was needed was an approach that would lay out clearly and systematically, but without recourse to 'physical hypothesis', the relationship that a future theory would have to explain:

> We must therefore discover some method of investigation which allows the mind at every step to lay hold of a clear physical conception, without being committed to any theory founded on the physical science from which that conception is borrowed, so that it is neither drawn aside from the subject in pursuit of analytical subtleties, nor carried beyond the truth by a favourite hypothesis. [16, p. 156]

Whereas Maxwell, in his 1856 paper, insists that he is 'not attempting to establish any physical theory' [16, p. 157], in his 1861 paper 'On Physical Lines of Force', his exploratory work is more overtly directed at offering explanations and developing a new theoretical framework. This is evident from his assessment of an 1847 paper by William Thomson which, similar to Maxwell's earlier work, aimed at illustrating the new phenomena by means of 'mechanical representations': 'The author of this method of representation', Maxwell writes about Thomson, 'does not attempt to explain the origin of the observed forces […], but makes use of the mathematical analogies of the two problems to assist the imagination in the study of both' [16, p. 453]. Maxwell's project, by then, had grown more ambitious, culminating in the development of his famous mechanical ether model, which visualized the lines of magnetic force around a magnet as though they were vortices within a continuous fluid (see Fig. 4.1).

The ether model combined easy visualizability with the conceptual clarity of the mechanical analogies employed, while staying largely clear of substantive ontological commitments with respect to the precise nature of the model's constituents. This rendered it an excellent tool for the exploration of the nascent theory of molecular vortices, and Maxwell guides his readers by giving them detailed instructions regarding how to interpret, and bring to life, the model of which they

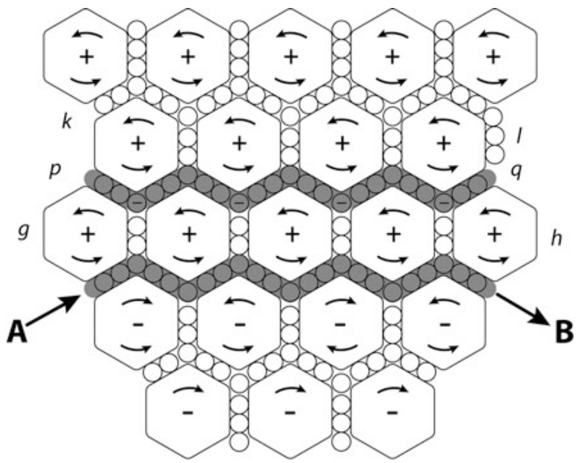

Fig. 4.1 Maxwell's vortex model, representing the flow of electricity from A to B; the large cells are the vortices, the smaller circles the 'idle wheels'

are only given a momentary 'snapshot': '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 row of vortices *kl* still at rest, then the layer of particles between these rows will be acted on by the row *gh*' [16, p. 477], and so forth. At the same time, Maxwell is explicit about the exploratory role of the model and its interim status, while a more fundamental theory of electromagnetism is still being pursued: as he sees it, anyone who 'understands the provisional and temporary character' of the mechanical ether model 'will find himself rather helped than hindered by it in his search after the true interpretation of the phenomena' [16, p. 486]. In his *Treatise* of 1873, reflecting on his earlier studies, Maxwell again acknowledges their provisional and exploratory character: 'The attempt which I then made to imagine a working model of this mechanism must be taken for no more than it really is, a demonstration that mechanism may be imagined capable of producing a connexion mechanically equivalent to the actual connexion of the parts of the electromagnetic field' [38, pp. 416–417]. Yet, *as a demonstration* of this kind, his exploratory model had proved unusually fruitful and explanatorily successful: not only was it able to account for light as an electromagnetic phenomenon, following the realization that the velocity of transversal waves in the stipulated medium would correspond to the ratio of the electric and the magnetic units (which, in turn, had been found to have the dimension of a velocity and a numerical value in the same range as the experimentally established value of the speed of light); it also helped explain other phenomena, such as the effect of magneto-optic rotation.

Let us turn to the second, more recent example of how exploratory models may provide potential explanations, which is drawn from early 20th century physical organic chemistry.[13] The late 19th century had seen a rapid growth of knowledge in physical and organic chemistry, accompanied by a growing recognition of the structural complexity of organic compounds. Early researchers, such as August Laurant in 1854, had assumed that transformations of organic compounds were subject to a *principle of minimal structural change*, which posited that changes would largely be limited to peripheral atoms. However, it quickly became clear that there were other more complex molecular rearrangements, many of which involved the migration of a whole group, e.g. a methyl group ($–CH_3$) or allyl group ($H_2C=CH–CH_2–$), from one carbon atom to another. The two-fold challenge of the nascent discipline of physical organic chemistry at the beginning of the 20th century therefore consisted in classifying the various types of structural transformations and explaining their occurrence by identifying the underlying molecular mechanisms. It is perhaps worth recalling that such discussions took place against the backdrop of an only gradually emerging qualitative understanding of the chemical bond in general. Gilbert N. Lewis and Irving Langmuir, during and after World War I, had developed the idea that the chemical bond involved pairs of electrons,

[13] My discussion of this example draws heavily on [40].

which would either be shared between two atoms ('covalent bond') or would result from one atom losing electrons to the other ('ionic bond').

Elements of Lewis and Langmuir's theory inspired other theoretical proposals, specifically in relation to reaction mechanisms for organic compounds. Thus Christopher Ingold, a former adherent of the (non-electronic) affinity view of chemical reactions, and his collaborator Edward Hughes developed an account of how intramolecular forces could give rise to molecular structures such as the benzene ring and their reconfigurations. As Ingold put it in a lecture in 1946: 'The idea of intramolecular electric interaction is the central theme of all our theories of reactivity.'[14] An important theoretical ingredient in the Hughes-Ingold account was the thought that organic reactions required, as part of their mechanism, the spontaneous emergence—through a posited 'inductive effect'—of highly polarized reaction sites, for which electrons either needed to 'be drawn away from [...] or be driven into, the rest of the molecule' [39, pp. 49–50]. Depending on direction and stipulated causes, these processes were to be described in terms of an elaborate terminology, which classified them as 'mesomeric', 'tautomeric', 'inductomeric', and 'synartetic', among others.[15] Though both accounts, the Lewis-Langmuir theory and the Hughes-Ingold theory, developed creative ways of conceptualizing molecular processes, they were largely speculative, given that intramolecular forces could not be observed in real time. Indeed, in a textbook two decades later, in 1953, Ingold [39] acknowledged, in relation to the rearrangement issue, that, 'if a cyclic transition state is internally set up, there is no conceivable way in which we could tell which way round the electrons moved in order to form it, or even whether they were displaced heterolytically in pairs, or in homolytic fashion by the breaking and constitution of pairs'.[16]

At the time, in the 1940s, the speculative character of the competing theoretical frameworks, and their evident incompleteness, did not stop researchers from attempting to extend it to ever more complicated molecular rearrangements. Several complex rearrangements had been known for some time, predating the emergence of the Lewis-Langmuir and the Hughes-Ingold theories. One such example is the Claisen rearrangement, named after Rainer Ludwig Claisen (1851–1930), which concerns isomeric changes in allyl ethers of enols and phenols, and essentially involves the shift of the allyl group ($H_2C{=}CH{-}CH_2{-}$) to the ortho-position (relative to the oxygen atom; see Fig. 4.2):

Claisen studied this type of rearrangement experimentally and found that the C=O=C group was essential; allyl ethers lacking it failed to rearrange. Together with his collaborator E. Tietze, he developed what was in essence a non-electronic mechanism in terms of the spatial proximity of the atoms involved, which allowed for the temporary formation of a cyclic arrangement of bonds linking the allyl group to the carbon atom in the ortho-position (see Fig. 4.3). Adherents of the emerging

[14] Quoted after [44, p. 222].

[15] For a discussion of Ingold's system of classifications, see [45].

[16] Quoted after [40, p. 567].

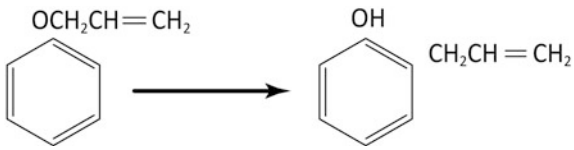

Fig. 4.2 The ortho-Claisen rearrangement of an allyl phenyl ether

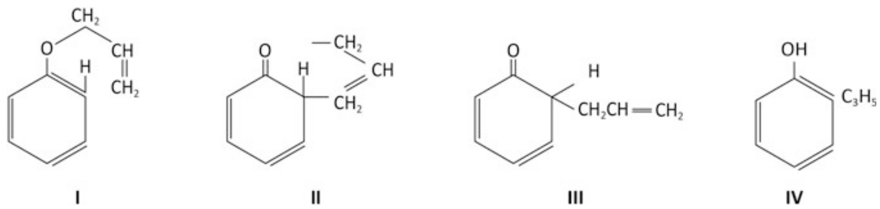

Fig. 4.3 The Claisen and Tietze model of the Claisen rearrangement

electronic paradigm of bond formation criticized this explanation for its spatial-mechanical emphasis on 'close contact' between the different constituents; as Stanley Tarbell wrote in a review in 1940, the two authors had given 'primarily a description of the rearrangement', when what was needed was an explanation 'in electronic terms'.[17]

Interestingly, however, when it came to explaining molecular rearrangements in electronic terms, even the proponents of the new theories continued to rely on the (non-electronic) considerations employed by their predecessors. Grant Fisher, in a case study of early 20th century models of molecular rearrangements, discusses the example of Arthur Cope and his student, Elizabeth Hardy, who set out, in 1940, to apply the Hughes-Ingold theory to a new rearrangement, which has since come to be known as the Cope rearrangement. Deploying the Hughes-Ingold terminology of the 'electromeric effect' and the 'inductive effect', they argued that such effects were the driving forces behind the allyl group's shifting from the electron-attracting α-carbon atom to the less electron-attracting γ-carbon atom, accompanied by the formation of a carbonyl group, via the migration of a double bond (see Fig. 4.4).

In order to explicate the (hypothetical) steps of how such a rearrangement could occur, Cope and Hardy modeled their mechanisms for the Cope rearrangements closely after Claisen and Tietze's earlier (non-electronic) model. Yet, as Fisher notes, 'what makes Cope and Hardy's approach puzzling is that before 1940 it had already been made clear that the Hughes-Ingold theory in particular was *not* a suitable candidate for explaining these rearrangements' [40, p. 575]. In other words, even though Cope and Hardy were committed to the *electronic* nature of the Cope rearrangement and proposed their model explicitly as an *application* of the Hughes-Ingold theory, their model nonetheless remained fundamentally indebted to

[17]Quoted after [40, p. 571].

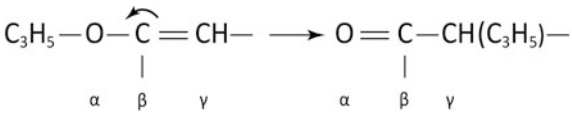

Fig. 4.4 Cope and Hardy's proposed mechanism of the Claisen rearrangement

the earlier *non-electronic* considerations by Claisen and Tietze, and proceeded *in the absence of* a unified electronic theory of molecular rearrangements.

What are we to make of this puzzling story of the Cope-Hardy model of molecular rearrangement? First, the case nicely illustrates a key ingredient of Morgan and Morrison's 'models-as-mediators' approach, according to which models 'are *not* situated in the middle of an hierarchical structure between theory and the world', but 'are made up from a mixture of elements', which often includes resources from outside the 'theory–world axis' [35, pp. 17f.; p. 23). (For a more detailed discussion of the 'models-as-mediators' approach, see Chap. 5, Sect. 5.1.) As Fisher observes, 'Cope and Hardy's use of analogies with the Claisen rearrangement is suggestive of the use of resources in model construction that lie outside of the theory-world axis' [40, p. 575]. Second, the Cope-Hardy model was clearly intended to be of explanatory value: 'The allyl group shift with inversion was what required explanation, and while it may not have been impossible given chemists' background assumptions, it was certainly puzzling, perhaps even baffling, to physical organic chemists' [40, p. 578]. However, unlike traditional explanations in terms of laws of nature, which aim to show how—given suitable background conditions—an observed phenomenon had to occur *necessarily*, the Cope-Hardy model provided an answer to a 'how-possibly'[18] question: while the Hughes-Ingold theory did not explain why, *necessarily*, the allyl group and double bonds shifted the way they did, it did not rule out the possibility either, and Cope and Hardy, by drawing on the analogy with the Claisen rearrangement, were able to suggest a way in which the rearrangement *could* have happened. Fisher acknowledges the exploratory character of devising *potential explanations* via *possible mechanisms* when he writes that 'Cope and Hardy built models […] that explored or tested out the Hughes-Ingold theory by generating how-possibly explanations' [40, p. 578]. However, where Fisher sees Cope and Hardy's models as 'resources for the exploration *of theory*' [40, p. 582; emphasis added], it may be more economical—not least in light of the failure of theory, in this case, to provide any independent guidance, at least not without relying on the earlier arguments by Claisen and Tietze—to deem exploration *simpliciter* the model's primary function, without making it

[18]The notion of a 'how-possibly' question goes back to the philosopher of history William Herbert Dray who had argued that 'the demand for explanation is, in some contexts, satisfactorily met if what happened is merely shown to have been *possible*; there is no need to go on to show that it was necessary as well' [46, p. 157] That is, how-possibly explanations are offered as genuine and complete explanations of particular phenomena without pretending to subsume them under general laws or generalizations (see [55] for a more recent discussion).

subservient to theory. Either way, the case of model-building in early physical organic chemistry serves as a good example of exploratory modeling that aims at devising *potential explanations* for phenomena that cannot readily be understood on the basis of existing (and often competing) theories of the domain in question.

4.5.4 Exploring the Suitability of the Target

Another application of exploratory modeling is for the purpose of assessing the suitability *of the target*. At first this may seem surprising, or even misguided: after all, shouldn't our models fit the world, rather than the other way round? When we have a good grasp of what constitutes the target phenomenon—even while we may not (yet) be able to explain it—tailoring the model to the target may be a promising strategy. But often in the context of exploratory research—that is: in the absence of comprehensive theoretical knowledge—determining where the target system begins and where it ends, reliably picking it out from background noise, and arriving at a stable 'research object' are not at all straightforward tasks. This is why scientists modeling empirical phenomena often spend considerable time and effort identifying appropriate initial and boundary conditions, which are required to render the fundamental laws, typically expressed in the form of differential equations, applicable to the case at hand.[19] In order to delineate a target phenomenon and converge upon a set of relevant properties and relations (which may subsequently come to define the target system or phenomenon), we must operate with some preconception of which factors are significant or salient; in the first instance, this will plausibly require grasping—tentatively, through an exercise of a well-trained theoretical imagination—certain facts about the way things present themselves to us. At an early stage of inquiry, before the stability of the target phenomenon has been ascertained, our conception of the target phenomenon will necessarily be subject to revision.[20] It is in such situations that exploratory modeling may lead to a reconsideration of the target system.

Taking the logistic equation (see Sect. 4.4) as his example, Taylor credits exploratory modeling with reminding us that it is not always clear 'that the quickest route to better generative representations relies on every new idea being framed as a hypothesis and directly tested'. Instead of treating the logistic equation as a tool for predictions, which may then be tested against empirical observations, 'we could move back to the level where the logistic is an exploratory tool and examine the effects of the model population being genetically heterogeneous or spatially

[19]Something like this seems to be Jordi Cat's point in [51], though he appears to think that scientists and philosophers alike have tended to overlook the importance of initial and boundary conditions.

[20]The difficulty of stabilizing phenomena in the absence of agreed-upon criteria for what counts as a successful experiment is known as the 'experimenter's regress'; for a discussion of its analogue in the case of scientific models and simulations ('simulationist's regress'), see [52].

distributed', or include new background assumptions: 'Out of such explorations might emerge ideas about the conditions under which the logistic might work as a generative representation' [29, p. 123]. Similar considerations apply in the engineering sciences where, as Susan Sterrett notes, modeling often leads to revisions of the original problem, sometimes leading to a reconsideration of the target system:

> But, occasionally, the person carrying out the model-making also gains some insight about the model, the target, and the behavior of interest that makes him or her question the suitability, not of the model, but of the target. The model-making points out a certain feature of the target, and questioning the suitability of the target for solving the larger engineering problem that occasioned interest in pursuing the problem in the first place can lead to a complete overhaul of the conception of the problem. [41, p. 36]

What this suggests, in line with general analyses of scientific practice, is that the traditional picture of modeling as a unidirectional activity—either leading from theory to phenomena, via simplifications and idealizations, or the other way round, by aggregating empirical data into a format that can be subsumed under theory—is inadequate; instead, modeling is a complex process of integration and exploration. Modeling need not always come *after* a fundamental theory has been established or an empirical phenomenon has been stabilized; as the examples above show, sometimes scientists devise models *in search of* an empirical phenomenon. Philosophers of science, in recent years, have become more careful to distinguish between modeling as an activity and scientific models as the products of that activity. This has led to a recognition that not all modeling is immediately tied to a specific target system or theory. As Arnon Levy writes:

> When a model is proposed it might not be clear at first what target it is tied to, and there might be a period in which the right target is sought. But later, assuming the model is retained, this issue is usually clarified. [42, p. 796]

It would, however, be unwise to view this indeterminacy with respect to a model's target as simply a shortcoming of early-stage modeling, standing in urgent need of clarification. Rather, assessing the suitability of a proposed target is precisely one of the functions of exploratory modeling, and it is as much its purpose to enable the zeroing in on a set of known phenomena as it is to open new lines of inquiry.

4.6 Exploratory Modeling: Prospects and Caveats

This chapter has discussed a variety of closely related uses of models which, I have argued, are best described collectively as *exploratory modeling*. As in the case of exploratory experimentation, applying the label 'exploratory' to models is intended to capture a particular mode of doing science; the same model—be it a physical model, or a set of equations—may be exploratory in one context, but not in another. Yet, as the discussion in the previous section has aimed to show, whether or not an instance of modeling is exploratory is not just in the eye of the beholder:

exploratory modeling serves specific, identifiable purposes, some of which—e.g., the generation of potential explanations—are closely aligned with the traditional goals of science, such as explanation and prediction. Others, however, go beyond the standard dimensions along which scientific activity is typically evaluated. In particular, as we have seen, exploratory modeling often serves the purpose of developing a grasp of (as yet theoretically inaccessible) phenomena and may lead to serious reconsideration of the suitability *of the target*. Such uses might previously have been brushed off as belonging to the 'context of discovery', yet looking at them through the lens of 'exploration' reveals their continuity with other, more established uses and functions of models. While the four typical functions of exploration listed in the previous section are meant to be neither exhaustive nor mutually exclusive, my contention is that they can be found across different disciplines—as indeed is also suggested by the examples I have discussed, which were drawn from physics, chemistry, biology, and the study of collective social phenomena. Beyond its uses in individual scientific disciplines, exploratory modeling is also relevant to broader questions in the history and philosophy of science. While acknowledging the exploratory character of much of scientific modeling holds out the promise of a richer and more complete understanding of scientific practice, one needs to be equally aware of the prospects of exploratory modeling and some important caveats. By way of conclusion, I wish to highlight only two such aspects, one of which illustrates how an acknowledgment of exploratory modeling may help improve our historical understanding of science, while the other speaks to a legitimate worry concerning the limitations of exploratory modeling.

Consider the case of the Ising model. Originally proposed in 1925 by the German physicist Ernst Ising (1900–1998) at the suggestion of his then supervisor Wilhelm Lenz (1888–1957), it was published under the modest title 'A Contribution to the Theory of Ferromagnetism' and its conclusions were negative throughout. According to a summary in that year's volume of *Science Abstracts*, the model is

> an attempt to modify Weiss' theory of ferromagnetism by consideration of the thermal behavior of a linear distribution of elementary magnets which … [have] only a non-magnetic action between neighboring elements. It is shown that such a model possesses no ferromagnetic properties, a conclusion extending to a three-dimensional field.[21]

The basic idea of the model was simple, in that it pictured a magnet as a collection of classical particles, each behaving as a tiny elementary magnet by pointing in one of two directions. If all the elementary magnets were aligned, the (macroscopic) system would display perfect magnetization, whereas in the case of total thermal disorder, no net magnetization would result. Ising's original hope was that a model along those lines would be able to explain magnetic phase transitions, i.e. the spontaneous emergence of a net magnetization below a certain critical temperature. The failure of the model to mimic such behaviour for any finite temperature is but one aspect of the story. For, as Niels Bohr (1911) and Hendrika

[21]Quoted after [47, p. 140].

Johanna van Leeuwen (1919) had rigorously shown, prior to the Ising model's development, no purely classical system—i.e. no system of classical particles obeying the laws of electrodynamics—could *ever* display spontaneous magnetization (though, of course, it might become magnetized in response to an external field). As a model of spontaneous ferromagnetism, the Ising model was therefore doomed to failure. Interestingly, however, this did not spell the end of the Ising model, either in physics or beyond. Indeed, contemporary applications of the Ising model range from the magnetic behaviour in linear polymer chains to belief polarization in social systems, and the swimming patterns of schools of fish. In most of these cases, the Ising model is not expected to make precise empirical predictions; rather, its flexibility and internal mathematical structure is used as a resource for the systematic exploration of disparate phenomena.[22] Without an acknowledgment of the exploratory character of much of what passes as scientific modeling, it would be hard to see how one might explain the longevity of a prima facie *unsuccessful* model such as the Ising model: it would remain an inexplicable historical anomaly.

Exploration in general is not without risk, though, and in the case of exploratory modeling, too, success is neither guaranteed nor without dangers. Precisely because exploratory modeling precedes the full articulation of an underlying fundamental theory, standards for judging an instance of exploration successful tend to be implicit, often depending on a tacit component. As discussed in Sect. 4.1, scientists themselves often use the phrase 'getting a feel for [a model or phenomenon]'. Models, in Morgan's turn of phrase, are 'resources for manipulation', and with an increase in skill and manipulative facility often comes an increased subjective sense of understanding. Mastery of mature mathematical formalisms and manipulative prowess tend to be markers of expertise and skill, yet they are not alone sufficient to guarantee the empirical validity of inferences made on their basis about the target system. Morgan's distinction between 'the world in the model' [13, p. 37]—for example the internal dynamics of a set of mathematical equations—and the outside world represented by the model comes to mind here. This is echoed by Taylor in his discussion of models in mathematical ecology: 'Strictly speaking, without a quantitative analysis of correspondence the insights from exploration are insights about a mathematical system' [29, p. 123]. Danger lurks when researchers mistake their facility at exploring the 'world in the model', e.g. via symbolic manipulation of mathematical equations, for an improved understanding *of the target system itself*. Indeed, as Jaakko Kuorikoski and Petri Ylikoski point out: 'Sometimes the ease of use increases the chances of the illusion of depth of understanding by making "toying with the model" too easy in such a way which crowds out thinking about the crucial background assumptions needed to reliably infer with the model to

[22]The same holds for other many-body models such as the Hubbard model, discussed in Chap. 3 (Sect. 3.3). As I have noted elsewhere, often 'the "exploratory" phase of understanding a proposed many-body model and cultivating intuitions about the interplay of themicroscopic mechanisms it is designed to represent is drawn out over many years; whether the model will in the end match an empirical phenomenon in many cases remains an open question' [53, p. 263].

a real world phenomenon' [43, p. 17]. While exploratory modeling may indeed be especially vulnerable in this respect, this only goes to show that exploration, if left unchecked by theory and observation, may go astray; it does not mean that exploratory research, whether in scientific experimentation or modeling, can be made redundant by reliance on theory and observation alone. Indeed, from the perspective of a functional account of scientific models, and in light of the wide range of examples discussed in this chapter, it seems far more compelling to acknowledge exploration as standing alongside explanation and prediction as one of the core functions of scientific modeling, and ineliminably so.

References

1. C.G. Hempel, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science* (Free Press, New York, 1965)
2. M. Scriven, Explanations, predictions, and laws, in *Scientific Explanation, Space, and Time*, ed. by H. Feigl, G. Maxwell (University of Minnesota Press, Minneapolis, 1962), pp. 170–230
3. W. Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton University Press, Princeton, 1984)
4. M. Strevens, No understanding without explanation. Stud. Hist. Philos. Sci. **44**(3), 510–515 (2013)
5. A. Gopnik, Explanation as orgasm and the drive for causal knowledge: the function, evolution, and phenomenology of the theory formation system, in *Cognition and Explanation*, ed. by F.C. Keil, R.A. Wilson (MIT Press, Cambridge, Mass., 2000), pp. 299–323
6. A. Alexandra, R. Northcott, It's just a feeling: why economic models do not explain. J. Econ. Method. **20**(3), 262–267 (2013)
7. P. Godfrey-Smith, The strategy of model-based science. Biol. Philos. **21**(5), 725–740 (2006)
8. W. Thomson, *Notes on Lectures on Molecular Dynamics and the Wave Theory of Light* (Johns Hopkins University Press, Baltimore, 1884)
9. S.W. Yi, The nature of model-based understanding in condensed matter physics. Mind Soc. **3** (1), 81–91 (2002)
10. H.W. de Regt, Understanding and scientific explanation, in *Scientific Understanding: Philosophical Perspectives*, ed. by H.W. de Regt, S. Leonelli, K. Eigner (University of Pittsburgh Press, Pittsburgh, 2009), pp. 21–42
11. T. Kisiel, Scientific discovery: logical, psychological, or hermeneutical?, in *Explorations in Phenomenology: Papers of the Society for Phenomenology and Existential Philosophy*, ed. by D. Carr, E.S. Casey (Martinus Nijhoff, The Hague, 1973), pp. 263–284
12. M. Stöckler, On modeling and simulations as instruments for the study of complex systems, in *Science at Century's End: Philosophical Questions on the Progress and Limits of Science*, ed. by M. Carrier, G.J. Massey, L. Ruetsche (University of Pittsburgh Press, Pittsburgh, 1997), pp. 355–373
13. M.S. Morgan, *The World in the Model: How Economists Work and Think* (Cambridge University Press, Cambridge, 2012)
14. D.E. Berlyne, *Conflict, Arousal and Curiosity* (McGraw-Hill, New York, 1960)
15. J.A. Simpson, E.S. Weiner (eds.), *The Oxford English Dictionary*, vol. V (Oxford University Press, Oxford, 1989)
16. J.C. Maxwell, in *The Scientific Papers of James Clerk Maxwell*, vol. 1, ed. by W.D. Niven. (Cambridge University Press, Cambridge, 1890)
17. F. Steinle, Entering new fields: exploratory uses of experimentation. Philos. Sci. **64**, S65–S74 (1997). (Proceedings of the PSA1996, Pt. II)

18. R.M. Burian, Exploratory experimentation and the role of histochemical techniques in the work of Jean Brachet, 1938–1952. Hist. Philos. Life Sci. **19**(1), 27–45 (1997)
19. U. Feest, Exploratory experiments, concept formation, and theory construction in psychology, in *Scientific Concepts and Investigative Practice*, ed. by U. Feest, F. Steinle (de Gruyter, Berlin, 2012), pp. 167–189
20. C.K. Waters, The nature and context of exploratory experimentation: an introduction to three case studies of exploratory research. Hist. Philos. Life Sci. **29**(3), 275–284 (2007)
21. K.C. Elliott, Varieties of exploratory experimentation in nanotoxicology. Hist. Philos. Life Sci. **29**(3), 313–336 (2007)
22. D. Gooding, *Experiment and the Making of Meaning: Human Agency in Scientific Observation and Experiment* (Kluwer, Dordrecht, 1990)
23. J.H. Holland, Emergence. Philosophica **59**(1), 11–40 (1997)
24. J. Roughgarden, A. Bergman, S. Shafir, C. Taylor, Adaptive computation in ecology and evolution: a guide for future research, in *Adaptive Individuals in Evolving Populations: Models and Algorithms*, ed. by R.K. Belew, M. Mitchell (Addison-Wesley, Boston, 1996), pp. 25–30
25. N. Goldenfeld, *Lectures on Phase Transitions and the Renormalization Group* (Addison-Wesley, Boston, 1992)
26. R. Batterman, Asymptotics and the role of minimal models. Br. J. Philos. Sci. **53**(1), 21–38 (2002)
27. D. Hausman, Economic methodology in a nutshell. J. Econ. Perspect. **3**(2), 115–127 (1989)
28. W.C. Wimsatt, *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality* (Harvard University Press, Cambridge, Mass., 2007)
29. P. Taylor, Revising models and generating theory. Oikos **54**(1), 121–126 (1989)
30. A. Gelfert, Mathematical formalisms in scientific practice: from denotation to model-based representation. Stud. Hist. Philos. Sci. **42**(2), 272–286 (2011)
31. M. Boumans, Built-in justification, in *Models as mediators: perspectives on natural and social science*, ed. by M.S. Morgan, M. Morrison (Cambridge University Press, Cambridge, 1999), pp. 66–96
32. M. Hesse, Models in physics. Br. J. Philos. Sci. **4**(15), 198–214 (1953)
33. A. Gelfert, Symbol systems as collective representational resources: Mary Hesse, Nelson Goodman, and the problem of scientific representation. Social Epistemology Review and Reply Collective **4**(6), 52–61 (2015)
34. P.G. Gipps, A behavioural car-following model for computer simulation. Transport. Res. Part B: Method. **15**(2), 105–111 (1981)
35. M. Morrison, M.S. Morgan, Models as mediating instruments, in *Models as Mediators: Perspectives on Natural and Social Science*, ed. by M.S. Morgan, M. Morrison (Cambridge University Press, Cambridge, 1999), pp. 10–37
36. V. Grimm, S.F. Railsback, *Individual-Based Modeling and Ecology* (Princeton University Press, Princeton, 2005)
37. N. Cartwright, *How the Laws of Physics Lie* (Oxford University Press, Oxford, 1983)
38. J.C. Maxwell, *A Treatise on Electricity and Magnetism*, vol. 2 (Clarendon Press, Oxford, 1873)
39. C. Ingold, *Structure and Mechanism in Organic Chemistry* (Cornell University Press, Ithaca, 1953)
40. G. Fisher, The autonomy of models and explanation: anomalous molecular rearrangements in early twentieth-century physical organic chemistry. Stud. Hist. Philos. Sci. **37**(4), 562–584 (2006)
41. S.G. Sterrett, Morals of model-making. Stud. Hist. Philos. Sci. **46**, 31–45 (2014)
42. A. Levy, Modeling without models. Philos. Stud. **172**(3), 781–798 (2015)
43. J. Kuorikoski, P. Ylikoski, in external representations and scientific understanding. Synthese, 1–21 (forthcoming)
44. M.J. Nye, *From Chemical Philosophy to Theoretical Chemistry: Dynamics of Matter and Dynamics of Disciplines, 1800–1950* (University of California Press, Berkeley, 1993)

45. J.F. Bunnet, Physical organic terminology, after Ingold. Bull. Hist. Chem. **19**, 33–42 (1996)
46. W.H. Dray, *Laws and Explanation in History* (Clarendon Press, Oxford, 1966)
47. R. Hughes, The Ising model, computer simulation, and universal physics, in *Models as Mediators: Perspectives on Natural and Social Science*, ed. by M.S. Morgan, M. Morrison (Cambridge University Press, Cambridge, 1999), pp. 97–145
48. D.C. Christensen, *Hans Christian Ørsted: Reading Nature's Mind* (Oxford University Press, Oxford, 2013)
49. A.-M. Ampère, in Correspondance d'Ampère, Lettre L596, CNRS, [Online]. Available: http://www.ampere.cnrs.fr/amp-corr596.html. Accessed 22 July 2015
50. F. van Wageningen-Kessels, H. van Lint, K. Vuik, S. Hoogendoorn, in Genealogy of traffic flow models. EURO J. Transport. Logistics, 1–29 (2014)
51. J. Cat, Modeling cracks and cracking models: structures, mechanisms, boundary conditions, constraints, inconsistencies and the proper domains of natural laws. Synthese **146**(3), 447–487 (2005)
52. A. Gelfert, Scientific models, simulation, and the experimenter's regress, in *Models, Simulations, and Representations*, ed. by P. Humphreys, C. Imbert (Routledge, London, 2012), pp. 145–167
53. A. Gelfert, Strategies of model-building in condensed matter physics: trade-offs as a demarcation criterion between physics and biology? Synthese **190**(2), 253–272 (2013)
54. M. Redhead, Models in physics. Br. J. Philos. Sci. **31**(2), 145–163 (1980)
55. P. Forber, Confirmation and explaining how possible. Stud. Hist. Philos. Biol. Biomed. Sci. **41**(1), 32–40 (2010)
56. U. Mäki, Remarks on models and their truth. Storia del Pensiero Economico **3**(1), 7–19 (2006)
57. H.W. de Regt, S. Leonelli, K. Eigner (eds.), *Scientific Understanding: Philosophical Perspectives* (University of Pittsburgh Press, Pittsburgh, 2009)
58. 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
59. S. Ahrens, *Experiment and Exploration: Forms of World-Disclosure* (Springer, Heidelberg, 2014)
60. H.W. de Regt (ed.) Understanding without explanation (Special Section). Stud. Hist. Philos. Sci. **44**(3), 505–538 (2013)

Chapter 5
Models as Mediators, Contributors, and Enablers of Scientific Knowledge

5.1 Models as Mediators

Over the course of the previous four chapters, we have encountered a range of uses and functions of models and a variety of ways in which models can be constructed, applied, and evaluated in their own right—that is, without treating models as a merely auxiliary step in the application of fundamental theory to the complex and messy real world. By acknowledging models as *autonomous* in this way, the argument developed in this book is in broad agreement with what has come to be called the 'models-as-mediators' view, due to Margaret Morrison and Mary Morgan—though, as we shall see, there are also a number of crucial differences [1].

The models-as-mediators view starts from an observation concerning how models are typically constructed. Model-building, Morrison and Morgan argue, is 'not only a craft but also an art, and thus not susceptible to rules' [1, p. 12]. When scientists build models, they draw on a range of different resources and (theoretical and material) ingredients, and 'it is because [models] are made up from a mixture of elements, including those from outside the original domain of investigation, that they maintain [their] partially independent status' [1, p. 14]. Any model that is intended to be applicable to a specific process or phenomenon necessarily depends on considerations that are extraneous to fundamental theory:

> Because models typically include other elements, and model building proceeds in part independently of theory and data, we construe models as being outside the theory–world axis. It is this feature which enables them to mediate effectively between the two. [1, pp. 17–18]

Of the various elements that together make up a scientific model, some may derive from theory, whereas others may originate from extra-theoretical considerations: 'model construction involves a complex activity of integration' [2, p. 44]. More often than not, for example when certain elements of a model are incompatible, this integration can be neither perfect nor complete. A case in point is the Bohr model's conflicting demands that the electrons in an atom should be

A. Gelfert, *How to Do Science with Models*, SpringerBriefs in Philosophy,
DOI 10.1007/978-3-319-27954-1_5

conceived of as orbiting the nucleus on circular paths without losing energy, while at the same time viewing them as objects of classical electrodynamics.

The thought that models 'mediate' between theory and data, while suggestive, requires some unpacking, lest it remain vague. As Morrison has recently emphasized, early versions of the models-as-mediators view involved two different senses of mediation. In the first case, 'the starting point is some physical system about which we have insufficient knowledge; so we construct a model in an attempt to learn more about its hypothetical features'. In other words, the model serves as a hypothetical realization of a scenario whose theoretical ramifications may then be studied in more detail (the example Morrison gives is of 'models that describe physics beyond the standard model'). In the second case, we begin by modeling a target system in a particular way and then find ways of fruitfully applying 'highly idealised and abstract laws to phenomena'—for example by modeling an electron encountering an impurity in a solid in terms of the wave function of a quantum particle in a potential well [3, pp. 119–120]. In the latter type of cases, it may be preferable to speak of the '*putative* target', given that not all situations to which highly idealized and abstract laws can be applied need to be, in fact, realized (for example, Morrison herself mentions the 'infinite potential well' as an example). More recently, the model-as-mediators view has been further generalized to also include cases where a model functions 'as a mediator in its role as the "object" of inquiry'; that is, 'the model itself, rather than the physical system, is the thing being investigated' [3, p. 120].

If models indeed routinely include 'additional "outside" elements'—i.e., elements that can neither be deduced from theory nor be found among the data—the question arises how they can nonetheless acquire the requisite degree of cohesion that we would expect from 'autonomous agents' [2]. This question is all the more pressing if one considers the range of different *types* of elements that are thought to feature in models. Thus Marcel Boumans, in his study of business-cycle models in economics, lists as some of the key ingredients 'theoretical notions, metaphors, analogies, mathematical concepts and techniques, policy views, stylised facts and empirical data' [4, p. 94]. In the case of economic models, many of which are formulated in the language of mathematics, much of the integrative work is due to what Boumans calls 'mathematical moulding', the definition of which also applies to mathematical models in other disciplines:

> Mathematical moulding is shaping the ingredients [of a model] in such a mathematical form that integration is possible, and contains two dominant elements. The first element is moulding the ingredient of mathematical formalism in such a way that it allows the other elements to be integrated. The second element is calibration, the choice of the parameter values, again for the purpose of integrating all the ingredients. [4, p. 90]

Boumans's explicit mention of 'mathematical formalism' as an ingredient in modeling is significant, and the specific contribution that mathematical formalisms make to the construction of models—already discussed in some detail in Chap. 3—will be taken up in more detail in the next section. For Boumans, formalisms stand alongside other, equally important 'ingredients', which include 'theoretical ideas, policy views, mathematisations of the [business] cycle, metaphors and empirical facts' [5, p. 4]. Model-builders therefore pursue 'an adaptive strategy until all ingredients [are] integrated', so as to achieve 'empirical significance' [5, p. 50]. Sometimes, mathematical moulding may be as simple as choosing suitable parameter values in order to accommodate certain model assumptions; often, however, it will require adapting existing formalisms to a particular type of target system. In a similar vein, Morgan likens the process of integration to the pulling together—'as ingredients of a recipe are'—of the various theoretical and empirical ingredients [6, p. 46].

When can we say that integration of the various elements into an autonomous model has been achieved, such that the latter is capable of 'mediating' between theory and data? Given that integration is a matter of degree, this will depend in part on the model user's purposes, which fits well with the earlier characterization (see Chap. 1) of the models-as-mediators approach as a pragmatic account of scientific modeling. This, however, does not mean that no more can be said; indeed, the question of what makes for successful integration is intimately related to the criteria for the appraisal of models as mediators more generally. Proponents of the models-as-mediators approach such as Boumans have argued that the main goal 'in the context of discovery is the successful integration of those items that satisfy the criteria of adequacy' [4, p. 67], thereby placing a premium on the empirical performance of a model and its various elements. This is thought to also explain why, 'in the integration process, "tuning" is essential': in Boumans's case study of models of the business cycle, much of the integration via mathematical moulding is achieved by ensuring that various parameter values are 'chosen such that the model could precisely mimic specific facts about the cycle' [5, p. 50].

The idea that models should be assessed by how well they can be moulded to fit specific empirical contexts can be traced back to early versions of the models-as-mediators view. In one of the first papers to explore the conception of models as mediators, Morrison emphasized that 'the proof or legitimacy of the representation arises as a result of the model's performance in experimental, engineering and other kinds of interventionist contexts—nothing more can be said!' This would suggest that model construction should primarily be driven by a concern for whether its products—the models—are empirically successful, that is: by how well they fit with observed phenomena. Yet, as we have seen—both in the context of specific examples from across the natural sciences (Chap. 3) and in the case of exploratory modeling more generally (Chap. 4)—models, including successful ones (as measured, for example, by their contribution to scientific understanding or insight), need not always be closely tied to specific empirical phenomena or contexts. Sometimes a model can afford insight, for example by exploring counterfactual (or even merely conceptual) dependencies, even in the absence of an actual target system. This is why Morrison's most recent acknowledgment that 'a model can also function as a mediator in its role as the "object" of inquiry' [3, p. 120] is significant: sometimes, for example when we lack in-principle access to (potential) target systems, all we know is how the model behaves, yet even this may be a source of significant scientific insight.

5.2 Mature Mathematical Formalisms as a Representational Resource

In order to better understand how models, considered as objects of inquiry, behave, it is often useful to look a how they were constructed. In this section, we shall discuss in detail a type of model-construction we already encountered, if only briefly, in Chap. 3 above: ‘formalism-driven’ model construction. The term ‘formalism-driven’ is meant to reflect the fact that models thus constructed are the result of deploying a ‘mature mathematical formalism’—that is, an integrated and entrenched system of rules and conventions for the manipulation of various symbols and terms, which are typically expressed in the language of mathematics and interpreted in accordance with a set of theoretical and methodological commitments.[1] The specific example we considered was the formalism of creation and annihilation operators, $\hat{a}_i^{\dagger}$ and $\hat{a}_j$, as applied to the physics of quantum many-body systems. The formalism of creation and annihilation operators originates from quantum field theory, which governs the behaviour of particles and fields at high energies, but has been fruitfully extended to strongly correlated, low-energy systems such as delocalized electrons in a metal. It is the latter type of system that, for example, the Hubbard model (see Chap. 3, Sect. 3.3) is intended for. In such a system, the electrons can be thought of—to a first (very crude) approximation—as a uniformly spread-out ‘sea’ of electrons, consisting of a quantum superposition of ‘electron waves’, each of which is characterized by a unique wave vector $\vec{k}$, indicating its energy and momentum, as well as a spin variable (with an electron’s spin σ pointing either ‘up’ or ‘down’ along the direction of measurement). Changes in the system, for example an energy excitation brought about by an external perturbation, may then be described in terms of the creation and annihilation of new electron waves—thereby adding or removing contributions to the system’s overall state. The formalism of quantum mechanics allows for the representation of such processes in terms of operators acting on the many-body quantum state of the system—hence the creation and annihilation operators $\hat{a}_{\vec{k}\sigma}^{\dagger}$ and $\hat{a}_{\vec{k}\sigma}$.

The mere deployment of a certain branch of mathematics—in this case, operator algebra—does not by itself constitute a mature mathematical formalism in the sense intended here. Obviously, from a purely mathematical viewpoint, speaking of the ‘creation’ or ‘annihilation’ of ‘particles’ in relation to the deployment of certain types of operators is quite meaningless: for such talk to make sense, the mathematical operations need to be given a physical interpretation—for example, of the type sketched in the previous paragraph. Furthermore, a (physically interpreted) mature mathematical formalism needs to constrain the wide range of mathematically permissible scenarios so as to ensure that it does not lead to (too many)

[1] see also [36].

expressions that lack a physical interpretation. This is why, in addition to their mathematical aspects, mature formalisms often incorporate a number of theoretical and methodological commitments—either explicitly, for example in the form of rules that prohibit certain operations within the formalism, or implicitly, through shared judgments about what constitutes an acceptable use of the formalism.

As an example of how mathematical formalisms need to be properly constrained, consider the example of the *conservation of particle number* in quantum many-body systems. At the fairly low energies involved in condensed-matter phenomena such as conductivity and magnetism—unlike, say, in the case of collisions in a particle accelerator—no elementary particles can be annihilated completely or created *ex nihilo*. Yet there is nothing in the mathematics of operator algebra—or indeed in the formalism of creation and annihilation operators *as used in high-energy physics*—that rules out the possibility of a varying number of particles in a system. If one were to deploy creation and annihilation operators 'in isolation', as it were, then (on the standard interpretation) their net effect would be such that they would seem to describe states where an individual electron has been created *ex nihilo*, or destroyed without a trace. This is why, in the case at hand, it is stipulated that, as a general rule, creation and annihilation operators cannot feature in isolation in the Hamiltonians describing quantum many-body models, but must always occur in pairs. This way, an electron that is 'annihilated' in one place will immediately be 'recreated' at a different point in the system, so that the system at no point violates conservation of particle number. This constraint on how operators may be deployed within the formalism is matched by a convention regarding how to interpret its output, such as the operator product $\hat{a}_i^\dagger \hat{a}_j$: although the formalism does not allow for the explicit representation of time, such that there appears to be no 'time lag' between the annihilation of an electron at site j and its reappearance at site i, physicists nonetheless choose to give a dynamic interpretation to the overall effect of the combination of an annihilation and a creation operator, by simply interpreting the expression $\hat{a}_i^\dagger \hat{a}_j$ as describing the *movement* of an electron from one place in the crystal lattice to another.

The example of creation and annihilation operators in many-body physics thus illustrates an important feature of *mature mathematical formalisms*: namely, their dual character as both *enabling* and *constraining* the development of scientific representations. On the one hand, the details of a formalism enable what can be easily represented by it; in other words, mature mathematical formalisms *afford* their users certain ways of model construction and thereby function as representational resources. Thus, the rules governing creation and annihilation operators shape how a many-body theorist will go about modeling complex phenomena, e.g. in the form of simple additive contributions to the overall Hamiltonian, and—as in the case of 'electron hopping' (mentioned in the previous paragraph and discussed in detail in connection with the Hubbard model in Chap. 3, Sect. 3.3)—they often suggest specific ways of representing physical processes, thereby contributing to the intelligibility of a model and a sense of understanding on the part of its user.[2] On the other hand, mature formalisms also impose various restrictions and come to have a constraining role. Many such restrictions are implicit in the rules of a formalism: as formalisms mature, they undergo a process of elaboration, enrichment, and entrenchment, and come to embody theoretical, ontological, and methodological commitments and assumptions. Importantly, these constraints and commitments may no longer be obvious to either the novice or the proficient user, since they are usually 'enshrined' in the formalism, cloaked behind purely notational rules which, in turn, determine which subset of possible processes and mechanisms a formalism lends itself to representing.

The importance of formalisms to the construction of scientific representations has been recognized before by others, but their precise function and characteristics have sometimes been obscured by a hasty assimilation of the issue of mathematical formalisms to the more general question of the applicability of mathematics to nature. One of the first to recognize that mathematical formalisms often have rather specific heuristic and theoretical roles in scientific inquiry was Mary Hesse. In her 1953 paper on models in physics, she notes: 'Mathematical formalisms, when used as hypotheses in the description of physical phenomena, may function like the mechanical models of an earlier stage in physics, without having in themselves any mechanical or other physical interpretations'. More specifically, according to Hesse, mathematical formalisms have an important role to play in those areas of modern physics, such as quantum theory, where 'we are told that we must not ask for picturable mechanical or electrical models', but should instead rely exclusively on 'formal mathematical hypotheses' [8, p. 199]. In such highly abstract areas of science, what is desperately needed, in the absence of visualizable aids to reasoning such as mechanical models, are *alternative* 'pointers towards further progress'. As Hesse sees it, what takes the place of easily pictured mechanical models

> is provided by the nature of the mathematical formalism itself—any particular piece of mathematics has its own ways of suggesting modification and generalisation; it is not an isolated collection of equations having no relation to anything else, but is a recognisable part of the whole structure of abstract mathematics, and this is true whether the symbols employed have any concrete physical interpretation or not. [8, p. 200]

Two points are worth highlighting especially in this context. First, Hesse argues that mathematical formalisms fulfil a similar heuristic function as picturable mechanical models, in that both have often been suggestive of new directions of research. The tentative language is important here: picturable mechanical models as well as mathematical formalisms 'point towards' progress and 'suggest' new steps of modification and generalization, rather than *logically entailing* them or making them otherwise inevitable. Second, mathematical formalisms do not merely accommodate whatever hypotheses one may already have formulated regarding the target system, but instead also *constrain* the ways in which a system can be represented. As a '*particular* piece of mathematics', each formalism 'has its *own* ways

[2]On this point, see also Chap. 3 (Sect. 3.2.3).

of suggesting modification and generalisation' (italics added); that is, every choice of a particular mathematical formalism involves a trade-off between certain affordances and constraints.

While there exists a great deal of affinity between Hesse's views and my suggestion that mature mathematical formalisms occupy a special place in model construction, there also remain several crucial differences. Thus, when Hesse equates a mathematical formalism with 'any particular piece of mathematics', this may be rather too global a characterization for our purposes. The mere deployment of the formal methods of one particular branch of mathematics does not by itself constitute a mature mathematical formalism. For this to be the case, a formalism must lend itself to physical interpretation, as in the case of creation and annihilation operators in many-body physics (this already marks a point of contrast with Hesse who, in the block quote above, remains neutral as to 'whether the symbols employed have any concrete physical interpretation or not'). Furthermore, a mature mathematical formalism must be sufficiently general to be applicable to a *range of* physical problems; at the same time, it will typically fall short of universality, in the sense that certain cases—though logically and physically permissible—do not lend themselves to being modeled using the formalism. This is simply the flip side of a mature mathematical formalism's having already incorporated certain theoretical, methodological, and heuristic assumptions that are not themselves part of either the mathematical framework at large or the 'underlying' physical theory. Much of the value of a mature formalism derives from precisely such built-in theoretical constraints: they ensure that its output—e.g. a many-body Hamiltonian—will automatically satisfy important conditions (e.g. preservation of particle number, Fermi's exclusion principle, etc.). This is why more is required for a mature formalism than just the wholesale application of mathematics-at-large to scientific problems; it takes additional rules and notational conventions, which are appropriate to the specific domain of application, for a 'particular piece of mathematics' to count as a mature formalism.

The emphasis on mathematical formalisms as representational vehicles should not obscure the fact that similar considerations also apply to other (e.g. visual or diagrammatic) symbol systems. While this is not the place for developing a full definition of what constitutes a 'symbol system', this much seems uncontroversial: typically, a symbol system requires that well-formed arrangements such as marks on paper, figures in a table, etc., be registered semantically as instances of a particular character (i.e., that such a system, in Nelson Goodman's terminology, be 'syntactically articulate'), and that certain other features (e.g., compositionality) allow for the manipulation and interpretation of expressions within that system. As an example, consider the case of the notational system of Feynman diagrams, which was developed with the goal of representing a potentially indefinite number of physical processes in quantum electrodynamics. Each Feynman diagram consists of points ('vertices') and arrows (of different orientation) attached to the vertices, representing interacting electrons and positrons, as well as wavy lines signifying photons that may be emitted or absorbed. Enshrined in the formalism of Feynman diagrams are both rules for the construction of new diagrams (e.g., 'At every vertex, conservation of energy and momentum among the interacting particles is required'), as well as for the interpretation of the diagrams thus generated (e.g., 'Lines in intermediate stages in the diagram represent "virtual particles", which may "temporarily" violate the relativistic energy-momentum relation, but which are in-principle unobservable if they do not'). While the formalism of Feynman diagrams was developed on the basis of an overarching theoretical conception—which takes each diagram to represent a contribution to the total amplitude for a (multiply realizable) quantum process—it has taken on 'a life of its own' in certain areas of high-energy physics, where it has developed from a mere shorthand to what one might call a notational *lingua franca*.

The ubiquity of mature notational systems across the sciences—ranging from mathematical formalisms in physics to structural formulas in chemistry and the use of cladograms in evolutionary biology—attests to their utility as aids to scientific inference. Much of this utility derives from the fact that mature formalisms and notations provide powerful representational resources, which scientists can draw on in their attempts to represent reality, explore the consequences of their models, and convince colleagues of the merits of their theories and hypotheses. Furthermore, mature formalisms and notations also function as a way of outsourcing inferential work, for example by ensuring that results derived within the formalism satisfy the requisite criteria of validity. As Lorenzo Magnani notes, while such 'external forms of representation'

> can give people access to knowledge and skills that are unavailable to internal representations, help researchers to easily identify aspects and to make further inferences, they constrain the range of possible cognitive outcomes in a way that some actions are allowed and others forbidden. [9, p. 443]

This is especially obvious in the case of mathematics, given the truth-preserving nature of logical and mathematical reasoning, but it applies equally to physically interpreted formalisms such as Feynman diagrams and the operator formalism in quantum physics, where adherence to certain syntactic rules of the formalism often automatically ensures that certain physical constraints—conservation of energy and momentum, or conformity with Pauli's exclusion principle—will be satisfied. Importantly, such formalisms and notations need to be entrenched and sustained through collective social practices; in the case of Feynman diagrams, as historians of science have noted, there was much initial resistance, with the method being perceived as 'too new, and too idiosyncratic' [10, p. 91]. Yet, although physicists initially 'practiced drawing and interpreting the diagrams in distinct ways, toward distinct ends' [11, p. 165], over time these methods converged into a powerful and near-universal representational medium. Once such representational formalisms have collectively solidified, and have been mastered at the individual level, they allow their users 'to alleviate the cognitive load and increase the reliability of [their] inferences' [12, p. 7]. While it would be an exaggeration to say that mature formalisms and notations 'do the thinking' for their users, neither would it be adequate to treat them simply as neutral tools; instead, they constitute powerful representational resources that both enable and constrain what can be done with them.

5.3 Models as Contributors

As argued in the previous section, the existence of a mature formalism—such as the formalism of creation and annihilation operators in many-body physics—can ensure that a model generated on its basis is, in various ways, sensitive to the phenomenon or target system it represents. Thus, in the case of models of strongly correlated electron systems, correctly deploying the formalism ensures that the resulting models will conform to certain basic theoretical commitments, such as the Pauli principle governing the behaviour of fermions. At the same time, the formalism frees the model from other types of constraints. Consider again the Hubbard model. As we saw in our earlier discussion, one of the core contributions to the Hubbard Hamiltonian arises from the movement ('hopping') of electrons from one site in the crystal lattice, i, to another, j. The extent of this contribution is reflected by the hopping integrals, T_{ij}, an estimate of which can be derived from 'first principles' by calculating the influence of the wave functions of neighbouring atoms. This would seem to impose an enormous burden on the modeler to get the numerical values for T_{ij} 'just right' (where this might mean reproducing empirical results—provided these are available—or employing quasi-exact numerical techniques). However, by switching perspectives and moving from a 'first-principles' approach to a formalism-driven approach, this burden can be easily alleviated: from a formalism-driven perspective, the hopping integrals are simply parameters that can be chosen largely arbitrarily (except perhaps for certain symmetry requirements). No matter what one's specific choice of numerical values for T_{ij}, the resulting model will count as a valid output of the formalism, and the modeler is free to adjust her choice of parameters depending on her goals of inquiry. Reliance on mature formalisms, thus, may not only serve to *enforce* conformity with certain (e.g. theoretical) constraints, thereby ensuring that the resulting model is sensitive to certain features of the target systems, but may also *liberate* the model from overly exacting demands, e.g. concerning the specific choice of parameter values.

The example discussed in the previous paragraph suggests that it would be quite misguided to think, as a naive conception of modeling would have it, that the success of a model need always be tied to how well they approximate, or instantiate, fundamental 'first principles'. Indeed, in the case at hand, a number of approximations have already gone into the derivation from 'first principles', for example by positing that interactions between atoms and their orbitals be limited to nearest neighbours. When the formalism-driven approach lifts restrictions on the hopping integrals T_{ij}, by treating them as parameters that can be freely chosen (rather than as values that need to be derived from fundamental theory), this need not entail any additional 'loss' of accuracy as compared with the 'first principles' approach. At the very least, given the unavailability of an exact solution to the full 'first principles' problem, there is no principled way of telling whether freely chosen parameters are further from the truth than values that have been derived via a series of approximations. More generally, any view that subjugates models to fundamental theory—or to empirical data, for that matter—risks overlooking the fact that models themselves often contribute new elements to the description and investigation of their target systems—elements which are not themselves part of the fundamental theory (or, as it were, cannot be 'read off' from it) but which may take on an interpretative or otherwise explanatorily valuable role.

Contributions of novel criteria, quantities, and structural features at the level of models can take many forms. Some of these novel contributions may come to inform the way scientists think about a class of target systems and may suggest new lines of research. Let us stick with the example of the Hubbard model in order to illustrate several kinds of such novel contributions. At a fairly basic level, new quantities and parameters may be generated by combining different elements of the model in novel ways. Thus, in the Hubbard model one finds two sets of parameters $U, \{T_{ij}\}$, with the former being a measure of the on-site Coulomb repulsion between electrons at the same lattice and the latter being the hopping integrals. In the interest of simplicity, let us assume that we are dealing with a cubic lattice and nearest-neighbour interaction, in which case the interaction T_{ij} between different lattice sites will either be zero or have the same fixed value t. By combining both parameters—U and t—into a single ratio, U/t, it is then possible to study how the *relative strength* of the interaction *between* electrons (as compared with their individual kinetic movement) affects the behaviour of the model. This is because, from 'within the world of the model' (to use Morgan's evocative phrase; [6, p. 238]), the ratio U/t is a unique and exact measure of the relative importance of the electrons' dynamics as compared with their on-site repulsion. Though initially introduced as parameters, the two quantities U and t are now linked by the model in a way that allows the modeler to explore the relative strength of the two processes: on-site repulsion and hopping. On the one hand, this gives more internal cohesion to the model, on the other hand it has implications for how one should, and should not, go about modifying the model. For example, an attempt to make the model more accurate by adding new (higher-order) terms to the model (perhaps accounting for higher-order interactions of strengths $V, W, X, Y, Z < U, t$), may be undesirable, as it may be more useful to have a *single* measure of the relative strength of the electron–electron interactions—U/t—rather than a whole set of different measures $\{U/t, V/t, W/t, \ldots\}$. If one's goal is to gain a deeper understanding of the model and its target, rather than merely making numerical predictions, it may simply not be desirable to replace an intuitively meaningful quantity with a set of parameters that lack a straightforward interpretation, even if such a move might lead to more accurate numerical results.

The 'active' contribution of the model, that is, its contributing new elements rather than merely integrating theoretical and experimental (as well as further, external) elements, is not only relevant to interpretative issues, but also opens up new dimensions of assessment and of linking seemingly unrelated models. In the case of many-body models, one especially salient class of novel contributions—known as *rigorous results*—illustrates both these points quite strikingly. The expression 'rigorous results', which is not without its problems, has become a term of art in theoretical physics, especially among practitioners of statistical and many-body physics (see for example [13]). It therefore calls for some clarification. What makes a result 'rigorous' is not the qualitative or numerical accuracy of a particular prediction of the theory or model. In fact, the kind of 'result' in question will often have no immediate empirical connection with the model's target system or phenomenon. Rather, it concerns an exact *mathematical* relationship between certain mathematical variables, or certain structural components, of the mathematical model—which may or may not reflect an empirical feature of the target system.[3] Put crudely, in much the same way as Pythagoras' theorem ($a^2 + b^2 = c^2$) is not merely true of a particular set of parameters, e.g. $\{a, b, c\} = \{3, 4, 5\}$, but holds for all rectangular triangles, so a rigorous result for a type of mathematical model holds for any instance of it, not just for a particular choice of parameter values. At the same time, rigorous results are true only *of a model* (or class of models) as defined by a particular Hamiltonian, unlike, say, certain symmetry or conservation principles that follow directly from fundamental theoretical considerations. In other words, they are genuinely novel contributions at the level of models, rather than at the level of fundamental theory or empirical data.

Rigorous results may also connect different models in unexpected ways, which can neither be readily deduced from fundamental theory (since rigorous results do not hold *generally* but only for certain types of models), nor be inferred from empirical data (since the corresponding target systems of the different models may be radically different). For our preferred example of the Hubbard model it can been rigorously shown that, at half filling—when half of the quantum states in the conduction band are occupied—and in the strong-coupling interaction limit, $U/t \rightarrow \infty$, the *Hubbard* model can be mapped on to the spin-1/2 antiferromagnetic *Heisenberg* model.[4] The Heisenberg model was first proposed by Werner Heisenberg in 1928 as a simple quantum mechanical model of magnetic insulators. Whereas the (classical) Ising model, discussed in Chap. 1 (Sect. 1.1), was based on the idea that a macroscopic magnet consists of 'elementary magnets', each of which can take only two values, $S_i = +1$ ('up') and $S_i = -1$ ('down'), the Heisenberg model replaces the classical variables S_i with quantum mechanical spin operators $\hat{S}_i$. Heisenberg was thus able to give a physical interpretation of the (previously merely posited) 'elementary magnets' in terms of the newly discovered quantum mechanical property of *spin*. The introduction of quantum operators also fundamentally alters the algebraic properties of the Heisenberg model, as compared to the classical Ising model. What matters in the present context, however, is not the extent to which the Heisenberg model constitutes an advance over the Ising model, but rather the fact that the Heisenberg model was proposed as a model of magnetic *insulators*, whereas the Hubbard model reflects the itinerant behaviour of electrons

[3]One way to think about rigorous results is to regard them as akin to mathematical theorems that are provable from within the model or theory under consideration. Indeed, the notions of 'theorem' and 'rigorous result' are frequently used interchangeably in scientific texts, especially in theoretical works such as [38].

[4]see, for example, [14].

in *conductors*. Conductors and insulators constitute very different classes of target systems, so it is surely surprising that a model for the former—the Hubbard model at half filling in the strong-coupling limit—maps onto a rather different type of model for the latter, the Heisenberg model. While the Hubbard model with *infinitely strong* electron–electron interaction ($U/t \to \infty$) cannot claim to describe an *actual* physical system (since the strength of any actual interaction is necessarily finite), various mathematical and numerical techniques can nonetheless be applied in the strong-coupling limit. This allows for the testing of the adequacy of the Hubbard model by exploring the numerically and analytically more tractable antiferromagnetic Heisenberg model—a truly astonishing result.

Rigorous relations between different many-body models not only provide fertile ground for testing of mathematical and numerical techniques, and for the 'exploration' (in the sense discussed in Chap. 4) of models more generally. They can also give rise to a transfer of empirical warrant across models that were intended to describe very different physical systems. The mapping, in the strong-coupling limit ($U/t \to \infty$), of the Hubbard model onto the spin-1/2 antiferromagnetic Heisenberg model is one such example. For, the latter—the antiferromagnetic Heisenberg model—has long been known as an empirically successful '"standard model" for the description of magnetic insulators' [14, p. 75], yet the Hubbard model at low coupling ($U/t = 0$, indicating zero electron–electron interaction) reduces to an ideal Fermi electron gas—a perfect conductor. It has therefore been suggested that, for some finite value between $U/t = 0$ and $U/t \to \infty$, the Hubbard model must describe a system that undergoes a transition from conductor to insulator. Such transitions, for varying strengths of electron–electron interaction, have indeed been observed in physical systems and are known as Mott insulators. Thanks to the existence of a rigorous relation between the two models, initial empirical support for the *Heisenberg* model as a model of a magnetic insulator thus translates into support for a new—and originally unintended—representational use of the *Hubbard* model, namely as a model of Mott insulators. In other words, 'empirical warrant first flows from one model to another, in virtue of their standing in an appropriate mathematically rigorous relation' [15, p. 516], from which one may then gain new insights regarding the empirical adequacy of the model.[5] As this example illustrates, rigorous results neither borrow their authority from fundamental theory nor do they always need to prove their mettle in experimental contexts; instead, they are genuine contributions of the models themselves and afford their users new ways of confirmation, assessment, and exploration—well beyond the narrow focus on a model's empirical performance 'in experimental, engineering and other kinds of interventionist contexts' [7, p. 81].

[5]This case of cross-model support between many-body models that were originally motivated by very different concerns is discussed in detail in [15].

5.4 Models as Epistemic Tools

The idea that models and their various functions in scientific inquiry are best analyzed by paying close attention to representational means and their 'characteristic limitations and affordances' [16, p. 695], has been developed more fully in recent work on models in the engineering sciences by Tarja Knuuttila and Mieke Boon. Starting from the realization that, in the engineering sciences, the goal of accurately representing extant target systems often takes second place to actively intervening in the world in order to generate new objects, Boon and Knuuttila argue that it is best to 'consider scientific models in engineering as *epistemic tools* for creating or optimizing concrete devices or materials'. This, they argue, generalizes to scientific modeling more broadly, which may be characterized 'as a specific scientific practice in which concrete entities, i.e. models, are constructed with the help of specific representational means and used in various ways, for example, for the purposes of scientific reasoning, theory construction and design of other artifacts and instruments' [16, p. 689]. By explicitly adopting a 'functional perspective' [16, p. 696], this view not only aligns itself with pragmatic approaches more generally (as discussed in Chap. 1, Sect. 1.6), but also coheres well with the overall approach developed in this book, which is based on a recognition of the diversity of uses and functions of scientific models.

Emphasizing their character as *concrete objects* is of central importance to the view of scientific models as epistemic tools. As Knuuttila puts it elsewhere, models ought to be regarded 'as concrete artefacts that are built by specific representational means and are constrained by their design in such a way that they facilitate the study of certain scientific questions […] by means of construction and manipulation' [17, p. 262]. Indeed, its recognition of the concrete material dimension of models is put forward as a criterion for what distinguishes this view from the models-as-mediators view with its emphasis on the partial autonomy of models from both theory and data; as Boon and Knuuttila argue, 'it is not sufficient that [models] are considered as autonomous; they also need to be concrete in the sense that they must have a tangible dimension that can be worked on' [16, p. 694]. One might worry that this emphasis on the concrete materiality of model is too limiting, in that it makes it difficult to see how, on this view, abstract models—including mathematical models—can function as tools of inquiry. Boon and Knuuttila aim to address this worry by explicitly extending their analysis to abstract models: 'when working with them we typically construct and manipulate external representational means such as diagrams or equations'. Abstract models may be given material form via 'scale models, pictures, diagrams, different symbolic formulas and mathematical formalisms', all of which 'suggests that the material dimension of models and the diverse representational means they make use of are crucial for their epistemic functioning' [16, p. 695]. The example of Feynman diagrams, discussed in Sect. 5.2 above, though not directly mentioned by Boon and Knuuttila, is a good illustration of how the use of external representational means—even when their material dimension consists only in marks on paper, combined with conventions regarding their use and interpretation—facilitates abstract reasoning and stabilizes inferences by anchoring them in a more accessible medium.

It is clear, from this brief summary, that there are strong affinities between this view of models as epistemic tools and the view, developed above and in the preceding chapters, that models are functionally diverse contributors to scientific inquiry, which draw on a range of representational resources. Both views acknowledge the dual nature of models as *enabling* and *constraining* what their user can do with them; indeed, as Boon and Knuuttila aptly put it, 'modellers typically proceed by turning the constraints […] built into the model into affordances' [16, p. 695]. At the same time, there remain certain differences between the two views, some of which are a matter of emphasis, while others point to contrasts that run deeper. For one, there remains a gap between viewing models as epistemic tools, which emphasizes their concreteness, and making sense of certain uses to which models are put. As we saw in Chap. 4 (esp. Sect. 4.4), models often do not aim at representing individual target systems, but aim at exploring universal features of *classes* of target systems—which, in many cases (such as higher-dimensional systems in physics, or multi-sex species in biology) are not, and perhaps could not be, instantiated. In the case of physical tools, although *multiple utilizability* is often considered a key characteristic of how tools function [18, p. 227], any particular use is causally tied to a particular target, not to a whole class of target system. How, then, is it possible for models to allow us to represent, explore, and otherwise investigate whole classes of models, including counterfactual scenarios? To be sure, when a user interacts with a specific external representation—e.g., when she simplifies an equation on the whiteboard, or when she manipulates a given Feynman diagram—such interactions are indisputably concrete; yet, acknowledging this only pushes the question back one step: at the very least, what is needed is an account of how it is that we can so easily move between the concrete world of representational media and the (real or fictitious) target systems that are being modeled. In the next section, an attempt will be made to develop a general framework for making sense of how models enable us to gain epistemic access to (information about) their—real or purported—target systems *qua* interacting with representational means that are typically different in character from the target.

None of this, however, should detract from the significance of realizing that there are important parallels between models and tools. For material models, of course, this similarity has long been noted and has been accompanied by the realization that physical manipulation, on occasion, can take the place of theoretical derivation. As Davis Baird observes in relation to orreries (i.e., mechanical models that 're-enact' the motion of various planets and their moons), the very way these models are materially constituted allows one to use them—in a hands-on way, by setting them in motion—'to demonstrate […] the shape of the moon's orbit relative to the sun'. Similarly, in the case of the physical ball-and-stick model of the DNA double helix (see Fig. 5.1), it is entirely appropriate to say that 'the sticks in Watson and Crick's model denote bond lengths, not rigid metallic connection' [19, p. 38].

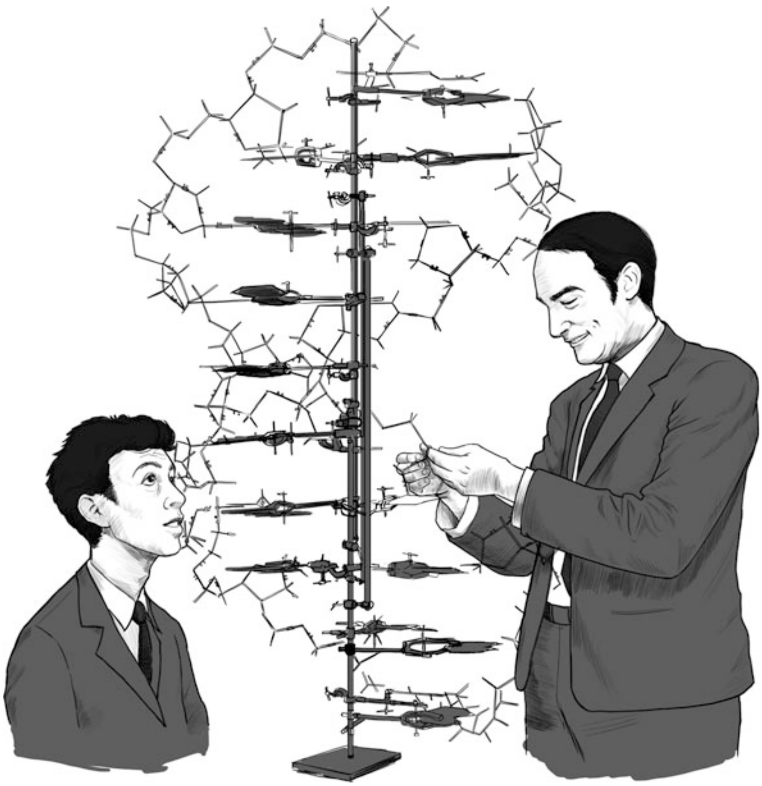

Fig. 5.1 James Watson (*left*) and Francis Crick (*right*) interacting with their physical model of the DNA double helix. Artwork by Jerry Teo

In other words, the interactions that the material realization of a model affords its users crucially shape what kinds of targets it is suitable for and what kinds of inferences it enables. Perhaps the strongest endorsement of the importance of representational media to the utility of models comes from scientific practitioners. James Watson, in his historically controversial first-person account of the discovery of the structure of DNA and the research leading up to it (including Linus Pauling's discovery of the protein α-helix), gives vivid expression to this insight: 'The α-helix had not been found by staring at X-ray pictures; the essential trick, instead, was to ask which atoms like to sit next to each other. In place of pencil and paper, the main working tools were a set of molecular models superficially resembling the toys of preschool children' [20, p. 34]. Whatever the shortcomings of Watson's historical narrative, not least regarding his portrayal of Rosalind Franklin's contribution, this much he is getting right: the ability to physically manipulate a material model—whether in the case of the α-helix, or the full-fledged DNA double helix model—was important not just for illustrative purposes, or because it somehow provided a convenient shortcut for theoretical reasoning based on empirical data, but because it afforded novel ways of exploring and approaching the target system.[6]

A final caveat about viewing scientific models as epistemic tools concerns the implied neutrality of models as instruments of inquiry. If models are thought of as (passive) tools, which a user may or may not help herself to, then it may be tempting to view the function of models in solely instrumental terms: on this account, a model's success would be defined by how seamlessly and efficiently it allows its user to achieve an intended outcome. Philosophers of technology have challenged this 'neutrality assumption' regarding the use of tools and technology in a twofold way. First, a close look at how technological devices are deployed reveals that, once they become sufficiently widely adopted, they can no longer be said to instrumentally realize pre-existing goals and intentions, but instead begin to reshape social reality in a way that also affects human desires and interests. There is no reason to think that the situation is any different in the case of epistemic tools; indeed, a number of case studies from the history of science have shown that, depending on social and cultural context, the 'same' models may be put to quite different uses and may have divergent effects on the collective development and assessment of theories and research programmes.[7] Second, the specific constraints and affordances of technological devices may directly affect how individual users perceive problem situations and what they focus their attention on. Notoriously, 'to someone with a hammer, everything looks like a nail'; similarly, a scientist's choice of a particular type of model—which, more often than not, will itself not be an unconstrained decision—may determine which scientific problems they judge significant and how they assess specific scientific hypotheses. Sometimes, as in the case of Watson and Crick's double helix model, a model can focus attention on, and give stability to, a theoretical hypothesis—in this case, concerning the molecular structure of DNA—even in the face of unclear or misleading empirical evidence (in this case, crystallographic evidence from X-ray diffraction). Sometimes, as in the case of models of the cell membrane, attachment to a particular structural model—the 'unit membrane' model, which portrayed the membranes of all cell types as having a trilaminar protein-lipid-protein structure—can steer inquiry away from a

[6]The allusion to 'exploration' as a function of material models is entirely deliberate: for, as Baird notes, material manipulation 'is particularly important when conceptual manipulations are impossible either for lack of a theory or because analytical manipulations would be too difficult' [19, p. 39]. Similarly, Magnani identifies as one of the core features of what he calls 'manipulative abduction' that 'manipulations have to be able to introduce potential inconsistencies in the received knowledge and so *to open new possible reasoning opportunities*' [9, p. 444]. Baird's and Magnani's remarks cohere well with my earlier characterization, in Chap. 4, of exploratory modeling as model-based research in the absence of a fully-formed theoretical framework; interestingly, Magnani explicitly refers to the experimental work of Ørsted (see the discussion in Chap. 4, Sect. 4.3) as an instance of successful 'manipulative abduction'.

[7]For example, Grant Fisher has demonstrated how cultural differences between communities of physical organic chemists in the UK and the United States impeded convergence in their modeling strategies, specifically in relation to the molecular rearrangements discussed in Chap. 4 (Sect. 4.5.3). See [35, pp. 580–581].

true account of the target system (which, in this case, turned out to be far messier than anticipated, with the cell membrane now known to consist of a lipid bilayer and an irregular distribution of proteins—some integral, some peripheral—rather than being made up of a highly ordered trilaminar structure).[8] In both of these cases, and across the natural and social sciences more generally, models actively shape the course of scientific inquiry, even as they themselves are subject to a range of empirical, social and historical factors. Therefore, rather than thinking of models first and foremost as (passive) tools—which then, in a second step, creates the need for addressing issues arising from their non-neutrality—it may perhaps be more apt to acknowledge their active role in inquiry from the start. This is what the proposed view of models as *contributors to inquiry* is intended to highlight, though it is important to stress that this amounts to a difference in emphasis rather than to a fundamental disagreement with Knuuttila's and Boon's view of models as epistemic tools.

5.5 Models as Enablers of Scientific Knowledge

What does it mean to say that a successful model functions as an 'enabler' of scientific knowledge? For one, it means that, for a model to be successful, more is required than that it stand in the right sort of objective relationship to its target system. Consider a 'black box' whose inner structure maps perfectly onto a target system, but which at the same time prevents any user from accessing its information-carrying inner relations. Such an object could hardly be said to be a good model of its target, given that, for lack of accessibility, it would be impossible for us to learn anything from it about the target system. A successful model should enable such learning, by making relevant information about its target accessible to us—not only in principle, but in a sufficiently salient way, such that a reasonably skilled user would be able to draw relevant inferences about the target system from interacting with the model via the representational means it employs. A putative model that failed to make salient to us any relevant connections that exist between the model and its target would be of as little use to us as a model that lacked such a connection in the first place. Hans-Jörg Rheinberger makes a similar point when he gives the following characterization of models, which can also be read as a normative statement about how successful models ought to function: 'What models basically enable is an overview *at one glance* of a multiplicity of data and of how they interrelate' [21, p. 325]. In other words, in order for a model to enable us to acquire scientific knowledge about aspects of the world, it must not only stand in the right sort of objective relation to its target, but must also afford us some kind of cognitive access to the information it contains.

[8]For a detailed discussion of models of the cell membrane, see [39].

This much is largely uncontroversial. What is more contentious is how one should best characterize this dual relation of models to their users and targets. One possible approach is to distinguish between, on the one hand, different degrees of *immediacy* in the way users can interact with a model (qua representational device) and, on the other hand, different degrees of *directness* in the relation between model and target system. Immediacy and directness, thus understood, are qualitatively different: the former relates to the phenomenology of our interaction with the representational means deployed by a model, whereas the latter concerns the link by which a model can be said to contain information about its target system. Because they concern different components in the complex 'user–model–target' relationship that gives rise to model-based representation, immediacy (in the way a user accesses a model) and directness (in the relation between model and world) can come apart. A model that is immediately accessible to its user may nonetheless stand only in a highly indirect relation to its target system, in that its construction requires many steps of idealization, abstraction, and approximation; similarly, a model may contain information directly obtained from its target system, without thereby being immediately accessible. As an example of the latter, consider a complex mathematical model that includes as parameters direct measurements obtained from its target system. Due to its high degree of mathematical complexity, the model may still require significant 'decoding' and interpretation on the part of its user; in other words, although it directly represents information about the target system, it may nonetheless lack immediacy from the user's perspective.

The point that immediacy and directness can come apart is, of course, entirely general and not limited to scientific models. Indeed, for certain modes of representation—such as visualization—it has attracted considerable attention. For example, Megan Delehanty notes that, in diagnostic imaging techniques such as PET or fMRI, the production of visual images often requires 'that extensive mathematical transformation occur to produce the data that can *then* be represented in the form of an image' [22, p. 161]. Thus, when a patient is sent through a PET scanner, the initially recorded output is a vast array of numerical values associated with spatial coordinates; in order to produce an image, it is necessary to translate the numerical data into visual stimuli, for example by displaying different intensities in different locations as coloured pixels on a three-dimensional grid. Visualization, then, is a deliberate strategy to increase *immediacy*: 'The epistemic value of cognitive accessibility, then, is not that images contain spatial information that is not present in the corresponding numerical data, but that they make it much easier to get it into our heads; to produce belief or knowledge' [22, p. 170]. Similarly, John Kulvicki notes that it is the '*immediate availability* of a great many pieces of abstract information [which] accounts for some of the epistemological weight given to images and graphs, not to mention photographs' [23, p. 298; italics added]. Visual immediacy is but one example of how representational means can determine cognitive accessibility; similar considerations apply to the modalities of touch, hearing, and so forth.

Moving from these global considerations to the case of scientific models, let us look at how immediacy and directness interact in material models, specifically in the case of wax models of biological organisms, which once were (and in some quarters continue to be) widely used as instructional tools. Such models may represent various stages of an organism's development, yet the most prominent example, which has received considerable attention from historians of science and medicine, are three-dimensional wax models of different stages of the embryo. In his study of the history of three-dimensional wax models in embryology, Nick Hopwood waxes lyrical (pardon the pun) about how 'scale, texture, and colour worked together to convert delicate and shimmering but tiny and elusive forms into solid and opaque but huge and memorable shapes' [24, p. 193]. The naturalistic effect of some such models, especially those created by Wilhelm His (1831–1904), is indeed striking (see Fig. 5.2). Having perfected his technique of 'plastic reconstruction'—which results in scale models that mimic the texture and appearance of the original specimen—His described the purpose of his models as 'giving body' to what would otherwise have been at best partial depictions of the organism. Hopwood aptly characterizes this representational strategy when he observes that, for His, '"the form of our body" was not a self-evident problem awaiting explanation', but instead required plastic reconstruction in order to reproduce a more immediate 'bodily apprehension of form' [25, p. 466]. Of course, a given wax

Fig. 5.2 Embryo model in black wax, by Wilhelm His. Reproduced with permission from Anatomisches Museum Basel

model need not, and rarely will, represent a specific (real-life) embryo: those creating the models would usually have aggregated information from different sources, including the study of actual specimens, in order to create a model of an 'average' embryo, which includes the *typical* features that would be expected at the corresponding stage of development, while leaving out individual variations. Yet, a given wax model nonetheless lends itself to being treated as a specific concrete object. Even if the model does not directly represent any particular target object, the kind of cognitive access it affords—at least in cases such as His's naturalistic wax models—closely mirrors the (visual or tactile) immediacy that we *would* experience, were we to encounter a real specimen. As one 19th-century anatomist put it, His's models presented students with different embryonic stages 'in corporeal and, we might even say, *graspable* form' [26, p. 545; emphasis added].[9]

How do models such as His's wax models function, and how does the interplay between immediacy and directness affect the way models work more generally? Given the variety of available formats and media, and the large number of ways in which we could potentially relate to them, it might seem that nothing much in general could be said about the immediacy and directness with which models allow us to learn about the world. Yet, some degree of unification, I want to suggest, can be achieved by shifting the focus away from the diverse material qualities of models towards the phenomenology of user–model relations. Taking the idea of models as 'tools of inquiry' seriously then allows us to draw on a rich body of work in the phenomenology of human–technology interaction. Just as tools and technologies afford their users different ways of accessing and manipulating the world, so models enable different kinds of *user–model–target relations*, or so I shall argue. Among philosophers of technology, Don Ihde stands out as one of the most thorough analysts of how humans relate to the world via technology. In particular, he develops a taxonomy of *human–technology–world relations*, at the core of which is a distinction between what Ihde calls *embodiment relations* and *hermeneutic relations*.[10] Though developed primarily with technological artefacts in mind, this distinction, as we shall see, also provides a fruitful framework for analyzing how users access and explore the world using scientific models. Before this framework can be developed further, however, a brief characterization of Ihde's twin notions of *embodied* versus *hermeneutic* human–technology–world relations is in order.

When we interact with technologies through *embodiment relations*, technologies are already incorporated *into* our experience 'in a particular way by way of perceiving *through* such technologies and through the reflexive transformation of [our] perceptual and body sense' [p. 72].[11] Optical technologies—such as glasses, binoculars, or telescopes—are a good example of this, since they literally allow

[9]Quoted after [24, p. 184].

[10]Ihde uses the slightly clumsy expression 'intentionality relations' as an umbrella term and distinguishes between embodiment, hermeneutic, alterity, and background relations. See [27].

[11]Unless otherwise stated, all page numbers in this and the next paragraph refer to [27].

their user to 'see through' the mediating layer of technology and, in the case of wearing glasses, are treated as de facto extensions of one's sensory organs. In such cases of embodiment, tools and technologies are not given focal attention *as technologies*; instead, they 'become part of the way I ordinarily experience my surroundings; they "withdraw" and are barely noticed' [p. 73]. Typically, embodiment relations require integration across multiple sensory modalities, along with a high degree of coordination between a technological device and its user. A good example is the technologically mediated activity of driving: 'One embodies the car, too, in such activities as parallel parking: when well embodied, one feels rather than sees the distance between car and curb—one's bodily sense is "extended" to the parameters of the driver-car "body"' [p. 74]. When it comes to technologies that essentially depend on embodiment—for example, in the case of remote manipulators, prosthetic devices, or endoscopic probes—it is clear 'that the design perfection is not one related to the machine alone but to the combination of machine and human' [p. 74].

By contrast, *hermeneutic relations* involve 'a special interpretive action within the technological context' that 'calls for special modes of action and perception, models analogous to the reading process' [p. 80]. Reading—in the sense of reading of a text or reading an instrument—'is a specialized perceptual activity and praxis', which involves one's body and sensory-motor system 'in certain distinctive ways' [p. 81]. At the same time, the target of one's attention—the text, formula, chart, or dial—is typically assumed to stand in a representational relation to an aspect of the world:

> If the object-correlate, the "text" in the broadest sense, is a chart, as in the navigational examples, what is represented retains a representational isomorphism with the natural features of the landscape. The chart represents the land- (or sea-)scape and insofar as the features are isomorphic, there is a kind of representational "transparency." The chart in a peculiar way "refers" beyond itself to what it represents. [p. 81]

As Ihde's choice of example—a graphical chart—indicates, the hermeneutic dimension of our interaction with representations goes well beyond interacting with linear text that is given in a natural (or formal) language. When we deal with language, it is clear that we cannot do without interpretation, since linguistic representations depend essentially on shared rules and conventions. Yet even graphical representations like a navigational chart require some hermeneutical work: *reading* a map is an activity that is structured by assumptions about the veracity of the representation, by conventions about the interpretation of the various graphical elements (e.g., lines, colours, symbols), and by an awareness of the interpretive process itself. While there exist similarities between the reading of a map and, say, the aerial viewing of the charted territory, one crucial difference (among others) is the fact that 'during the act of reading, the perceptual focus is the chart itself, a substitute for the landscape' [p. 81]. Readable technologies—including such examples as measurement instruments, computing devices, dials, gauges, charts, maps, etc.—tell us about the world, not by allowing us to access it immediately, but by serving as an interface with it.[12] Both the printed map and the handheld telescope are visual technologies of sorts; but whereas in the case of a telescope, we can 'become one with' and, through skilled embodied use—as an *extended* self, we might say—look *through* it *at* the world, in the case of the map the representational medium itself occupies the focal point of our attention: when we read a map, we are looking *at* the map, not *through* it *at* the world; at best, we might be said to be encountering the world indirectly, as a composite technology–world system.

One might expect the distinction between *immediacy* of access and *directness* of representation to map onto the distinction between embodiment and hermeneutic relations. After all, the attraction of highly accessible models such as His's wax models consists in their seemingly replicating putative target systems so naturalistically, and along the same sensory modalities (visual, tactile, etc.), that they may be treated almost as though they were individual specimens of the kinds that they represent. However, even when the representational means employed by the model ensure a high degree of immediacy, it does not follow that we can access the 'world out there' in an embodied fashion; modeling, after all, is often a highly indirect way of representing the world. And yet, there is a sense in which we can more immediately gain knowledge about an organism via a life-like wax model than if we were to consult, say, a sequence of diagrams and data points. One potential explanation of this difference lies in the fact that visual and tactile media more easily support our entering into a 'game of make-believe'.[13] When faced with a wax replica of an embryo, we can successfully pretend that what is, in fact, a stand-in—the material model—may, within the context of model-based inquiry, be treated as 'the real thing': that is, the target. For example, if we inspect the wax model closely and find what looks like a distinct anatomical structure in the object in front of us, we are to assume that *the embryo*—not just the lump of wax we are examining—really does exhibit this feature at the particular stage of development that is being represented.[14] What makes it possible to treat *some* models in this way is the fact that they successfully replicate the empirical richness of the original and that they utilize the same sensory modalities as the target system to convey information about them; His's wax models invite us to pretend that, *if* we were to look at an actual embryo, *this* is what we would see. The case, then, is similar to the embodiment relations

[12]Patrick Heelan, in a similar vein, notes that in natural and artificial languages, while 'the meanings of words and sentences are given *directly*, that is, *non-inferentially*, they are nevertheless *mediated* by the forms and structures of the words used' [42, p. 190]. Heelan extends this analysis to the structured output of measurement instruments, which a scientist may routinely (and, over time, effortlessly) use in order to learn about the world and about the unobservable theoretical entities it contains. Thus, a scientist might use 'standard scientific instruments, and when these function for a perceiver as readable technologies, they become what Merleau-Ponty calls "detachable [sense] organs", and they make possible new forms of perception capable of detecting particulars of the kinds named by the scientific theory' [42, p. 193].

[13]See the discussion in Chap. 1 (Sect. 1.5).

[14]Recall that, on the 'models as make-believe' account, physical models may be regarded 'as props in games of make-believe, which represent their objects by prescribing imaginings about them' [40, p. 82].

discussed earlier, with the exception that, for as long as the user immerses himself in the world of the model, it is the model which ‘becomes’ the target system. There is no contradiction between this state of make-believe immersion and the tacit acknowledgment that the model is simply a proxy for the target system. Indeed, it will often be necessary to suspend the make-believe attitude and look at the model ‘from the outside’, as it were—treating it as an *artefactual representation* rather than as a *substitute target*. Equally importantly, it takes more than just a decision on the part of the model user for such an embodied state of immersion to be possible. In particular, the model must be so constituted as to allow for the suspension of the usual distinction between model and target.

So far, our focus has been on material models, in particular those that aim to replicate aspects of their target systems such that they allow their users to immerse themselves in the world of the model by treating it *as though* it were the target. What about models that lack the requisite sensory continuity between the representational means and the material qualities of the target system? In previous chapters, a great deal of the discussion was devoted to mathematical models. Such models, arguably, are very different in character from material objects like His’s wax models. While the view of models as epistemic tools (see Sect. 5.4 above) has successfully emphasized that mathematical models, too, must draw on external representational means, not even the defenders of this view would argue that the types of representational means typically employed in such a cases—e.g. diagrams on paper, symbol systems, etc.—are of a kind that would be able to sustain anything like an embodied state of immersion on the part of the model user. Equations on a piece of paper or Feynman diagrams on a blackboard do not replicate their target systems along the same sensory modalities; rather, they represent aspects of these systems in a symbolic language. As a result, when dealing with mathematical models, the hermeneutic dimension of the user–model–target relationship will prevail. Given that hermeneutic relations differ in their phenomenology from embodied relations insofar as the medium itself becomes a more salient focus of attention, one might worry that this might make extracting information more difficult. But lack of immediacy need not come at the expense of epistemic utility. On the contrary: once we have mastered the requisite rules of interpretation, a hermeneutic technology holds out the promise of giving us more precise—and more explicit—information than its embodied analogue. Thus, when we read a thermometer, we can learn the precise temperature of our surroundings in a more reliable and explicit manner than if we were to rely on the embodied (felt) sense of hot and cold. Precisely because hermeneutic technologies can single out relevant bits of information from the totality of our experience as a whole, they allow for a precise and reliable uptake of information, even if at the expense of a somewhat impoverished phenomenology: ‘There is an instantaneity to such reading, as it is an already constituted intuition (in phenomenological terms)’ [27, p. 85]. It is also worth noting that the act of reading need not always be strenuous, even in the case of mathematical equations or diagrams: just as reading and interpreting linear text has, for most of us, become ‘second nature’, so the hermeneutic activity of reading an equation or writing down a series of Feynman diagrams, for those who use them on a daily basis, may over time become almost effortless.[15]

Earlier in this section, I claimed that, similar to the way tools afford their users different ways of accessing and manipulating the world, scientific models enable different forms of user–model–target relations. This claim can now be made more precise. For one, the preceding discussion has identified the two central types of relations in question: embodiment relations and hermeneutic relations. Different types of models have greater affinities with one or the other: material models lend themselves more readily to embodied forms of interaction, while mathematical models require specific modes of interpretive action, more akin to reading. But in real-life cases of scientific models, this is rarely an either/or affair: as proponents of the models-as-mediators account have convincingly shown, models 'are made up from a *mixture* of elements' [28, p. 14], and this heterogeneity often entails that some parts of a model may be continuous with our ordinary sensory modalities, whereas others require significant interpretation. More often than not, actual scientific models will be characterized by a coexistence of embodied and hermeneutic user–model–target relations. Consider the case of scale models in engineering, which are commonly used to study the structural, mechanical, and/or aerodynamic features of their targets.[16] An engineer designing a new type of aeroplane might begin by constructing a model that has the appearance of the full-scale aircraft, including its geometrical proportions, only to find that not all relevant properties (such as drag, weight, friction etc.) scale proportionately with size; in such cases, one would need to suspend immersive engagement with what looked to be a good stand-in for the target system and 'read' the model in a more detached way: for example, by taking measurements, making appropriate modifications (e.g. adjusting the relative wing size), or adding further elements (e.g. additional background assumptions) to it. Working with models often requires such 'switching' between embodied and hermeneutic modes of interaction. This leads to the second modification of my general claim: not only do scientific models support different types of user–model–target relations, but they often *enable* their users to switch back and forth between them. In short, models may function as *mediators between different user–model–target relations*.

In order to substantiate this claim, let us consider two 'mixed' cases—that is, instances of models which combine features that contribute to a more embodied model–user interaction with elements that require a more interpretive stance. As we shall see, each case requires some degree of switching of the sort discussed in the previous paragraph, not because of any malfunctioning of the model, but simply as a matter of successfully using the model as a way of deriving knowledge from it. The examples in question are the Phillips machine, developed in the late 1940s as a

[15]So much so that, as discussed in Chap. 4 (Sect. 4.1), scientists often speak of having developed 'a feel for' a model, even in the case of highly abstract mathematical models.

[16]For an insightful discussion of scale models in engineering and their dependence on similarity principles, which supply criteria for characterizing situations and translating information from one situation to another, see [41].

hydraulic model of the economy, and the use of interactive computer graphics in contemporary protein modeling. A full discussion of these examples would be well beyond the scope of this book; here, I shall restrict myself to certain aspects of how users interact with each model and how, in each case, they are able to acquire knowledge about the model's putative target system.[17]

Described variously as 'legendary', 'unique', and altogether 'best seen rather than described' [29, p. 21], the Phillips machine was constructed by William Phillips in 1949 as a hydraulic model of the macro-economic relationships between stocks and flows in an open economy. The machine operates via the flow of water (laced with dye to allow for easier observation), which is pumped up from a tank (named 'transaction balances') and then makes its way—representing the flow of income—through a system of glass tubes, valves, and tanks (which represent stocks, taxes, and money flows in the economy). Various elements of the model can be adjusted by its user: for example, the experimenter can adjust how much of their own disposable income individuals spend, thereby modeling a range of different scenarios. Though thoroughly mechanical (and therefore non-digital) in its operation, it can nonetheless be 'programmed', as it were, by changing its settings; it may thus function either as a model of Keynesian economic theory or as a neoclassical model, depending on the impact its settings allow fiscal policy (as represented in the model) to have on aggregate demand and output. From the vantage point of economic theory today, which puts a premium on highly abstract models, the Phillips machine may seem eccentric, inasmuch as it was not formulated in the language of mathematics, but instead 'involved metal, liquid, plastic, electricity and glass' [30, p. 118]. But its materiality is key to how the machine models economic processes. For example, the viscosity of water, which affects the speed with which it flows through the model, introduces a time lag that reflects 'the necessary time gap, or lag, in the interactions between investment and income' [6, p. 210].

To the uninitiated user, the Phillips machine may look like a curiosity; without reading the labels and instructions, it is certainly not obvious how to make sense of its various flows, water levels, and other elements. Clearly, then, some degree of interpretive 'reading' of the machine is required: although the materiality of the model is just as important to the Phillips machine as it was for His's wax models, in the former case it does not readily support an immersive experience on the part of its user. Or does it? To be sure, we cannot expect any material model of the economy to represent its target system with the same degree of immediacy as a wax replica of an organism is able to represent its target. But the Phillips model is often credited with making vivid—in a manner that approaches the degree of embodiment and accessibility afforded by other material models—otherwise abstract relationships and difficult-to-grasp processes. The economist David Vines argues that, while '[i]t is not true that "everything is in the machine" […] there is in fact

[17]Readers interested in analyses of these models in their own right may wish to turn to [6], esp. Chap. 5, and [32] for the Phillips model, and to [33, 34] for the use of interactive computer graphics in protein modeling.

more in the machine […] than is allowed in macroeconomic conventional wisdom. *And it is immensely visible*' [31, p. 49; italics original.]. Others concur with this assessment, noting that '[a] student of economics can never forget the difference between an economic stock and flow after seeing the machine in operation' [29, p. 21], and the same applies to the more complex operations it performs: 'The Phillips machine is capable, not simply of representing these concepts visually, it can also vividly show the inter-relationships between all these aggregates' [29, p. 26]. As Mary Morgan and Marcel Boumans put it: 'Phillips built the machine to learn and understand economics through his eyes and fingers rather than through mathematics and word' [32, p. 388]. Rather than viewing the Phillips machine as a way of merely *illustrating* macroeconomic relationships, as some of its early detractors did, perhaps it would be more apt to say that, once the machine has been set in motion, it *performs* these relationships in such a way that users can access them with greater immediacy, and in a more embodied fashion, than would be possible by relying on symbolic means of representation alone.

The second example of how users may gain knowledge from a model by moving back and forth between an embodied and a hermeneutic stance is the increasingly widespread use of interactive computer graphics in protein modeling. Modeling the structure of proteins is an exceedingly difficult task and requires not only significant background knowledge, but also imagination and considerable skill. Before the advent of computer technology, figuring out the structure of a particular protein required the construction of material models, as in the already mentioned case of Pauling's use of paper models in the run-up to his discovery of the α-helix. In such cases, the embodied dimension of models is evident. As Natasha Myers puts it in her study of molecular 'body-work' in protein crystallography:

> The 'haptic' dimension involved in the manipulation and handling of physical materials was key for the production of models that could give modelers a sense of the structure and dynamics of the molecule, and offered a means for researchers to use their bodies to incorporate structural knowledge. [33, p. 174]

One of the difficulties of modeling the structure of a protein lies in the fact that there is no straightforward way of predicting its three-dimensional shape on the basis of the linear sequence of amino acids. A sequence of amino acids will 'fold' into the most energy-efficient three-dimensional structure, yet determining this structure involves running numerically demanding simulations which, in turn, requires the extensive use of computer technology. In the early days of computer-aided protein modeling, researchers would write computer programmes for specific problems (e.g. specific molecules) and such code would, on occasion, be passed from one lab to another. Molecular structure, within this framework of modeling, is then represented by the spatial coordinates of the individual atoms (or groups of atoms), and the overall shape of the protein can be 'read off' only by decoding this information: 'In the absence of any other form of interface, the interaction between user and computer was in the form of these numerical representations' [34, p. 413]. In its early days, and to a lesser extent also today, computational protein modeling thus involves an ineliminable hermeneutic element in the way users access potential molecular structures.

Yet, in recent years, thanks to ever-greater computing power, more and more sophisticated visualization and modeling techniques have developed, which are sometimes referred to collectively as *molecular graphics*. Integrated software packages now offer their users full-scale simulation and visualization environments, in which users can simply 'drag and drop' atoms (or molecular groups such as amino acids) on a computer screen, while an algorithm calculates likely spatial arrangements in real time. These can then be displayed, either in vector graphics format or as space-filling models which can be rotated and manipulated using various computer-based commands, and which may be further visually enhanced using virtual lighting techniques such as ambient occlusion. As Myers describes it, onscreen protein modeling involves pulling up multiple windows on the computer and moving back and forth between 'restless gestures of the mouse and the quick-paced and sometimes clumsy tapping out of keyboard commands':

> In one window, data will be streaming up the screen, and in another, the crystallographer holds the skeleton-like interactive rendering of a model. She keeps it alive in space and depth, rotating it onscreen and zooming in and out, keeping it visible at multiple angles, constantly shifting her visual and haptic relationship to it. [33, p. 179]

It is such rapid shifting between different ways of relating to the model—from visual to haptic and, one might add: from embodied to hermeneutic—that enables protein modelers to 'learn how to see, feel, and build protein structures through their embodied interactions' and 'to acquire their "feeling for the molecule"' [33, p. 181].

Whereas the heterogeneity of models in science and the diversity of their uses and functions are nowadays widely acknowledged, what has perhaps been overlooked is that not only do models come in various forms and shapes and may be used for all sorts of purposes, but they also give unity to this diversity by mediating not just between theory and data, but also between the different *kinds of relations* into which we enter with the world. Models, then, are not simply neutral tools that we use at will to represent aspects of the world; they both constrain and enable our knowledge and experience of the world around us: models are mediators, contributors, and enablers of scientific knowledge, all at the same time.

References

1. M. Morrison, M.S. Morgan, Models as mediating instruments, in *Models as Mediators: Perspectives on Natural and Social Science*, ed. by M.S. Morgan, M. Morrison (Cambridge University Press, Cambridge, 1999), pp. 10–37
2. M. Morrison, Models as autonomous agents, in *Models as Mediators: Perspectives on Natural and Social Science*, ed. by M. Morgan, M. Morrison (Cambridge University Press, Cambridge, 1999), pp. 38–65

3. M. Morrison, *Reconstructing Reality: Models, Mathematics, and Simulations* (Oxford University Press, New York, 2015)
4. M. Boumans, Built-in Justification, in *Models as Mediators: Perspectives on Natural and Social Science*, ed. by M.S. Morgan, M. Morrison (Cambridge University Press, Cambridge, 1999), pp. 66–96
5. M. Boumans, *How Economists Model the World into Numbers* (Routledge, Abingdon, 2005)
6. M.S. Morgan, *The World in the Model: How Economists Work and Think* (Cambridge University Press, Cambridge, 2012)
7. M. Morrison, Modelling nature: between physics and the physical world. Philosophia Naturalis **35**(1), 65–85 (1998)
8. M. Hesse, Models in physics. Br. J. Philos. Sci. **4**(15), 198–214 (1953)
9. L. Magnani, Reasoning through doing. epistemic mediators in scientific discovery. J. Appl. Logic **2**(4), 439–450 (2004)
10. G. Gramelsberger, Symbol systems as cognitive and performative hybrids: a reply to Axel Gelfert. Social Epistemology Review and Reply Collective **4**(8), 89–94 (2015)
11. D. Kaiser, Physics and Feynman's diagrams. Am. Sci. **93**(2), 156–165 (2005)
12. J. Kuorikoski, P. Ylikoski, External representations and scientific understanding. Synthese, 1–21 (forthcoming)
13. R.J. Baxter, Exactly Solved Models in Statistical Mechanics (Academic Press, New York, 1982)
14. F. Gebhard, *The Mott Metal-Insulator Transition: Models and Methods* (Springer, Heidelberg, 1997)
15. A. Gelfert, Rigorous results, cross-model justification, and the transfer of empirical warrant. Synthese **169**(3), 497–519 (2009)
16. M. Boon, T. Knuuttila, Models as epistemic tools in engineering sciences: a pragmatic approach, in *Philosophy of Technology and Engineering Sciences (Handbook of the Philosophy of Science)*, vol. 9, ed. by A. Meijers (Elsevier, Amsterdam, 2008), pp. 687–720
17. T. Knuuttila, Modelling and representing: an artefactual approach to model-based representation. Stud. Hist. Philos. Sci. **42**(2), 262–271 (2011)
18. B. Preston, Philosophical theories of artifact function, in *Philosophy of Technology and Engineering Sciences (Handbook of the Philosophy of Science)*, vol. 9, ed. by A. Meijers (Elsevier, Amsterdam, 2008), pp. 213–233
19. D. Baird, *Thing Knowledge: A Philosophy of Scientific Instruments* (University of California Press, Berkeley, 2004)
20. J.D. Watson, *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*. (Norton, New York, 1981 [1968])
21. H.-J. Rheinberger, Preparations, models, and simulations. Hist. Philos. Life Sci. **36**(3), 321–334 (2015)
22. M. Delehanty, Why images? Medicine Studies **2**(3), 161–173 (2010)
23. J. Kulvicki, Knowing with images: medium and message. Philos. Sci. **77**(2), 295–313 (2010)
24. N. Hopwood, Plastic publishing in embryology, in *Models: The Third Dimension of Science*, ed. by S. de Chadarevian, N. Hopwood (Stanford University Press, Stanford, 2004), pp. 170–206
25. N. Hopwood, 'Giving body' to embryos: modeling, mechanism, and the microtome in late nineteenth-century anatomy. Isis **90**(3), 462–496 (1999)
26. R. Wiedersheim, Notiz [of Adolf Ziegler's series after Philipp Stöhr's models on the development of the skull]. Zoologischer Anzeiger **2**, 545–546 (1882)
27. D. Ihde, *Technology and the Lifeworld* (Indiana University Press, Bloomington, 1990)
28. M. Morrison, M. Morgan, Models as mediating instruments, in *Models as Mediators: Perspectives on Natural and Social Science*, ed. by M.S. Morgan, M. Morrison (Cambridge University Press, Cambridge, 1999), pp. 10–37
29. R. Moghadam, C. Carter, The restoration of the Phillips machine: pumping up the economy. Econ. Aff. **10**(1), 21–27 (1989)

30. R.M. Goodwin, A superb explanatory device, in *A.W.H. Phillips: Collected Works in Contemporary Perspective*, ed. by R. Leeson (Cambridge University Press, Cambridge, 2000), pp. 118–119
31. D. Vines, The Phillips machine as a 'progressive' model, in *A.W.H. Phillips: Collected Works in Contemporary Perspective*, ed. by R. Leeson (Cambridge University Press, Cambridge, 2000), pp. 39–67
32. M.S. Morgan, M. Boumans, Secrets hidden by two-dimensionality: the economy as a hydraulic machine, in *Models: The Third Dimension of Science*, ed. by S. de Chadarevian, N. Hopwood (Stanford University Press, Stanford, 2004), pp. 369–401
33. N. Myers, Molecular embodiments and the body-work of modeling in protein crystallography. Soc. Stud. Sci. **38**(2), 163–199 (2008)
34. E. Francoeur, J. Segal, From model kits to interactive computer graphics, in *Models: The Third Dimension of Science*, ed. by S. de Chadarevian, N. Hopwood (Stanford University Press, Stanford, 2004), pp. 402–429
35. G. Fisher, The autonomy of models and explanation: anomalous molecular rearrangements in early twentieth-century physical organic chemistry. Stud. Hist. Philos. Sci. **37**(4), 562–584 (2006)
36. A. Gelfert, Mathematical formalisms in scientific practice: from denotation to model-based representation. Stud. Hist. Philos. Sci. **42**(2), 272–286 (2011)
37. A. Toon, The ontology of theoretical modelling: models as make-believe. Synthese **172**(2), 301–315 (2010)
38. R.B. Griffiths, Rigorous results and theorems, in *Phase Transitions and Critical Phenomena*, ed. by C. Domb, M.S. Green (Academic Press, New York, 1972), pp. 8–109
39. A. Gelfert, J. Mok, Saving Models from Phenomena: A Cautionary Tale from Membrane and Cell Biology, in *Integrated History and Philosophy of Science. Proceedings of the 5th Conference*, ed. by F. Stadler. (Springer, Dordrecht, forthcoming)
40. A. Toon, Models as make-believe, in *Beyond Mimesis and Convention: Representation in Art and Science*, ed. by R. Frigg, M.C. Hunter (Springer, Dordrecht, 2010), pp. 71–96
41. S.G. Sterrett, Models of machines and models of phenomena. Int. Stud. Philos. Sci. **20**(1), 69–80 (2006)
42. P. Heelan, Interpretation and the structure of space in scientific theory and in perception. Res. Phenomenol. **16**(1), 187–199 (1986)

Index

A. Gelfert, *How to Do Science with Models*, SpringerBriefs in Philosophy,
DOI 10.1007/978-3-319-27954-1