Biol Philos (2006) 21:725–740
DOI 10.1007/s10539-006-9054-6

ORIGINAL PAPER

The strategy of model-based science

Peter Godfrey-Smith

Received: 6 September 2005 / Accepted: 30 September 2005 /
Published online: 13 December 2006

Introduction

My title refers to Richard Levins' famous paper on models in population biology (1966). Here Levins presented his three-way distinction between kinds of model-building, and also introduced a set of more fundamental ideas about trade-offs that constrain and guide scientific work. For Levins, these trade-offs derive from the relationships between three different theoretical goals: realism, precision, and generality.

The talk of "strategies" within Levins' paper concerns alternative strategies *within* the enterprise of model-building. My topic here is broader; I will treat models and model-building as characteristic of one particular approach to theorizing, a *strategy of model-based science*.

The ideas presented here are indebted to Michael Weisberg's work (forthcoming, 2003), and to discussions with him at Stanford. They are also indebted to a scattered tradition of other philosophical and scientific work, especially by Levins, Giere, and Wimsatt. The aim of the paper is to give a general sketch of model-based science as I see it, but a sketch focusing particularly on divergences between myself and those on whose work I draw.

The term "model" is surely one of the most contested in all of philosophy of science. What sense of the term do I have in mind? My starting point is the account given by Giere (1988). For Giere, models are idealized structures that we use to

Presented at a conference on Richard Levins at the University of Pennsylvania, 2005.

P. Godfrey-Smith (✉)
Department of Philosophy, Harvard University, 208 Emerson Hall, Cambridge, MA 02138, USA
e-mail: pgs@fas.harvard.edu

represent the world, via resemblance relations between the model and real-world target systems.[1] Various modifications to Giere's picture will be made along the way, and in some respects Giere's view itself will be treated as an idealized model of one kind of science. But Giere's view is uniquely useful as a point of departure. My aim is to treat the use of models, in roughly Giere's sense, as enabling a particular approach to representing and understanding the world. What is most distinctive of model-based science is a strategy of *indirect* representation of the world—here I agree with Weisberg. The modeler's strategy is to gain understanding of a complex real-world system *via* an understanding of simpler, hypothetical system that resembles it in relevant respects. This is an approach with both strengths and weaknesses, with effects on the sociology of scientific work, and perhaps with a distinctive historical signature.

The examples used to motivate the analysis will be drawn almost exclusively from biology. This reflects both limitations on my expertise and the occasion for which the paper was written. The phenomenon I am trying to describe—model-based science as a strategy—is well-marked within biology. It is also clear within psychology and at least some of the social sciences. Weisberg has developed ideas of this kind in the context of chemistry. It is possible that things may look different in the case of physics itself. Giere's view, which is used as my starting point, was specifically developed for the case of physics, but Giere's handling of the core ideas is somewhat different. In any case, as I will criticize those who make simple extrapolations from the physical sciences to the biological, it will be wise to resist simple extrapolations in the other direction.

Model-based science, the semantic view, and mental modeling

When models are disussed in the philosophy of science, people usually think first of "the semantic view of theories." The analysis of model-based science I will give is intended to be distinct from the leading ideas of the semantic view, and the contrasts between the two are essential to my project.

If the word "model" makes people think first of the semantic view, they will probably think second of a project in the cognitive science of science that uses the notion of a "mental model" to give an account of conceptual change (Nersessian 1999). That second project is closer in spirit to my view than the semantic view is, but once again the contrasts are significant. Nersessian is trying to describe something that may well be real and important, but which is also narrower and more psychologistic than the phenomenon I focus on.

I will discuss the semantic view of theories first. The semantic view is an attempt to represent "the structure of theories" using the notion of a model (Suppes 1960; Suppe 1977; Van Fraassen 1980). The semantic view is often treated as something of a "big tent" in the philosophy of science, but I will regard the following two ideas as central to the view:

[1] In this paper I will not worry about the case of actual physical models constructed in the lab, such as model wings in wind tunnels. These seem to me to be of secondary importance, when they are truly distinct from the kinds of models discussed here. Some other treatments of models in philosophy of science take care to include these cases, however (e.g., Griesemer 1990).

(i) The analysis is applicable to *all* scientific theories.
(ii) The sense of ‘‘model’’ that is applied is either the logician’s sense, or something relevantly close to it.

Both of these ideas are rejected in the treatment of model-based science developed here. The second claim is one that some people within the semantic view have had doubts about for some time. The first claim seems to have been rarely questioned. I will comment in some detail on each.

The semantic view began as an application of meta-logic and set theory to the general structure of theories (Suppes 1960). The initial role played by the idea of a model was derived from the role of this concept in those formal disciplines. A model, basically, is a set of objects (and relations between them) that functions as an *interpreting structure* for a set of sentences. A model is used as something for a set of sentences to be *true of*.[2]

When models are treated in this way, they can be used to help us understand *any* theories, any sentences, perhaps any representations with indicative content at all. They can be used to think formally about scientific theories, and they can be used to think formally about the sports pages of the *New York Times*. What we have here is a general way of thinking about meaning; we think about the meaning of a representation by thinking about the formal characteristics of structures that the representation can be seen as true of.

This is a way of thinking about meaning that is based on the notion of a certain kind of interpretive act, the act of *treating* a representation as about some domain of objects. In this kind of interpretation, we ignore the question of natural-world connections between the representation’s physical instantiation, and the domain of objects in question. When treated as the basis of a complete theory of meaning, this approach rapidly leads to serious trouble (Putnam 1982; Lewis 1984). But the framework does have utility in some contexts. For advocates of the semantic view, a key contrast is between the semantic view of theories and an older language-bound view of theories, as ‘‘syntactic’’ objects.[3] When *that* contrast is laid before us, the appeal of the semantic view is understandable. But this contrast, which drives a lot of the early work on the semantic view, is one that only arises within a strong commitment to the project of highly formal analysis of scientific theories. The question is one about which *kind* of formal approach is to be taken, a ‘‘syntactic’’ or ‘‘semantic’’ approach. For philosophers like Suppes, the development of model theory and set theory provided a superior formal framework.

After a while, philosophers began to make ‘‘the semantic view’’ more responsive to features of actual scientific work. Scientists themselves are interested in things called ‘‘models,’’ and these usually seem to be hypothetical objects of some kind. Suppes had claimed in 1960 that the meta-logical concept of model could be used to describe all the things called ‘‘models’’ by scientists, but later work started to discuss models in ways that more closely resembled the ways scientists actually talk about them. So this later work continued the general philosophical project of the semantic

2 See also Thompson-Jones (2006) for a careful and useful discussion of these aims; he makes a further distinction within this general project, between richer and more trivial senses in which something can function as a truth-making structure.

3 French and Ladyman (1999) is a more recent defence of the semantic view that emphasizes, again, the importance of language-independence.

view, but using an account of models that aimed to be closer to scientific practice.[4] This looked like progress, but the result was really a growing equivocation on the word "model."

This problem was ably diagnosed in a (1992) paper by Stephen Downes. Downes unpacked the various senses of "model" that have been used in the literature, and argued that neither the logician's sense nor any other comparably general sense could be used to unify all the diverse discussions of models that had come to be associated with "the semantic view." The semantic view had gone wrong in trying to simultaneously see "models" as an important real-world scientific tool, *and* as a concept that could be used in an abstract way to describe all of theoretical science.

Downes called his eventual position a "deflationary semantic view." This was the minimal claim that modeling, of various kinds, is an important tool in theoretical science. There is also a tendency, in more recent discussion, to treat the semantic view itself in a "big tent" sort of way; almost any philosophical enthusiasm for models or modeling is categorized as part of the same tradition as Suppes. We should not be too concerned about the labels, and various ideas developed by proponents of the semantic view have real utility.[5] But it should now be clear that my analysis of model-based science is intended to be neither "semantic" (in the sense of drawing on model theory) nor a "view of (all) theories."

We can also improve our "fix" on the phenomenon I aim to describe via a contrast with another treatment of models in science, the analysis of "model-based reasoning" given by Nersessian (1999) and others. Nersessian's primary aim is to give an account of the psychological mechanisms involved in particular kinds of change in science, which she and others call "conceptual" change. Here she draws on a large body of work in cognitive psychology (Johnson-Laird 1983; Gentner 2002). Nersessian argues that the distinctive mental mechanisms that this kind of psychological work has tried to describe may have great importance in understanding science.

This kind of psychological theorizing is likely to be an important component in an account of model-based science as I conceive it. But I emphasize the term "component". The aim of this work is to describe a psychological process that is at most one part of model-based science in my broader sense. Model-based science, in the broader sense, has sociological and formal features, as well as psychological ones. What is distinctive about model-based science is a particular approach

[4] For me, this quote in Suppes (1960) exemplifies where things went wrong. "It is true that many physicists want to think of a model of the orbital theory of the atom as being more than a certain kind of set-theoretical entity. They envisage it as being a very concrete physical thing built on the analogy of the solar system. I think it is important to point out that there is no real incompatibility in these viewpoints. To formally define a model as a set-theoretical entity which is a certain kind of ordered tuple consisting of a set of objects and relations and operations on these objects is not to rule out the physical model of the kind which is appealing to physicists, for the physical model may simply be taken to define the set of objects in the set-theoretical model." (Suppes 1960, p. 290)

[5] The "state-space" framework, for example, has been used extensively in philosophy of physics, and it can also be fruitfully applied to questions about evolutionary models in biology (Lewontin 1963; Lloyd 1988; Godfrey-Smith and Lewontin 1993).

to representing and understanding the world, not a particular kind of cognitive change.[6]

But is there really a genuine phenomenon of interest, that essentially involves models, lying ‘‘midway’’ between the over-broad semantic view and the psychologistic project of Nersessian? The idea that there is a genuine phenomenon in this location, a ‘‘natural kind’’ within the larger domain of theoretical science, is a guiding idea of this paper. How can we show that such a phenomenon is real?

Showing that there is *something* real and distinctive in that location is easy to do. Some scientists now are trained, hired, and assessed as modelers; that is their job description. Modelers have their own subculture within science, to some extent, and their own language. But not all scientists who engage in theoretical work are in this category, and some are quite resistant to being described this way. Surely this is enough to show that model-based work is a proper subset of scientific theorizing?

For some philosophers, it will not be enough. In response to the evident sociological facts about the development of modeling as a distinctive style of science, a philosopher might argue that the difference between what gets labeled ‘‘modeling’’ and what gets described differently is an epistemologically shallow one. Perhaps the term ‘‘modeling’’ is used in a context-sensitive way to pick out whatever theorizing in a discipline happens to be especially abstract, speculative, or provisional. Once we are in the business of giving a philosophical account of science, as opposed to attending merely to the sociology, it might be argued that there is little difference between these alleged varieties of theoretical work, and they can receive a uniform philosophical treatment. This is how I imagine an advocate of the ‘‘semantic view,’’ as traditionally conceived, might respond to my arguments as presented so far.

Replying to this challenge, and related ones, will be the task of the next section. There I use a particular pair of examples to argue that the distinction between model-based work and other theorizing is not a superficial one in this way. Before turning to the example, though, I will finish this section by clarifying how the distinction between model-based and other scientific work might fit into an overall philosophical picture.

The strategy of modeling is one strand or current in the larger history of science. In some contexts it shows up in a quite ‘‘pure’’ form, as self-conscious modeling of a kind that has its own skill-set, subculture, and language. What we see in those cases may be the scientific elaboration and formalization of a more general and psychologically deep capacity for model-based understanding. (Here the project of Nersessian and others, sketched above, may be very important.) But this capacity is one of several that have contributed to the development of theoretical science. Much scientific work may be influenced by, may draw on, several disparate investigative strategies at once. In much scientific theorizing, we may find a complex mixture, or a to-and-fro movement, between the strategy of modeling and other, more ‘‘direct’’ approaches to representing and understanding the world.

[6] Another treatment of models in science, distinct from both the semantic view and Nersessian's project, is the ‘‘models as mediators’’ approach of Morgan and Morrison (1999), which is related also to Cartwright's work (1999). Morgan and Morrison see models as ‘‘mediating instruments,’’ partially autonomous from both theory and real-world phenomena. I am unsure how to treat the relations between their view and mine, but the idea of the ‘‘autonomy’’ of model from theory seems different from, and more specific than, the idea of a strategy of indirect represention via similarity between model and target.

Further, a recognizable *style* of model-based science (self-conscious or not) may be increasing in prominence within science, on a decade-long and perhaps centuries-long time-scale. This has eventually led to the development of a self-conscious practice of a model-driven form of science, guided by a distinctive set of ideals. This last development involves, among other things, a rejection of the idea that modeling is a mere heuristic adjuct to the real business of theory-construction. Instead, we have an embracing of the idea that models themselves can be the tools by which we represent and understand the world. Levins' work is of some importance in this historical process.

Modeling as optional: Buss versus Maynard Smith and Szathmáry

Model-based science is fundamentally a strategy of indirect representation of the world. In understanding a real-world system, the modeler's first move is the specification and investigation of a hypothetical system, or structure. The second is consideration of resemblance relations between this hypothetical system and the real world "target system" that we are trying to understand.

In much day-to-day work in model-based science, the second stage is left implicit; a lot of the everyday focus is on understanding how various interesting model systems behave. Which evolutionary scenarios in which selection acts at only one locus generate stable equilibria? Which neural networks can learn a simple grammar? The obvious strategy that contrasts with model-based science is to avoid any deliberate detour through merely hypothetical systems, and seek to directly represent the workings of the real-world system we are trying to understand. The details of the role of hypothetical model systems, and their status, will be discussed in the section to follow this one. First, however, I will use some biological examples to support the claim that there *is* a real contrast between model-based theorizing and other kinds of theorizing. This case is intended to make clearer the fact that treating all theorizing as involving models, in the style of the "semantic view," is misguided.

Two of the most important books in evolutionary theory written over the last twenty years are Leo Buss' *The Evolution of Individuality* (1987) and Maynard Smith and Szathmáry's *The Major Transitions in Evolution* (1995). These works provide an ideal example here, because they were written around the same time and are partly about the same topic, but exemplify completely different theoretical methodologies. The Maynard Smith and Szathmáry book is very largely based on models. The book isolates eight "major transitions" in the evolution of life so far, and each of them is addressed via the construction of models that give us a sense of how a natural process might take us from the "before" to the "after" state associated with that transition. The transitions include the origin of life, the origin of chromosomal organization of replicators, the origin of the eukaryotic cell, the transition to multi-cellularity, the formation of adapted social groups, and the evolution of language. Much of the book focuses on the early transitions. The Buss book is concerned, roughly speaking, with the events around the middle of the Maynard Smith and Szathmáry transitions. His focus is on how coherent multi-cellular individuals arose, and remain stable, in the face of potentially subversive lower-level competition at the level of the cell lineage. The Buss book appeared first and is acknowledged as a key influence by Maynard Smith and Szathmáry.

The strategies employed by the two books could hardly be more different. Buss' approach is based on a broad and detailed examination of the *actual* relations between cellular reproduction and whole-organism reproduction in known organisms. He sets up the problem by noting that what are usually seen as familiar, routine facts about the "sequestration" of germ cells in large multi-cellular organisms, like humans, are quite anomalous when we consider the whole of extant life. It is the sequestration of a specialized set of germ cells that seems to make highly adapted cell-level cooperation possible in organisms like us, as it ensures that any cell lineage that tries to "go it alone" and reproduce freely at the expense of whole-organism activities must be an evolutionary dead-end, with no way of reaching the next generation.[7] But this sequestration is rare across all of life, and must have evolved from a more anarchic state of cellular competition between and within multi-cellular collectives. The details of Buss' arguments are based on general selectionist principles, but also on a key role for constraints deriving from the nature of the cellular machinery that early organisms would have had available. An example is a constraint that would have made it hard for single-celled organisms to both move and divide at the same time, owing to competing demands on microtubule organizing centers in the cell.

There are no formal mathematical models in Buss' book. And further, there are no overt models of any other kind. Buss' entire argument is based on the causal roles and consequences of actual cellular machineries, actual environmental circumstances, and actual developmental sequences. When I say "actual," I do not mean that Buss takes himself to be infallible about such matters; many of Buss' claims are cautious and speculative, especially when they concern extinct organisms. But there is no significant role for *deliberate* consideration of fictional, idealized, or merely schematic organisms, and the distinction between cautious exposition and deliberate fiction is a crucial one here.[8] Buss' aim is to reconstruct the causal facts that connect actual present-day organisms with their ancestors. Buss' book is also important as a demonstration that, despite the high degree of mathematicization in some parts of evolutionary theory, it is possible to do leading-edge work on evolution without constructing a mathematical model. Buss' alternative is the marshalling of a synoptic knowledge of actual facts, and careful causal reasoning.

In this respect, Buss' work is similar to Darwin's. When philosophers whose sole focus is physics claim that mathematics is *the* key to scientific theorizing, the case of Darwin has always been the best immediate response. But perhaps it will be claimed that Darwin's time is long past, and he is hence of secondary importance as a case of a non-mathematical theoretician. The example of Buss shows that this kind of work is still possible. The whole language by which the "semantic view" analyzes theories is completely inapplicable to Buss. There is no state-space, no dynamic equations, and nothing even resembling these. It would be possible to *re-do* some of Buss' work using this alternative approach. But that would be to turn Buss' work into something quite different.

[7] The recently discovered "canine transmissible venereal tumour" is a fascinating exception.

[8] I can find only two exceptions in the book, two places where deliberately fictional model cases are introduced. One is a three-page passage outlining some processes in the Cambrian that are admitted as "fantasy" (pp. 79–81) and the other is a footnote in which Buss describes and extends an elegant thought-experiment by Medawar, concerning the evolution of aging (p. 143).

In the Maynard Smith and Szathmáry book, the currency of theoretical argument at each stage is the model. Interestingly, these are often not formal mathematical models, though some are. Many of the models instead proceed by describing an idealized, schematic causal mechanism, noting how it will and will not behave, and exploring plausible evolutionary paths from one situation to another. State-spaces and the like are not explicitly employed, but it would not be too hard to express at least some of the arguments in something like those terms.[9] Much of the argument could be described as the development of tightly constrained ''how-possibly'' explanations, where the systems whose behaviors are being described are idealized analogues of real-world stages in the history of life. In contrast to Buss, rather few specific organisms are named and described.[10]

This gives the Maynard Smith and Szathmáry work quite different modal properties from those of Buss' work. Maynard Smith and Szathmáry's explanations, if they work at all, would work just as well in a range of nearby possible worlds that happen to be inhabited by different organisms. Buss' explanations do have some modal ''reach,'' but are generally more closely tied to contingent features of actual organisms.

Both sides could advance criticisms of the other. Maynard Smith and Szathmáry explicitly acknowledge Buss and criticize him on points of detail (pp. 244–245). The most tendentious side of Buss' work, which treats the developmental processes of multicellular organisms as a partial *replay* of early episodes of suborganismal selection (p. 78), has been largely rejected, as far as I can tell. A more general criticism of Buss has been advanced from the side of modelers. It has been argued that at several points his treatment of within-organism competition between cells is flawed because of insufficient consideration of role of relatedness between cells, and the mathematical complexities of kin selection (Queller 1997 explicitly compares the two books on this point.) But Buss would be within his rights to argue that many of the standard scenarios treated by modelers, when addressing the transition to multicellularity, pay insufficient attention to constraints on the process deriving from the actual cellular mechanics. (Here the dual role of microtubule organizing centers is again a good example.)

So a consideration of these two books yields a number of philosophical morals. One is that it is (as it has always been) absurd to write as if *all* scientific theorizing is best understood in terms of state-spaces and equations. That is the easy moral. The more telling one in this context is that science does contain a clear strategic distinction between model-based and non-model-based theoretical work. These two books are extreme cases on each side of the divide, but they are not unimportant or marginal works within science. On the contrary, they are front-rank works. So the same scientific problem can be addressed around the same time by people with the same background theories, via the employment of completely different scientific strategies, where one strategy is based on models and one is not.

[9] Yet a third and more recent book on the same topic, Michod's *Darwinian Dynamics* (1999) is based almost entirely around formal mathematical models with well-defined state-spaces. Michod focuses on the same ''middle'' transitions as Buss.

[10] Buss' methodology is nicely encapsulated in his acknowledgements, where he thanks a teacher (J. Jackson) for teaching him to see the endless mass of detail in comparative biology as ''fodder upon which any truly general idea will ravenously feed'' (p. xi).

The status of model systems

This section and the next will proceed on the assumption that model-based science is a real phenomenon worth analyzing, and will look in more detail at how this sort of theorizing works. As indicated earlier, the most useful starting point is Giere's (1988) account. Giere himself is usually associated with ''the semantic view.'' He, like others, is too inclined to see his picture as giving an account of all science. But the basic structure of Giere's account is useable within the alternative project developed here.

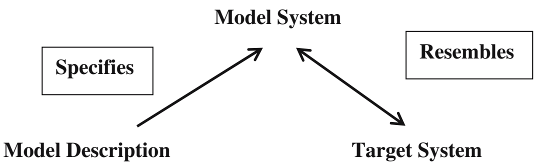
Giere's account can be summarized in a diagram (slightly modified from Giere 1988). Representation of a real-world system involves two distinct relations, the *specification* of a model system and some relevant *similarity* between model system and the world itself (Fig. 1).

So the word ''model'' tends to be ambiguous here, between what we can call the model *system* and the model *description*. The model description exists in some representational medium (mathematical formalism, words, pictures). Its role is to specify a model system, or a class of model systems. ''Specification'' and ''resemblance'' are both distinct from the usual semantic relations employed in philosophy of science. (Those relations include reference, understood via a causal or descriptive theory, and various empiricist concepts of ''cognitive significance.'')

Many of the special features of model-based science come from the role played by resemblance relations between model system and target. Philosophers tend to distrust resemblance relations because they are seen as vague, context-sensitive, and slippery. Here, those features of resemblance are indeed important, but they are not necessarily problematic. They are the source of both distinctive strengths and weaknesses of model-based science. In particular, it is important that two scientists might use same model for the same target system, but with different resemblance relations in mind. I call these different ''construals'' of the model. For one scientist, the model might operate merely as a predictive device. All that matters is its input–output profile. For another, the model is not just a device for predicting what the system will do, but a causal map of the target system that tells us something about *why* it does what it does. This more realist attitude does not present itself as a ''yes or no'' choice, or even as a single gradient. There is usually something more like a multidimensional *space* of different ways in which a model system might resemble a target. Following Giere (especially 1999), I reject attempts to give a uniform formal analysis of these resemblance relations using a concept of isomorphism, or some relative of that concept.

If the account given above describes one ''strategy'' for theorists, then what are the others? The indirect, two-step structure of model based science shows the obvious alternative: attempted *direct* representation of the target system, using whatever representational tools come to hand. There need be no routing *via* a model system. This is what Weisberg calls ''ADR'' or abstract direct representation.

Fig. 1 Modified from Giere (1988)



It is possible to put pressure on the distinction between model-based and direct representation. One might object that whatever ''representation'' might be, it is *always* indirect, never direct. A person cannot *choose* to have an intimate, unproblematic semantic contact with the world. This objection mistakes the nature of the position being defended. Whatever one's general philosophical theory of representation might be, there are strategic choices that can be made by people trying to understand a complex and unknown system. One approach is to immediately try to identify and describe the actual system's parts and their workings. A distinct approach is to deliberately describe another system, a simpler hypothetical system, and try to understand that other system's workings first. This is the contrast that we saw in a sharp form in the previous section, between the approaches of Buss on one hand and Maynard Smith and Szathmáry on the other. In response to a problem that arises, the Buss response might be ''Let's see how this works in sponges and cnidarians.'' The Maynard Smith and Szathmáry response is to imagine the simplest possible system in which the problem arises, and work from there. Philosophically, it might be a difficult task to give a good analysis of what is really going on when someone tries to ''directly'' describe a real-world system, in the face of complexity and uncertainty. But the two strategies are still distinct, and one does involve a kind of indirectness that is not seen in the other.

Still, it would be a mistake to say that the distinction is always so easy to draw. Scientists may move backwards and forwards between the two approaches very rapidly. They may insist that they are doing one thing when they seem to be doing the other. (''I am describing the actual Pacific ocean, but I suppose here that it has uniform depth.'') There are unresolved problems to tackle in this area, especially involving the role of idealization. It would be an error to associate model-based science with the presence of idealization of any kind.

A second challenge comes from the fact that work that appears to be of one kind can often be re-interpreted as work of the other kind. A person might take an attempted ''direct'' representation, and treat it as a model, for example. But this does not mean that the distinction collapses. One kind of work can indeed often be re-interpreted as the other, but this re-interpretation is itself a new move, a strategic choice of its own. Such moves can be important within science itself. One way in which an old scientific theory can be recovered and mined to inform new work is by *recasting it as a model* in my sense. Someone can take an old theory and treat it not as a failed direct description of reality, but as a specification of an interesting fictional system that can be studied in its own right. Perhaps the modern-day handling of Newtonianism is an example.

In the remainder of this section, and the next one, I will look in more detail at the status of ''model systems'' themselves. There has been a lot of uncertainty in the literature on this point. The move that most people have been tempted to make is to say that model systems are ''abstract mathematical objects'' of some kind. This general outlook is familiar from the literature on the ''semantic view,'' and it has been taken over by some writers within what I see as the alternative project of analyzing model-based science. Giere seems attracted to a view of this kind, and so is Weisberg (2003).

My aim is not to reject this idea outright, but I will argue that it is deficient in at least some cases. It is important to the practice of model-based science, at least some of the time, that model systems can be conceived and treated in a more concrete way. Roughly, we might say that model systems are often treated as ''imagined concrete things''—things that are imaginary or hypothetical, but which would be concrete if they were real. Metaphysical puzzles abound in this area, but it is common to say that in the case of purely mathematical objects there is no distinction that can be made between actual and possible existence. Model systems in science, however, are often treated quite differently from this.

In making this argument, I take at face value the fact that modelers often *take* themselves to be describing imaginary biological populations, imaginary neural networks, or imaginary economies. An imaginary population is something that, if it was real, would be a flesh-and-blood population, not a mathematical object. Although these imagined entities are puzzling, I suggest that at least much of the time they might be treated as similar to something that we are all familiar with, the imagined objects of literary fiction. Here I have in mind entities like Sherlock Holmes' London, and Tolkein's Middle Earth. These are imaginary things that we can, somehow, talk about in a fairly constrained and often communal way. On the view I am developing, the model systems of science often work similarly to these familiar fictions. The model systems of science will often be described in mathematical terms (we could do the same to Middle Earth), but they are not just mathematical objects.

At the end of the day, of course, some general account must be given of the imaginary objects of both ordinary fiction and scientific modeling. I would favor a naturalistic approach, and this account will probably be quite deflationary in some respects. But 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. For general philosophical reasons, we may eventually want to re-interpret this talk. But one kind of understanding of model-based science requires that we take this ''folk ontology'' seriously, as part of the scientific stategy.

Once we embark on this project, however, a number of complexities arise. If model systems are conceived concretely, what is important is usually not a single imagined system but a *collection* of them. Often these will be collections of systems built from common ingredients (all the imaginary sexual populations that differ genetically at a single locus, for example). Further, though I use the term ''concrete,'' a model system is usually quite schematic, only partly specified. When asked to imagine an evolving population, we will usually be told what the mating system is, but not the number of toes that the organisms have.

From the point of view of a foundational treatment, this might look strange. Again, though, the general phenomenon is familiar from the case of ordinary fictions. When most people read *Lord of the Rings*, before the films were made, they imagined the ''Orcs'' of the book in a very schematic way. Few would have bothered to work out how many toes an Orc has. Peter Jackson, the director of the films, had to imagine his orcs with much more specificity than most of us did; he had to specify the number of toes on an Orc.

A philosopher might immediately respond to this situation as follows: we can handle both the ordinary fictions and the scientific cases with sets of possible worlds. In a sense, that is quite right. We can ''handle'' them this way, for some theoretical purposes. But saying this also misses the point in some respects. It is a striking feature of our psychological capacities that we can engage in this process of schematic imagining. The skill is put to one kind of use in recreational and literary fiction, and to another kind of use in science. This is a capacity we should certainly try to analyze, but not try to analyze away.

I have suggested that we take seriously the apparent treatment of model systems as imagined concrete things by scientists. But it would be a mistake to apply this view too strongly and uniformly. In areas where mathematical methods can be used, one part of the practice of modelers is to discuss mathematical objects and structures in their own right. In these areas, what we often see might be described as a kind of oscillation between thinking of a model system in very concrete terms, and moving to a description of purely mathematical structure. It would be a mistake to insist that one of these is "the model" and the other is not. Each kind of talk can constrain the other. A person might say: "That's the wrong equation to use if you have in mind a sexual population." But a person might also say: "If those are the equations you want to study, you should think of them as describing spatial variation in environment, not temporal variation." So the "concrete" treatment of model systems is part of the story, but it is not the only part.

I have placed considerable emphasis on this issue here because it seems that much of the existing literature has shied away from thinking about model systems in the more concrete way, even when it seems to be a clear part of scientific practice.[11] Giere and Weisberg, who are closer than any others to the view I am presenting, both seem reluctant on this issue. One reason was given to me by Giere (in discussion). Giere said that to see imagined concrete entities as important in philosophy of science is to move too close to the extravagent metaphysics of Meinong and Lewis. Opting for abstract mathematical structures seems less metaphysically dubious.

In response, I accept that the foundational questions are real. But we also have to make sense of the fact that talk of imagined concrete things seems to be important in the actual practice of much model-building. This talk seems intregral to discussion, communication, and the discovery of novel phenomena. The models in Maynard Smith and Szathmáry's *Major Transitions*, discussed in the previous section, certainly have this character. Some of them might be seen as mathematical objects, but are instead described very concretely (the "hypercycle" model used in origin of life problems is an example of this). Others (like their models of the origin of membranes and compartmentalization) do not seem to have an abstract mathematical character at all. Once we move to modeling in fields like cognitive science (for example, models of memory and concept formation) the problem gets worse, as this is a *bona fide* form of modeling that often has even less of a link to mathematical structures.

An embrace of a more "concrete" view of model systems, and the comparison with ordinary fictions, would also help philosophers like Giere reply to some of their critics. Recall that Giere insisted on using an informal notion of resemblance, rather than isomorphism or something similarly strict, in his treatment of model/target relations. He also steers away from a concrete treatment of model systems. The result was a view that was easily criticized, as so little was said (or could be said) about the crucial resemblance relations. But one of the striking things about

[11] This does not hold of much earlier discussions of models in science, like Hesse's work (1966). But in that work, models were seen as adjuncts (perhaps very important ones) to the core structure of a scientific theory.

magined objects in ordinary fiction is that we have an *effortless* informal facility with the assessment of resemblance relations involving these systems. We often assess similarities between two imagined systems (Middle Earth and Narnia), and between imagined and real-world systems (Middle Earth and Medieval Europe). Whatever the foundational status of these objects is, we are all able to trade in these complex assessements of resemblance with little effort and with considerable consensus much of the time. The apparently strange idea of a vague and context-sensitive resemblance relation between model system and target is very unintimidating from this point of view.

The next section will illustrate the themes of this section with some of Richard Levins' scientific work. Following Wimsatt, I will focus especially on Levins' discussions of ''robustness'' as a feature of model-based science.

The example of Levins' evolutionary models

In this section, I will look at some of the models of evolution in variable environments that Levins developed in the 1960s (Levins 1962, 1963, 1968). These models are also used by Levins as a key example in his (1966) paper on model-building. They are used there to illustrate the features of ''Type III'' models, those that sacrifice precision for realism and generality.

Levins does not say explicitly in his paper what he thinks models are. And his focus is generally on mathematical structure. But to me, it is most natural to think of Levins' work as describing big sets of imagined concrete systems—imagined populations, just as they appear to be—with very subtle resemblance relations to various real-world systems. In fact, it is hard to treat some features of Levins' discussion in any other way. In particular, consider the treatment of ''robustness'' in the 1966 paper.

For Levins, a ''robust'' result is one that appears across a range of models of some domain that make different idealizations and have different structures (see also Wimsatt 1987). To give an example of this robustness, Levins describes three evolutionary models that all generate the result that ''in an uncertain environment species will evolve broad niches and tend towards polymorphism.'' As first expressed, this is intuitive but quite vague. The core of Levins' work here is the treatment of the roles of different *kinds* of environmental uncertainty, and the resulting distinctions made between cases where a single generalist is favored, as opposed to a single specialist, or a mixture of specialists.

The three models discussed in the 1966 paper are as follows. The first supposes two discrete environmental states. Each environment is associated with a function specifying fitness as a function of phenotype, with different optimal phenotypes for each environment. There is no genetics in the model. An optimality assumption is used to predict what evolution will do: the population will allegedly maximize its ''rate of increase.'' Mixed populations can have special properties from this point of view.

The second model does not assume discrete environmental variation, as the first does. There is also no specification of the shapes of functions giving the fitness of a given phenotype as a function of environment. Instead, says Levins, we fix the *areas* under the graphs for all the functions describing the fitness of a given phenotype in various environmental conditions, and argue from there using optimization assumptions. (I have never understood this second model.)

The third model uses explicit genetic assumptions, including diploidy. We assume two environmental states. The main aim of the model is to distinguish two kinds of polymorphic populations. One kind comes from ordinary heterosis in a ''fine-grained'' regime of environment variation. In the ''coarse-grained'' regime we predict a special kind of polymorphism deriving from the role of the two homozygotes in giving niche breadth.

My focus here is not on whether these are good models that achieve their aims.[12] My focus is on the *kind* of analysis that Levins is engaged in here. I suggest that in order to do the comparison of results from different mathematical structures that Levins engages in here, it is most natural to think of model systems themselves as something like as *sets of imagined-concrete systems*. In an analysis like this, we are not comparing three different mathematical objects, like three sets of equations. We are imagining three kinds of schematic populations, in slightly different circumstances, and tracking a complicated network of similarity relations across the three. The first and third cases are similar in one way (discrete environments) but differ in the role of genetics. The second and third are most different. The first and second are different in their treatment of environment but similar in the abstraction away from genetics.

If we look at this work through the ''lens'' of the treatment of model systems given above, we can describe it and see how it is supposed to operate as an argument for Levins' robust theorem. Perhaps there is another way to look at it; maybe we can describe the sense in which the three models relevantly converge, while treating the models strictly as mathematical abstractions.[13] But this seems considerably less obvious and natural than the way sketched here, where we think in terms of mathematical *descriptions* of imagined *populations*. So the argument offered here is that the ''imagined concrete'' view of model systems is not just a picturesque and intuitive way of thinking about models, but something that is linked to their distinctive epistemic role within science.

Further directions and conclusion

If the account offered above is on the right track, then a number of topics suggest themselves for further work. In particular, there is the possibility that the distinctive representational features of model-based science might tend to generate particular sociological and historical patterns.

Within model-based science, in the course of everyday work, much talk is *about the model*. Many of the hypotheses under discussion will be hypotheses about how model systems of various kinds behave. For example, evolutionary biology has been very concerned with questions about which model systems can include the evolution of various kinds of altruism, cooperation, and reproductive restraint. Cognitive science has been very concerned with questions about which kinds of idealized neural nets can learn the solutions to various kinds of problems (Marcus 1997).

[12] The first, central model is criticized in Godfrey-Smith (1996, Chapter 9).

[13] Weisberg suggests that this could be done by embedding the three models within a single richer state space.

When much day-to-day discussion is about model systems, disagreement about the nature of a target system is less able to impede communication. The model acts as a ‘‘buffer,’’ enabling communication and cooperative work across scientists who have different commitments about the target system. So some situations that might potentially generate linguistic ‘‘incommensurability,’’ in Kuhn’s (1970) sense, are prevented from doing so, because the model system provides common ground. It would be interesting to compare models with other devices and practices that overcome potential communication problems in science (Galison 1997).

The flip-side of this solution to incommensurability problems, however, might be a special kind of *inertia* visible in some scientific work. Suppose a field makes *full use* of its ability to hold onto a model while the construal shifts, and surrounding knowledge changes. There is the possibility that features of the background context that made the model initially seem reasonable and appropriate may be lost or jetisoned, while the model is nonetheless retained, with a more subtle or low-key construal of its relation to the target system. The model is kept on, but largely through a kind of inertia; that model would never have been the result of work that started afresh from the new set of background beliefs. So the ‘‘buffering’’ seen in model-based science, stemming from the role of flexibility in construals, may have both good and bad features. It enables continuity, makes communication easier, and makes retooling rarer. But sometimes it will be an option that scientists would be better off not taking.

I will briefly summarize my main points. An understanding of models within recent philosophy of science has been distorted by the formalist priorities of the ‘‘semantic view.’’ This has resulted in attempts to force all scientific theorizing into a procrustean bed with a shape determined by the notions of ‘‘model’’ that have proven useful in logic and set theory, and sometimes in physics. Model-theoretic ideas can, of course, be used as the basis for a general theory of meaning, but such an approach has well-known problems in its pure forms, and in any case has no particular connection to representations in science as opposed to elsewhere. Once this project is abandoned, it becomes possible to recognize and analyze model-based science as a distinctive style of theoretical work, which yields particular kinds of representations, explanations, and patterns of change. A contrast between recent classics in evolutionary theory by Buss and Maynard Smith and Szathmáry shows how a single scientific problem can be approached both with and without models functioning as the currency of theorizing and explanation. Giere’s (1988) account is a good starting point for the description of how model-based science works. Scientific modelers often treat model systems in a ‘‘concrete’’ way that suggests a strong analogy with ordinary fictions. The style of robustness analysis championed by Levins shows that this way of thinking of model systems is not merely picturesque but may have a definite role within the development of scientific ideas.

Acknowledgements I am indebted to all those present at the Penn conference, but especially to Michael Weisberg and Deena Skolnick, for comments on these ideas.

References

Buss L (1987) The evolution of individuality. Princeton University Press, Princeton

Cartwright N (1999) The dappled world: a study of the boundaries of science. Cambridge University Press, Cambridge

Downes S (1992) The importance of models in theorizing: a deflationary semantic view. In: Hull D, Forbes M, Okruhlik K (eds) PSA 1992, vol. 1. Philosophy of Science Association, East lansing, pp 142–153
French S, Ladyman J (1999) Reinflating the semantic approach. Intl Stud Philos Sci 13:103–121
Galison P (1997) Image and logic. A material culture of microphysics. Harvard University Press, Cambridge
Gentner D (2002) Mental models, psychology of. In: Bates P, Smelser N (eds) International encyclopedia of the cognitive and behavioral sciences. Elsevier, Amsterdam, pp 9683–9687
Giere R (1988) Explaining science: a cognitive approach. Chicago University Press, Chicago
Giere (1999) Using models to represent reality. In: Magnani L, Nersessian NJ, Thagard P (eds) Model-based reasoning in scientific discovery. Kluwer/Plenum, New York, pp 41–57
Godfrey-Smith P (1996) complexity and the function of mind in nature. Cambridge University Press, Cambridge
Godfrey-Smith P, Lewontin RC (1993) The dimensions of selection. Philos Sci 60:373–395
Griesemer J (1990) Material models in biology. In: PSA: Proceedings of the Biennial meeting of the Philosophy of Science Association, vol. 2, pp 79–93
Hesse M (1966) Models and analogies in science. University of Notre Dame Press, Notre Dame
Johnson-Laird P (1983) Mental models: towards a cognitive science of language, inference, and consciousness. Harvard University Press, Cambridge, MA
Kuhn TS (1970) The structure of scientific revolutions, 2nd ed. Chicago University Press, Chicago
Levins R (1962) Theory of fitness in a heterogeneous environment. I. The fitness set and adaptive function. Amer Nat 96:361–373
Levins R (1963) Theory of fitness in a hetergeneous environment. II. Developmental flexibility and niche selection. Amer Nat 97:75–90
Levins R (1966) The strategy of model-building in population biology. Amer Sci 54:421–31
Levins R (1968) Evolution in changing environments. Princeton University Press, Princeton
Lewis DL (1984) Putnam's paradox. Austr J Philos 62:221–236
Lewontin RC (1963) Models, mathematics, and metaphors. Synthese 15:222–244
Lloyd EA (1988) The structure and confirmation of evolutionary theory. Greenwood Press, Boulder
Marcus G (1997) The algebraic mind. MIT Press, Cambridge, MA
Maynard Smith J, Szathmáry E (1995) The major transitions in evolution. Oxford University Press, Oxford
Michod R (1999) Darwinian dynamics: evolutionary transitions in fitness and individuality. Princeton University Press, Princeton
Morgan M, Morrison M (1999) Models as mediators: perspectives on natural and social science. Cambridge University Press, Cambridge
Nersessian N (1999) Model-based reasoning in conceptual change. In: Magani L, Nersessian N, Thagard P (eds) Model-based reasoning in scientific discovery. Kluwer/Plenum, New York, pp 5–22
Putnam H (1982) Reason, truth and history. Cambridge University Press, Cambridge
Queller D (1997) Cooperators since life began. Quart Rev Biol 72:184–88
Suppe F (ed.) (1977) The structure of scientific theories. 2nd ed. University of Illinois Press
Suppes P (1960) A comparison of the meaning and uses of models in mathematics and the empirical sciences. Synthese 12:287–301
Thompson-Jones M (2006) Models and the semantic view. Philos Sci (Forthcoming) (Proceedings of PSA 2004)
Van Fraassen B (1980) The scientific image. Oxford University Press, Oxford
Weisberg, M (forthcoming) Who is a modeler? Brit J Philos Sci, to appear
Weisberg M (2003) When less is more. PhD dissertation, Philosophy Department, Stanford University
Wimsatt WC (1987) False models as a means to truer theories. In Nitecki M, Hoffmann A (eds) Neutral models in biology. Oxford University Press, Oxford, pp 23–55