1 Models in Theoretical Biology

Manfred D. Laubichler and Gerd B. Müller

All branches of the biological sciences rely on models. This simple observation comes as no surprise to anyone who has either participated in or thought about biological research. Experimental studies investigate model systems or model organisms that are carefully selected for a particular research problem; these experimental data are assembled and incorporated into computational, diagrammatic, or conceptual models that are used to organize and analyze these results; and even abstract formal theories—those elusive targets of scientific research—come to life in the form of formal and predictive models. There are thus many different types of models within the biological sciences that fulfill a variety of important practical, conceptual, and epistemological functions. The literature on this topic is extensive; see, for instance, chapters 2 and 3 in this volume, and Morgan and Morrison (1999), de Chadarevian and Hopwood (2004); Hacking (1983), Giere (1988), Cartwright (1983), Wimsatt (1987), Krohs and Callebaut (2007), and Wolkenhauer and Ullah (2007).

Among the reasons why models are so important within the biological sciences are the challenges posed by the rapidly growing amount of experimental results and the increasing complexity of the phenomena under investigation. Conceptual abstractions and mathematical modeling have always been part of the interpretation and integration of these accumulated data. Today, due to the improvement of computational methods, models of biological phenomena have reached, in many ways, a new dimension. It is now possible to model and simulate complex biological structures and processes in hitherto unknown detail. At the same time, bottom-up models, such as those adapted from the physical sciences, elucidate the properties of natural biological systems. Therefore, as can be seen in all chapters of this volume, modeling and simulation have been the key to most major advances in the biological sciences in recent decades.

The complex dynamics of biological processes at all levels of organization represent one of the key issues of modeling biology. These dynamical processes can be best understood in terms of a set of relations between individual entities and their environment, which often includes other individual entities. For example, while a specific cell inside an organism interacts with its neighbor cells as well as with the extracellular medium, and specific organs interact with the rest of the body, the organism itself interacts with the surrounding environment. All these processes, taking place along several spatial and temporal scales, are essential for defining the reproduction and survival not only of specific individual organisms, but also of their species as a whole. And because the environment itself is affected by such interactions, important feedback loops that ultimately define the biological world are established. In addition, a better understanding of these interactions is essential for most human concerns, ranging from the predictability and treatment of diseases to the proper management of the environment.

The study of form, both organismal and molecular, is another important area in which modeling approaches have become central. Mainstream research during the last decades has focused largely on molecular, biochemical, and environmental issues, which has ultimately contributed to the consolidation of genetics, molecular biology, evolutionary biology, ecology, and animal behavior as well-established fields. At the same time, only limited attention has been paid to the important problem of biological form (i.e., the characterization and analysis of the geometrical and topological properties of biological entities). Together with the respective biochemical features, the morphological properties of any biological entity represent an important element in the dynamic interactions between individuals and their environment. The shape of a specific cell, for instance, constrains and is constrained by the morphology of the neighboring cells.

At a more macroscopic level, the density of branching structures, such as blood vessels or trees, determines the efficiency of material and energy exchanges with the environment. And at the organismal scale, the shape of the avian beak, for example, is critical for access to sources of food. Thus, generally speaking, the morphology of biological entities mediates in a critical fashion a large number of interactions between the individual and the environment. Recent breakthroughs in modeling biological forms (several discussed in this volume) have provided important insights into these crucial relationships between form and function.

Understanding the structural properties of biological forms is one important dimension, but it is not complete without considering the fact that all biological systems develop from less complex precursors. Development is thus another central aspect of modeling biological systems. The essential features of this temporal, and in a sense also historical, dimension of biological systems can be understood only through models, since any detailed reconstruction would face the Borges dilemma in biology—the map of the system would soon be as complex as the system itself, and therefore would be of no use (Laubichler and Pyne, 2006). Models that are based on the right kind of abstractions, both in the form of the relevant parts of systems as well as of their causal interactions, are thus the only way we can arrive at an understanding of development. These models combine the structural features of models of form with those interactive aspects that are characteristic of dynamical models. Like models of form, models of development also combine biological and physical properties of complex organismal systems.

Organisms not only develop, they also interact with each other and with their environment. In studying these forms of behaviors, researchers have always relied on models that enabled them to go beyond mere descriptions and to develop causal hypotheses about the triggers and purposes of observed behaviors. Behavior has been modeled in the context of physiology and even of mechanics—here the focus was on understanding the triggers and stimuli of behaviors, of their evolution—with an emphasis on the adaptive value of behavior, and of economic theory—that brought a perspective of strategies, gain, and "rational decision-making" to explanations of behavior. More recently a number of integrative approaches introduced robots into the study of behavior. This no longer represents merely a new conceptual approach or a new methodological tool; rather, these studies create a physical-technical model system with the goal to embody rather than to describe behavior.

Mathematical models have a long tradition in evolutionary biology. They are also one of the prime examples of successful conceptual abstractions within the biological sciences. These abstractions—from the concept of a population to that of a replicator—also have enabled a more generalized mathematical theory of evolving systems captured, for instance, by the replicator or the Price equations. The fundamental problem of modeling phenotypic evolution is how to combine these abstract dynamical models of population dynamics with concrete properties of organisms as developing systems that interact with each other and their environment. This will require a new class of models that integrate biological processes at different scales. In order to accomplish this goal, a new set of models covering the whole spectrum from experimental model systems and new kinds of model organisms to conceptual and mathematical models are currently being developed. The details of these resulting models differ, depending on whether they focus on morphology (part III), development (part IV), behavior (part V), or evolution (part VI), but they all span different scales and levels of complexity.

Recently philosophers and historians of science concerned with actual scientific practice and knowledge production have focused their attention on models and modeling strategies (e.g., Hacking, 1983; Giere, 1988; Morgan and Morrison, 1999). This literature has already yielded a more complex understanding of how knowledge has been generated in particular experimental contexts, and it has contributed to a more detailed and less schematic philosophical conception of the structure of scientific theories and explanations. This new focus on conceptual issues in the biological sciences, triggered to no small degree by the emphasis on models, is indeed a profound addition to theoretical biology at large. The chapters in part II of this volume discuss models and modeling strategies from the perspective of philosophy, history, and applied mathematics. These chapters also provide a critical evaluation of the theoretical literature on modeling.

Models in Theoretical Biology

Today theoretical approaches are gaining in prominence throughout the biological sciences, contributing to what can, without exaggeration, be called a renaissance of theoretical biology. Models and modeling strategies are an important part of this trend. Indeed, in many instances, developments in theory are a direct outgrowth of modeling efforts. However, the relationship between models and theories, and between modeling and theoretical biology, are complex and manifold. Following Laubichler et al. (2005), we distinguish at least four different, yet interrelated, areas of theoretical biology in the postgenomic age. Different types of modeling and modeling strategies define the distinctions between these areas.

The first dimension of theoretical biology reflects the need to analyze, organize, and manage large amounts of data. This is the domain of bioinformatics and of certain areas of computational biology, which deal with the substantial challenges of data analysis and representation. The problem here is not only that biology is currently experiencing a phase of exponential growth in the amount of data available; there are also different kinds of data that need to be connected (integrating sequence information and medical histories, for instance). Consequently, data mining is nowadays a desired skill for theoretical and experimental biologists alike. But many of the bioinformatics tools, which the average user simply wants to apply to analyze data, actually represent heuristic and theoretical models of the underlying complexities of biological systems.

Even though one can look at the computational and mathematical tools developed in bioinformatics as the microscope of the twenty-first century, allowing us to "see" new connections and structures (Laubichler et al., 2005), these nevertheless embody more conceptual and theory-driven dimensions of biological research and are in fact models that represent certain aspects of biological systems. A DNA chip and the computational tools that enable us to analyze molecular data are, in a way, a model of the transcriptional state of a cell or a tissue. These and other aspects of merging new data-rich fields with traditional modeling strategies are particularly relevant in the emerging field of systems biology (Krohs and Callebaut, 2007).

A second area of theoretical biology—for many, the primary aspect of theoretical biology—involves mathematical and computational model-building and analysis (see also Levins, 1984). Over the last decades, simulation and modeling approaches have increased dramatically. Today these routinely include standard systems of differential equations, agent-based models, and spatial models. All this has been made possible by the substantial increase in computational capacity. While model-building and simulation have become an integral part of most areas of biology, they also remain a core part of theoretical biology proper.

In our view, the main emphasis of *theoretical* biology is not so much on creating an accurate mathematical representation of a particular event, but rather on gaining insights of a more general nature by means of abstract representations of biological processes. These representations are, of course, a form of model of the phenomenon in question. Only from the perspective of such a model is it possible to inquire to what extent different phenomena can be understood as instances of similar underlying processes or dynamics. This is a primary goal of theoretical biology.

A focus on concept formation and conceptual analysis constitutes the third area of theoretical biology. While this is an important and sometimes overlooked part of biology, it is also related to many ongoing efforts within the philosophy of biology. Conceptual analysis is often closely linked to mathematical analysis, including several mathematical modeling strategies—even more so as many theoretical ideas in biology become expressed in formal terms. However, conceptual analysis is by no means restricted to mathematical models. As several examples in this volume demonstrate, it is precisely through the application of a *concept*, such as constraint or optimization, to a wide range of phenomena that new insights will be gained.

Darwin's own work, for instance, contains several examples of this kind. Probably the best-known one is the conceptual model provided by artificial selection, which he knew from firsthand experience and then employed it to infer processes in natural populations. We submit—and the chapters in this volume support this claim—that as more data and more powerful mathematical and computational tools become available, the importance of this sort of conceptual biology to aid in the discovery of fundamental similarities between different systems and processes is likely to increase.

The fourth area of theoretical biology involves theory integration. This function arises from the fact that formerly distinctly separate disciplines and research domains have begun to merge, both methodologically and conceptually. Theory integration requires different types of models, most prominently heuristic and conceptual models. The goal is to arrive at formal and theoretical models that capture the essential features of a wide range of biological phenomena. Theory integration is often connected to the idea of a "synthesis." One prominent example of this kind of integration is evolutionary developmental biology (EvoDevo), which combines several fields of investigation in a practical and theoretical sense, leading to new research strategies and agendas (Müller, 2005, 2007). As the discussion in chapter 16 of this volume shows, achieving such a synthesis depends in crucial ways on the availability of the right kinds of models, both conceptual and material. It is all the more important today because the life sciences are in a phase of rapid expansion and ever increasing specialization, while at the same time biological theories and concepts have become increasingly more widespread outside the traditional boundaries of the life sciences (from evolutionary psychology to artificial life).

In summary, in our conception, theoretical biology is both foundational and practical, mathematical and conceptual, and always in a close partnership with empirical research. Furthermore, all areas of theoretical biology involve different kinds of models and, indeed, represent different types of modeling strategies within biology. Theoretical biologists are engaged in all areas of biological research: managing and analyzing data, developing new strategies for representing and visualizing data, building mathematical models and developing simulations that are designed to capture essential and general aspects of biological processes, as well as fine-tuning these models to allow for precise predictions; formulating concepts that adequately represent the underlying biological phenomena and their associated mathematical representations; and, finally, contributing to the theoretical integration of the life sciences. This complexity and diversity of theoretical biology is matched by the diversity and complexity of the attempts to model biology.

Classes of Biological Modeling

Models have many different functions within biology and even, as we have seen, within theoretical biology. These functions are reflected in a variety of attempts to classify and analyze different types of models and modeling strategies (e.g., Hacking, 1983; Levins, 1984; Wimsatt, 1987; Giere, 1988; Morgan and Morrison, 1999). Each of these schemes is guided by a particular question or philosophical approach. Here we provide a different sort of classification, one that is derived from an analysis of the various and varied uses of models within theoretical biology as represented in this volume (table 1.1). Our classification is thus a preliminary summary of modeling strategies in biology derived from what practitioners in different fields emphasize as being central to their efforts. But we would argue that while it certainly is not complete or the only way of classifying models, table 1.1 is a good representation of bio-

Table 1.1
Classes of biological modeling

Material	Physical	objects, robots, model organisms	Chapters 3, 11, 13
	Computational	2-, 3-, 4-D virtual models and simulations	Chapters 6, 7, 9, 10, 14, 15
Theoretical	Diagrammatic	functional, organizational	Chapters 4, 5, 7, 8, 13, 15
	Mathematical	formal, analytical, predictive	Chapters 4, 5, 6, 8, 9, 10, 12, 14
Heuristic	Conceptual	interpretive, heuristic	Chapters 3, 4, 5, 11, 12, 14, 16

logical modeling in general. It is also ambiguous in the sense that our descriptions are more centers of emphasis than categorical distinctions, which is a more adequate reflection of the varied uses of models within biology.

We distinguish primarily between different reference frames and functional roles of models in biology. Models can be *material* (i.e., concrete physical models, such as model organisms and robots, or virtual reconstructions and simulations of concrete situations (such as the virtual 3- and 4-D reconstruction of a developing embryo or the simulation of the population dynamics of a given species). Models can be *theoretical*, which implies a higher degree of abstraction and/or generality, and includes diagrammatic models of the functional and organizational properties of organisms and organ systems (such as the nervous system or the vasculature, cells, and intracellular structures), as well as abstract representations of relational properties of biological systems (such as network models and graphical representations), and analytical and predictive models of dynamical systems (such as models of natural selection). And models can be *heuristic*, which tends to be a conceptual representation of a particular problem. These conceptual models are often the first descriptions of a research problem and stand at the beginning of investigative pathways (then they serve mainly as heuristic tools), but they also complete a research program when problems are really understood and can be summarized within a conceptual interpretative framework (then often including other types of models, such as predictive formal models, simulations, or representative models).

The function of models within biology is thus dynamic rather than static. At all stages of biological research, models play an important role, whether explicitly or implicitly. Often this means that we see a progression from heuristic to material and theoretical models of a particular problem. But just as often it means that one particular model is transformed in light of ongoing research, and different dimensions (heuristic, theoretical, etc.) gain in prominence. The success of research programs thus depends on the right kind of modeling strategies. In that sense almost all research in biology amounts to "modeling biology."

References

Cartwright N (1983) How the Laws of Physics Lie. New York: Oxford University Press.

De Chadarevian S, Hopwood N (eds) (2004) Models: The Third Dimension of Science. Stanford, Calif.: Stanford University Press.

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

Hacking I (1983) Representing and Intervening: Introductory Topics in the Philosophy of Science. Cambridge: Cambridge University Press.

Krohs U, Callebaut W (2007) Data without models merging with models without data. In: Systems Biology: Philosophical Foundations (Boogerd FC, Bruggeman FJ, Hofmeyer J-HS, Westerhoff HV, eds), 181–213. Amsterdam, Reed-Elsevier.

Laubichler MD, Hammerstein P, Hagen E (2005) The concept of strategies and John Maynard Smith's influence on theoretical biology. Biol Philos 20: 1041–1050.

Laubichler MD, Pyne L (2006) The Borges challenge in biology: Review of Kenneth M. Weiss and Anne V. Buchanan, *Genetics and the Logic of Evolution*. BioEssays 28: 768–769.

Levins R (1984) The strategy of model building in population biology. In: Conceptual Issues in Evolutionary Biology (Sober E, ed), 18–27. Cambridge, Mass.: MIT Press.

Morgan MS, Morrison M (eds) (1999) Models as Mediators: Perspectives on Natural and Social Science. Cambridge: Cambridge University Press.

Müller GB (2007) Six memos for EvoDevo. In: From Embryology to EvoDevo: A History of Embryology in the 20th Century (Laubichler MD, Maienschein J, eds), 499–524. Cambridge, Mass.: MIT Press.

Müller GB (2005) Evolutionary developmental biology. In: Handbook of Evolution, vol. 2 (Wuketits FM, Ayala FJ, eds), 87–115. San Diego: Wiley.

Wimsatt WC (1987) False models as means to truer theories. In: Neural Models in Biology (Nitecki MH, Hoffman A, eds), 23–55. New York: Oxford University Press.

Wolkenhauer O, Ullah M (2007) All models are wrong. In: Systems Biology: Philosophical Foundations (Boogerd FC, Bruggeman FJ, Hofmeyer J-HS, Westerhoff HV, eds). Amsterdam: Reed-Elsevier.