

Models in Biology

Jay Odenbaugh, Lewis and Clark College

Summary

In recent years, there has been much attention given by philosophers to the ubiquitous role of models and modeling in the biological sciences. Philosophical debates have focused on several areas of discussion. First, what are models in the biological sciences? The term 'model' is applied to mathematical structures, graphical displays, computer simulations, and even concrete organisms. Is there an account which unifies these disparate structures? Second, scientists routinely distinguish between theories and models; however, this distinction is more difficult to draw in the biological sciences since biologists often only have a variety of models and rarely have something like a fundamental theory. What then is a theory in biology? Third, how are models related to empirical or "target" systems?

1. Introducing Models

Biologists use the term 'model' in several different ways; however, most often it means a *idealized representation of empirical systems*. For example, an evolutionary biologist might model natural selection independent of other evolutionary "forces" like random genetic drift. To do so, they assume the population is infinitely large so that actual and expected reproductive success is the same. Clearly no population of organisms is infinite in size; hence, this is an idealization. A second use of the term 'model' concerns organisms or physical objects which are particularly useful for study due to their simplicity – a model organism like *Drosophila melanogaster* (fruitfly), *Tribolium castaneum* (flour beetle), *E. coli*, or physical DNA double-helical structure. Whether these two uses can be given a unified treatment is something discussed below.

As an example of a mathematical model, consider a very simple metapopulation model. A metapopulation can be thought of as a "population of populations" which are subdivided spatially but connected by migrating organisms. If P is the proportion of patches occupied by a species, c is the rate of colonization, and e is the rate of extinction, then the rate of change of the proportion of occupied patches is the proportion of patches colonized minus that of the patches where extinction occurs; or $dP/dt = cP(1-P) - eP$. If $dP/dt = 0$, then the equilibrium is $P^* = e/c = 1 - e/c$. This equation entails both that the rate of colonization must be greater than the rate of extinction otherwise the equilibrium proportion of patches will decay to zero and that if the rate of extinction is non-zero then there must be habitats unoccupied. There are several idealizations on which this model is based. For example, we are assuming the local populations are identical in their chances of being colonized and going extinct,

rates of colonization and extinction are constant, and distance between patches is irrelevant.

In this entry we will consider several philosophical questions concerning both senses of 'model'. First, what exactly is a model? Second, scientists routinely distinguish between theories and models; however, this distinction is more difficult to draw in the biological sciences since biologists often only have a variety of models. What then is a theory in biology? Third, how are models related to empirical or "target" systems?

2. What are Models?

Biologists describe models as having three components – *parameters*, *variables*, and *laws* (Otto and Day 2007). Parameters are the properties of the biological objects which are not changing; variables are the properties of these objects which are changing, and laws are the relationships between the two. In the above metapopulation model, c and e are parameters of the metapopulation, P is the variable, and the first-order differential equation is the "law of succession". The laws in biological models may be deterministic or stochastic and they may describe relationships which are temporally continuous or discrete.

Biologists sometimes describe models simply *as* equations. However, this is problematic since the equations are syntactical and even a simple change in the language of use produces a distinct model. It is more reasonable to define a model as whatever *satisfies* the relevant equations or descriptions; i.e., whatever makes the equation or description true. Biologists' models are highly idealized. Thus, biological systems will rarely satisfy the equations or descriptions and hence are not themselves models. Mathematical models in biology are *abstract objects* (see ABSTRACT OBJECTS) or mathematical structures. A customary suggestion is that models are *phase* or *state spaces*. Suppose the system consists in n variables which are the possible states of the system. Geometrically the system is an n -dimensional space where the state of the system is the value of each of the variables at a specific time. The laws are rules of succession which describe how the system's state changes with time given the parameters of the system. In our metapopulation model, we have a one-dimensional state space consisting in the possible values of P . However, we might add complexity by supposing we have two species which exist as metapopulations where the first species is a superior competitor and the second is a superior colonist. Thus we would have a two-dimensional state space consisting in the possible values of P_1 and P_2 of the two species.

It is important to note that the notion of a "law" here is distinct from that discussed by philosophers of science and metaphysicians. First, the laws in biologists' models need not be exceptionless regularities, necessary relations between universals, or some other metaphysical notion (see LAW OF NATURE). Second, there is an ongoing controversy amongst philosophers of biology as to whether biology has laws in this sense. Suppose a law of nature is an exceptionless generalization which supports counterfactuals. Several

philosophers have argued biology has no such laws since biological generalizations are either mathematical truisms (Sober 1992), true counterfactual supporting generalizations with exceptions (Brandon 1992), or are the product of evolutionary history and whose truth values change over time (Beatty 1992). However, if biological laws are mathematical truisms or “excepted” generalizations, then one can argue that the subject has essentially been changed since we are no longer discussing traditional conceptions of lawhood. Likewise, the suggestion that biological laws “change” in their truth value over time due to evolutionary history apparently confuses the truth value of a conditional of the form “if F , then G ” and the satisfaction of its antecedent. Others argue that many of the problems which call into question the existence of biological laws do not do so with respect to physics and chemistry – hence, they should not in biology (Colvyan and Ginsberg 2006). However, there are philosophers of physics who are equally skeptical of physical laws (Cartwright 1983, van Fraassen 1989) which may undercut this argument.

We saw above that models are sometimes concrete as opposed to abstract objects. For example, one might use a physical specimen like *Saccharomyces cerevisiae* (budding yeast) or a physical double helical structure to represent a more complex empirical system like the cell cycle in humans and the DNA molecule. Here we still have an idealized representation of an empirical system since the concrete system is simplified or idealized relative to its represented and more complex counterpart. The simplifications usually involve properties like generation time, size, or manipulability. Thus, the notion of a model as an idealized representation applies *both* to mathematical and concrete objects; however, the structures which serve as representations differ.

3. Models and Theory

If biology lacks laws traditionally construed and a scientific theory consists in a set of such laws, then apparently biology also lacks scientific theories. One popular alternative account of scientific theories is what is sometimes called the *model-based view* (Giere 2002). On this view, scientific theories consist in *theoretical definitions* and *theoretical hypotheses*. A theoretical definition characterizes the model itself; in effect, it is the relevant set of equations or descriptions. The theoretical hypothesis is a claim of the form: “Model M is similar to system S in certain respects to certain degrees”. So, the equation above describes a mathematical structure which may be claimed to be similar to the spatial dynamics or persistence times of metapopulations of checkerspot butterflies in Santa Clara, California or the Glanville fritillary butterflies on the Åland Islands in the Baltic Sea. Likewise, budding yeast may be claimed to be similar to the cell cycle in humans. Many philosophers of biology have embraced this account of theories. Biology contains respectable theories even if does not contain laws and the model-based view provides an account of the structure of these theories and the practice that undergirds them.

As we have seen, our models are often very simple or idealized when compared to biological systems; moreover, biologists devise many different models for the same systems – what some philosophers of physics call a “patchwork” (Cartwright 1995). Some models are designed to be more realistic and predictively accurate of very specific systems – they are “high fidelity” models. For example, they may predict point values of the variables or parameters. Others are much more general leaving precision behind and represent “qualitative” features of the systems of interest such as whether they possess stable equilibria, limit cycles, or chaos. By combining these various models and noting their similarities to the world, we have the contours of biological theory. In metapopulation theory, we have the qualitative “classic model” discussed above, ones that do not assume that the rate of colonization is constant (i.e., a “propagule rain”) or that extinction is constant but is affected by the number of occupied sites (i.e., “rescue effect”), and ones which are multispecies systems combining colonization and competition.

Whether biological theories *necessarily* consist in clusters or patchworks of models is the subject of current controversy. Biological systems consist in a large number of variables which weakly interact. Some have argued that this fact coupled with our own cognitive abilities and resources requires that we must use multiple models. First, the parameters and variables cannot be measured in real time with so few field or laboratory biologists. Second, the “true equations” would be incredibly difficult to understand and utilize given that they would consist in so many partial differential equations with hundreds of parameters and variables. Third, the resulting equations will be analytically insoluble. If biologists aim to have models which represent biological systems realistically, generally, and precisely, then there will be tradeoffs with respect to these model desiderata (Levins 1966). It can be argued that these difficulties can be mitigated. For example, whether a model is too complex crucially depends on how those equations are represented – often parameters are tokens of the same type. Likewise, even insoluble equations can be simulated numerically on computers providing an inductive survey of their dynamical behavior. Thus, this debate is still ongoing.

It is also significant to note that the model-based view of theories applies to model organisms as well to mathematical models. In effect, theoretical definitions characterize the model organism and relevant similarities between the specimen and target system are noted. Thus, analogical reasoning broadly construed is used with regard to both mathematical structures and physical objects.

4. Models and Target Systems

Biologists empirically evaluate their models in three specific ways: goodness-of-fit, truth of assumptions, and variety of evidence (Lloyd 1988). First, they may determine the goodness-of-fit of the predictions of the models with respect to empirical systems. Specifically they compare expected values of the variables and parameters values using

various statistical or model selection tools like maximum likelihood. Likewise, they directly or indirectly test the truth of their model assumptions. For example, simple metapopulation models assume a “propagule rain” where there is constant number of colonists. However, this need not be true of all populations. Lastly, they examine their model’s goodness-of-fit and assumptions across a variety of different systems of the same or of a different type. Each of these elements is an important component of the empirical accuracy of biological models. However, it is not the case that all models can be evaluated in this way. For example, the classic metapopulation model is rarely tested against data directly. Rather it serves a “template” for constructing more realistic models. Thus, some models serve a heuristic role in providing conceptual frameworks for developing models. Some biologists argue a model can be successful only if it is empirically tested. However, by recognizing that models provide heuristic functions or templates provides a beginning response to this worry (Odenbaugh 2005, Wimsatt).

We have seen how the equations and descriptions describing our models are not literally true of the target systems of interest. However, many biologists claim that their models *explain* the phenomena of interest. If models incorporate false assumptions, then in what sense do they explain it? One suggestion is that those equations are “approximately true” of the system (see APPROXIMATE TRUTH). Unfortunately, an adequate account of approximate truth is elusive since customary accounts have fatal flaws. Another response to this worry is to show that a model represents the causal structure of a phenomenon and that its idealizations do not affect this causal relationship. If the same effect would be produced across independent assumptions describing this causal relationship, then this effect is robust. This form of explanatory reasoning applies equally well to model organisms. If incorporating complexities with respect to size, age structure, generation time, and so do not affect the causal structure generating some trait of interest, then the effect is robust over different properties. Nevertheless, we are left with a neglected question – how do false models explain?

Bibliography

- Beatty, J. (1995) "The Evolutionary Contingency Thesis," In G. Wolters and J. Lennox (eds.), *Concepts, Theories, and Rationality in the Biological Sciences*, University of Pittsburgh Press. This essay challenges the claim that there are biological laws given the fact the these systems evolve through time.
- Brandon, R.N. (1997) “Does biology have laws? The experimental evidence”, *PSA 1996*, vol. 2, 444–457. Here the author contends that there are biological laws provided we relax the notion of lawhood.
- Cartwright, N. (1999) *The Dappled World: A Study of the Boundaries of Science*. Cambridge University Press. A defense of the view that physics contains a patchwork of models rather than some fundamental theory.
- Giere, R. (1999) *Science without Laws*. Chicago: University of Chicago Press.
A recent defense of a “model-based” view of theories.

- Levins, R. (1966) "The Strategy of Model Building in Population Biology," *American Scientist* 54: 421-431. A important examination of the methodology of model-building by a renowned biologist.
- Lloyd, E. (1988) *The Structure of Evolutionary Theory*. Princeton University Press.
A defense of a model-based account of theories in the context of evolutionary theory and population genetics.
- Odenbaugh, J. (2005) "Idealized, Inaccurate, and Successful: A Pragmatic Approach to Evaluating Models in Theoretical Ecology," *Biology and Philosophy* 20: 231-255.
A examination of the heuristic roles that mathematical models play in the sciences of ecology.
- Otto, S. and T. Day (2007) *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*. Princeton University Press. A useful textbook on modeling in ecology and evolutionary biology.
- Sober, E. (1997) "Two Outbreaks of Lawlessness in Recent Philosophy of Science," *Philosophy of Science* vol. 2, 458-467. The author claims that evolutionary biology has laws; however, they are often mathematical truisms.
- Wimsatt, W. (1987) "False Models as means to Truer Theories, " In M. Nitecki and A. Hoffman, eds., *Neutral Models in Biology*. London: Oxford University Press, pp. 23-55. An important discussion of the heuristic roles of modeling in the biological sciences.