COMMENTARY

A RESPONSE TO ORZACK AND SOBER: FORMAL ANALYSIS AND THE FLUIDITY OF SCIENCE

RICHARD LEVINS

Department of Population Sciences and International Health, Harvard School of Public Health Boston, Massachusetts 02115 USA

MODELS OR MODEL BUILDING?

IN THE BOOKSTORES of my youth it was common to find titles that included the term "foundations of science." These were not books about science at all but about logic, heirs of the Russell-Whitehead program to derive mathematics from logic, and science from mathematics. The hope persisted that objectivity could be achieved by analytical methods: clear definitions, unambiguous categories, sharp measurement, and the discovery of algorithms that could substitute for the caprice of human judgment. But the program as a whole has been a failure, as indeed it had to be. The foundations of science are to be found in history and sociology, not formal analysis. Orzack and Sober's critique of my 1966 paper falls within the tradition of formal analysis, and our disagreements fall along the axis from formal analytical to dialectical views of the scientific process (Levins and Lewontin, 1985).

Formal analysis freezes moments of a process into things. My essay was concerned with model building as a process, embedded in the larger processes of scientific investigation. It looked at the decisions that population biologists were making at the time (the middle 1960s) in order to solve different kinds of problems. It was concerned with two difficulties: one was that attempts to optimize different criteria for satisfactory models interfere with each other, and the other was that all models are partly false. I identified three tendencies in model building in population biology, each of which involved the sacrifice of one of the desiderata: generality, realism and precision. Other criteria were mentioned (manageability and understandability) but not pursued. Nor did I include in the listing strategies that sacrificed two criteria for the third. It is clear from the context that the listing of three strategies was not intended to be exhaustive of all possible population biology practice or of science as a whole. What was important was the notion of trade-offs in model building.

Orzack and Sober reify the discussion of model *building* into a discussion of models, and turn the trends I noted into "Levins's taxonomy of models" or "Levins's trichotomy."

Formal analysis prefers fixed definitions of objects, free of their context, in order to allow for unambiguous measurement and ranking. But in science definitions evolve with the problem. The formal mode of thought prefers sharply separated disjunct categories and has indeed encouraged the long list of false dichotomies that have plagued our biology: heredity versus environment, physical versus biotic control of population abundance, internal versus external determination, random versus determin-

The Quarterly Review of Biology, December 1993, Vol. 68, No. 4

0033-5770/93/6804-0003$1.00

istic processes, and equilibrium versus non-equilibrium systems.

Formal analysis breaks the world, and the scientific processes that study the world, into mutually exclusive categories, although the world is fluid and categories form, dissolve, overlap, and interpenetrate. Orzack and Sober do not define quantity and quality, but they do treat them as mutually exclusive alternatives and decide that there are no "purely" qualitative mathematical models. They conclude from this that "Type II and Type III would distinguish different kinds of models if Type III models entailed no quantitative predictions at all" (p. 538). Since the model types were originally offered to indicate different directions of model development, they were not intended to be disjunct categories. Neither are quantity and quality. I could not have suggested that Type III models are "pure" qualitative models.

Even the most purely quantitative of objects, numbers themselves, have qualitative properties (Engels, 1940). In ordinary difference equations period three implies chaos. Four is the highest-order polynomial equation that can be solved in terms of its coefficients. Six is the sum of its prime factors. There are topological theorems applicable to dimensions that are multiples of 7 ± 2. In nature and society qualitative differences often emerge abruptly from threshold transitions or accumulate gradually with quantitative change. Quantity and quality are aspects of the same unitary processes viewed from different perspectives and at different times. This precludes that distinction from separating any disjunct classes of models.

CRITERIA FOR CHOOSING MODELS

Orzack and Sober offer definitions of the criteria for choosing models: "If one model applies to more real world systems than another, it is *more general*" (p. 534). There are two difficulties with this. First, scientific generality is not the same as logical or mathematical generality. A mathematical proposition derived for some set of objects is generalized if it can be shown to apply to all members of a more inclusive set. When we apply that generalization to another object in that inclusive set we specify that object more completely, and all of its relevant properties follow once we have done the specifying. A scientific generalization is quite different. It may apply to "most" tropical forests, or the majority of beetles, or to planktonic communities "most of the time." Such statements make scientific sense, but it would be of no mathematical interest to prove that a class of equations has a real root "often" or even most of the time. Second, the term "applies" must be qualified to take relevance into account. In science a model "applies to" or "fits" a situation if it is capable of analysing those properties that are of interest.

Formal analysis often lumps distinct phenomena under a single heading because of formal similarities. Orzack and Sober define realism as follows: "If one model takes account of more independent variables known to have an effect than another model, it is *more realistic*" (p. 534). This is far too narrow: The adding of independent variables is only one way of attempting to increase realism. Their definition seems to be derived from regression models where the distinction between independent and dependent variables is central to the conceptual framework. Then "taking account" means simply adding independent variables to the model, each with a coefficient that is left undetermined in the uninstantiated case and assigned numerical values when instantiated.

But we can attempt to make models more realistic in other ways as well:

(1) We can add new variables that mutually affect each other, such as the predators or food species affecting the species of interest, relative preferences for one or another habitat, and physiological states. These are not independent variables but covariables. The effects on generality and precision are ambiguous.

(2) We can add a new link between variables already present. For example, in a three-tiered trophic model with a plant, an herbivore and a predator, the predator consumes the herbivore, which consumes the plant. This is the core of the model, the part that is there by definition. But the plant's abundance may serve as a signal that attracts the predator and therefore increases predation. We take account of this by recognizing a direct positive link from the plant to the predator. This increase in realism will reduce generality. The effect on precision will be ambiguous.

(3) We can relax simplifying assumptions, such as symmetry, or the logistic form of the growth equation or constant selection. Realism and generality will be gained at the expense of precision.

(4) We can restrict the domain of application. Thus the growth equation

$$dN/dt = rN$$

is very unrealistic as a general description of population growth, but bacteriologists have used it for differentiating species by restricting the model to the logarithmic phase of growth, where it applies by definition. Realism and precision will be gained at the expense of generality. To make the addition of a coefficient or independent variable stand for all modifications of models is itself to sacrifice generality and realism for spurious precision.

Whether an additional factor increases realism depends on the state of the science at the time. In the original essay I discussed this aspect as the distinction between simplification and oversimplification: In the early years of population genetics, the assumption of constant selection coefficients was realistic enough for the question: "Can weak selection change populations?" The addition of variable environments would have added only complication to this model. But after the question "In addition to genetic drift, can a variable environment cause loss of genetic heterogeneity?" was posed, models with selection coefficients represented as stochastic "white noise" were appropriate. The answer was, yes; a varying selection coefficient may result in the loss of heterogeneity.

This model in turn became unrealistic for considering the problem, "Can the response to selection improve fitness?" Then a white-noise environment is completely unrealistic, and only when an autocorrelated environment was introduced could populations be seen to track their environments. The process continues, modeling frequency-dependent selection, demographic covariables, multiple loci, and so on. At each stage a new question is asked and the assumptions, realistic for the previous question, become unrealistic in the new context.

The four procedures noted above often increase realism, but not always. A paradoxical situation occurs when adding specification to a model actually reduces realism. This happens when the added variables or connections among variables change the level of abstraction of a model. It therefore opens up a whole new domain of variables for possible inclusion in the model and is a kind of silence about those variables at that same level that are not included, an implicit decision that they are not as important as the ones we included.

In this way, the addition of variables can reduce realism, not logically and necessarily but historically and practically. To say that plant growth needs nutrients is a general and realistic statement, and it would be legitimate to represent "nutrients" as a variable in a model. But if we go on to specify only zinc and magnesium explicitly as variables in the model, the situation changes. We have moved to the level of individual nutrients. The silence about nitrogen, phosphorus, potassium, and other nutrients has become a serious misrepresentation. Thus a change in the level of abstraction changes the realism of a model.

Therefore, the more closely the assumptions of a model correspond to the processes and level of abstraction being studied, the more realistic the model is; and the more closely the characteristics of interest correspond to the outcomes of the model, the more realistic the model is. The factor of relevance means that the "same" model will differ in realism for investigators studying different problems and for the same investigator at different times as an investigation proceeds. This makes it impossible to rank models as such for their realism.

Formal analysis prefers to work with properties that belong to the object in itself, independent of its context. Orzack and Sober find "unsatisfactory" what I find delightful: that the level of generality of the Hardy-Weinberg Law and of other laws and models depends on how they are used. They understand this in a very limited sense. They only distinguish instantiated cases from uninstantiated models. But the dependence of a model on how it is applied is more broadly useful: A theoretical result may be applied as a specific claim about a particular object. It may be invoked more generally and realistically but less precisely as an "if nothing else interferes" claim. This may be tied to an argument that nothing else does interfere sufficiently to change the outcome. Or it may be used still more generally as a claim about process rather than result, identifying the direction in which the process being studied acts on the variable without predicting the outcome. All these uses depend on the agenda of the science at the time, the goal of the researcher, and the stage of the investigation.

The Hardy-Weinberg Law, the heterotic theory of polymorphism, the MacArthur-Wilson theory of island biogeography, and others can all be used in these different ways and will be general, realistic and precise according to how they are used.

Orzack and Sober consider precision to be dichotomous. This allows them to make a sharp distinction between uninstantiated models that do not give point predictions and instantiated ones that do. This is too rigid and narrow. Precision is comparative. Models in which equations are specified symbolically in terms of their coefficients are more precise than models that only offer a graph of flow between compartments, but are less precise than numerical equations. Mixed models assign numerical values to some coefficients and leave others as symbols. Numerical solutions also differ in precision depending on the efficacy of measurement and the degrees of freedom. Therefore I see the precision of a model as a degree of specification, with instantiation as a final stage along a continuum.

Nowhere is the unsuitability of formal analysis more apparent than in the study of scientific processes whether in the large (the trajectory of a scientific field or problem area), or in the small (in studying the unfolding of a particular line of research).

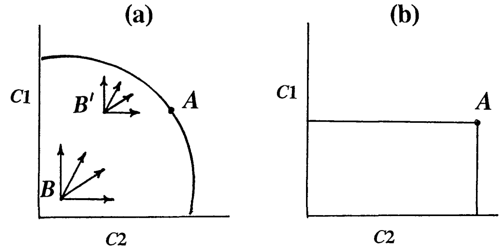

FIG. 1. A MODEL FOR MODEL BUILDING
The axes $C1$ and $C2$ represent two criteria, such as generality and realism. The enclosed space is the set of possible models at a particular time. Fig. 1(a) The initial model B can be improved with regard to both criteria by shifting it upward and to the right, as shown by the arrows. B' is superior to model B. It is closer to the boundary. Here only small improvements are possible with regard to both criteria together, but either one can be greatly improved at the expense of the other. At point A on the boundary any improvement with regard to one criterion must reduce the other. Fig. 1(b) In this model a unique point A is optimal with respect to both criteria.

INCREASING OR MAXIMIZING?

In the original article I made two claims that at the time seemed obvious to the point of triviality, and therefore I did not give them supporting argument. The first was that it is not possible to maximize generality, realism and precision simultaneously. This is a particular case of the more general truth that maxima with constraints cannot be greater than maxima without constraints, and except for very special circumstances will be less. The second claim was about robustness: The reliability of an inference is increased when it is the joint inference of multiple models.

My first assertion was that these three desiderata could not be *maximized* simultaneously. I did not claim that they could never be *increased* together. Indeed that claim would be absurd: We could always change a model in a direction that would make it less general, less realistic, and less precise, so that undoing that procedure would increase all three at once. And indeed adding a term to an equation may sometimes increase all three. But if adding higher-order terms to an equation always increased generality, realism and precision, why would anyone ever stop adding terms?

Figure 1 represents the process of model improvement in a space whose axes are measures of two criteria. The available models are points in this space. Point B represents a poor model near the origin. Almost all changes move B upward and to the right improving both measures. As the model improves toward B', the range of options that increases both criteria becomes more limited, and only small improvements are possible. But options are still available that make big improvements in one at the expense of the other. At the boundary it is possible to improve one measure only by reducing the other. Point A is one possible choice, but it is not better than other choices on the boundary with regard to both criteria.

But it seems that they have a different picture in mind. In Figure 1(b) we show a special configuration of the model space in which all models lie within the rectangle bounded by the axes, the origin, and point A. Then A is a unique best model and there is no trade-off. The burden of proof would be borne by any claimant arguing that the model spaces of population biology are the sort shown in Figure 1(b). Orzack and Sober attempt to do this by the device of the added variable or coefficient discussed below.

Therefore in a formal sense, any examples in which Orzack and Sober show model changes that increase all three criteria at once are irrelevant and would simply be examples of points B and B'.

SOME CONTEXTS OF MODELING

In real scientific practice we regularly make decisions that improve one or more of these desiderata at the expense of another. In each of the contexts of modeling below, the relationships among the desiderata change in the course of an investigation. Suppose we have to account for the dynamics of a particular object of study, a population of fish or the microbial community of a rainforest or the geographic range of a bird, which has already been observed. Then we would start with what we know in general about such objects and discard from the model what is irrelevant to the particular case. We would express relationships such as predation by particular equations, and estimate the appropriate parameters in order to fit the model to the observations.

This is a modeling sequence from the general to the particular, adapting general knowledge to particular cases by successive specifications that might reach the end point of assigning numerical values. At each stage of specification the model is more precise and realistic but less general.

Suppose the task is to provide a general model for a class of objects such as lakes or dry forests. This can be approached from opposite ends. If we start with a general model of ecosystems, the procedure is one of increasing specification, leaving out what is relevant only to other ecosystems in order to study the particular properties of the one at hand. From the other end, we might start from a single lake or forest, observe its properties, and ask how general the results are. Then we go through an opposite procedure of reducing specification, relaxing assumptions, and losing precision to gain generality.

We might want to examine the ecological or evolutionary significance of a particular phenomenon such as mutualism or heterosis. Here we would usually start with the phenomenon in a pure form, without complicating factors. We might use several very special cases for ease of handling in order to get a feel for the problem. If a conclusion seems to hold for all of them we might try to prove it generally, with fewer special assumptions. Or we may opt for increased realism by using more complex and intractable numerical models that are once again precise but not general. Many investigators now combine these approaches, scanning a phenomenon with simplified models and then verifying the results with simulations.

We might be trying to find ways to recognize a given phenomenon such as density dependence or heterosis or to distinguish among alternative explanations. This is similar to the previous case except that now we examine the consequences of different models in order to find conclusions that differentiate among them.

Suppose finally that the task is to show that a common sense expectation is not necessarily so. Here the first models are usually quite simple. If we need a life table we might use a negative exponential. Age-specific reproductive values might be represented by a triangle, latency by a fixed delay. If we guess that some result depends on a periodic environment we may chose alternating conditions or a sinusoidal curve. Since we are trying to model a process rather than a specific case, we are not concerned with additional factors that might be present in some situations and not others. Thus we have become inclined toward more general, unrealistic and precise models. Factors we know to be operating are ignored, circumstances such as symmetry that are uncommon in reality are assumed in order to reveal more clearly the operation of the factors we are interested in.

These different activities are idealized types. Real research may shift among them in a zigzag course, now modeling the rich interactions of a system in exquisite detail, now reducing magnification and seeing only the outline structure, moving from the general to particular cases, finding results that might be generalized, specifying the terms of a model for computer simulation or leaving them qualitative for analytic work. It is quite common now to use analytic models in conjunction with simulation models.

As we make the model more specific we are losing generality in two ways: Each step that adds new variables or new relations among them excludes those cases in which these particular variables or relations are not operative. And once we have introduced new variables we may have to say something about the mathematical forms of their interactions, the exact equations, or the magnitudes of their parameters. The last stage in specification is instantiation.

Orzack and Sober note that an instantiated model (one in which numerical values are assigned to the parameters so that numerical predictions can be made about the variables) is less general than an uninstantiated one, but they call this observation trivial. Full instantiation, however, is the end point of a continuum of increasing specificity, and generality is lost at each step.

It is plausible to expect that as we add variables, relations, or parameters to a model, we increase precision. Indeed this is often the intention, but not always the result. Some of the variables may not be measurable or only very poorly measurable. Our understanding of physiology now includes emotional states that are much more difficult to measure than conventional physiological variables, but should be included in the models. Epidemiological models may have to include the behavior of medical bureaucracies in response to the progression of an epidemic. Since model building usually begins with variables we are more familiar with, the addition of variables often (not logically and necessarily but historically and in practice) improves realism but reduces precision.

For many models increasing complexity magnifies the sensitivity of the outcome of the model to the parameters. The end point of this progression can be chaotic trajectories. Therefore, even if each additional parameter is measurable to the same degree of accuracy, the outcome may be less predictable.

Because of these practical difficulties, the very large and complex models used for biome studies, the world economy, or epidemics have not been conspicuously more precise in their predictions than simpler models.

UNDETERMINED COEFFICIENTS

A key to their argument is the use of the method of adding undetermined coefficients. Each of the additional factors that may be relevant to some particular cases of a general phenomenon can be included in the model, multiplied by a coefficient representing the magnitude of its effect. If that effect can be allowed to be zero, then the expanded model "includes" the previous one as a special case. Each of these factors may be relevant somewhere, but in each particular case most of them will be zero. In that way a large number of quite different situations can be covered formally by the same model, but in each model most of the factors have zero effect. The model will become cluttered by predators with predation rate zero, competitors with competition coefficients of zero, zero time lags, and so on. Mathematically, zero can be thought of as just any old number, and when we fit parameters it is no big deal if we decide that some are zero. Thus Orzack and Sober compare two growth equations:

$$dN/dt = rN$$

and

$$dN/dt = rN + \alpha N^2.$$

They claim that the second is more general than the first since it includes the case of

$\alpha = 0.0$. Mathematically this is true: Zero is just another number. But biologically it is not true. Each factor in the model must be taken into account, looked for, thought about, perhaps excluded or enhanced experimentally. It is a serious matter to include another factor in a model. And in terms of their consequences, nonpredation is not just a special kind of predation, noncompetition is not a special kind of competition, nor is density independence a special limiting case of density dependence.

When coefficients such as α are added, we do not treat zero as just another possible parameter value. We usually test for the coefficient being "significantly different from zero," not different from say -0.017 or any other negative number. That procedure decides between density dependent and density independent models. An alternative approach would be simply to accept the best estimate of α, which will almost always be different from zero. Then we are implicitly adopting density dependence. Even the smallest negative α gives qualitatively different results from $\alpha = 0.0$ in terms of equilibrium, or of sensitivity to different fixed values of r in different populations, or of dependence on past values of r, or of correlations between population size and environmental factors if r varies over time, or of the limiting probability distribution of N with random r.

So far I have been focusing the discussion on mathematical models, but much the same situation obtains in experimental design. If the problem is a general one, such as the role of crowding or of a fluctuating environment, the choice of species may be determined by ease of manipulation and also representativeness. That is, if we want to be able to generalize the results we choose a species that is in some sense "typical," at least in relation to the problem being studied. Enclosed populations in the laboratory can be kept at predetermined temperatures, whereas humidity is more difficult to manage. Therefore, temperature is used as a typical nonconsumable environmental factor. Experimental variables can be measured frequently, and manipulations performed according to plan. The desired phenomenon may be amplified above the background noise of nature for easy observation. Here precision is gained at the expense of realism. Field observations and experiments, on the other hand, are done under more realistic conditions, but with less precision.

ROBUSTNESS

Orzack and Sober deal with robustness with a logical model. They distinguish three cases: If we know that one of a set of models $M_1, M_2, \ldots$ is true and that each M_i implies R, then R is true. They then consider cases where we know that each is false, and assert that in these cases the fact that R is implied by all of them does not make it true. And finally, if we do not know whether any of the set is true, then their joint implication is also not proven. They then illustrate their analysis.

This analysis, however, is not relevant to the problem of robustness that was posed: In order for a model to be useful it must be both similar to and different from the object or process it models. Therefore it contains parts that may be true and parts that are either false or are of much more limited validity. The problem is how to identify the implications of the true part either for prediction or testing.

The logical structure of the argument is quite different from Orzack and Sober's representation of it. Let C be the common part of all the models, the core relationships we are either confident of or wish to test. Let V_i be the variable part of the model introduced for convenience or because each might hold for some cases. Then if C together with V_1 (the intersection of C with V_1 or their logical product CV_1) implies R, and if CV_2 implies $R, \ldots$, it follows that the intersection of C with the union of the V_i implies R:

$$CV_1 + CV_2 + CV_3 + \cdots = C(V_1 + V_2 + V_3 + \cdots) => R.$$

If the set of V_i exhausts all the admissible alternatives, then

$$V_1 + V_2 + V_3 + \cdots = 1 \text{ and } C \text{ implies } R:$$

$$\sum_{i=1}^{\infty} V_i \xrightarrow{\lim} 1 \left[C\left(\sum_{i=1}^{\infty} V_i\right) => R \right] = (C => R).$$

If the V_i's do not exhaust the possible alternatives, at least the more inclusive the set of V's, the more we can have confidence that C implies R. If we feel that the set of V_i's spans a wide enough range of possibilities, then we may generalize to claim that C usually implies R, a result that is not very exciting as a mathematical theorem but may be good biology.

The point is that all systems have more going on than the process we are interested in, and each system has some things we have not taken into account. The search for robust theorems reflects the strategy of determining how much we can get away with not knowing, and still understand the system.

Orzack and Sober are worried that the robustness strategy seems to propose a way to truth independent of observation. This is not the case. Observation enters first in the choice of the core model and the selection of plausible variable parts, and later in the testing of the predictions that follow from the core model. Multiple models sharing that same core help to find the consequences of that core when we are unable to offer a general proof that C implies R. Thus the search for robustness as understood here is a valid strategy for separating conclusions that depend on the common biological core of a model from the simplifications, distortions and omissions introduced to facilitate the analysis, and for arriving at the implications of partial truths. The use of multiple models is a common practice for this reason, either to strengthen the conclusion or to guide us in looking for a general result.

CONCLUSIONS

(1) Formal analysis depends on imposing a sharp differentiation among objects that do not have clear boundaries. It attempts to evaluate models outside of their context to develop an investigation or the history of their science. It extinguishes differences among quite different procedures that share a formal property.

(2) Models do not fall into mutually exclusive classes but lie on a multidimensional continuum. Three of these axes are generality, realism and precision. Others are manageability and understandability. Instantiation is an end point on the axis of precision.

(3) The location of a model on this continuum is not determined by the model itself but depends on the changing contexts in which it is used. The problem to be solved and the level of abstraction of the investigation can change the generality, realism and precision of a given model. The same formal act, for example, adding terms to a model, may increase or decrease generality, realism, or precision.

(4) The same model can have different meanings depending not only on how it is used but also on the stage of an investigation and the state of the science. It can be taken as abstract mathematical relations. It can be a claim about what happens when nothing else interferes. This may be coupled to claims that nothing else does interfere in some particular case. It can be interpreted as identifying one of the processes that always influences a trajectory but only sometimes is sufficient to determine the outcome.

(5) The strategy of model building consists of deciding how to move along this continuum. It is concerned with the processes by which different desiderata support and interfere with each other. It is not concerned with ranking or measuring models.

(6) In a wide range of scientific practices, including modeling a particular object or a class of objects, examining the significance of a phenomenon, exploring whether an accepted proposition really holds, deciding among alternative explanations, or determining how to recognize a phenomenon, there is a tradeoff among generality, realism and precision.

(7) In formal analysis, a model that is partly true and partly false is false. Therefore, models can be divided into true and false. In science, most of the models we work with are partly true and partly false.

(8) We may strengthen our confidence in the implications of some assumptions by using ensembles of models that share a common core of these assumptions but also differ as widely as possible in assumptions about other aspects. Then the more the variable part spans the range of plausible assumptions, the more valid the claim that the conclusions shared by all of them depend on the constant part. If we also have confidence that the constant part is true, then we have strong support for the

claim that the conclusion is generally true. This gives robustness to the conclusions.

A formal/analytic framework is not the appropriate domain for evaluating a model of research strategy.

ACKNOWLEDGMENT

I am indebted to Rosario Morales for her help in clarifying and expressing the concepts in this response.

REFERENCES

Engels, F. 1940. *The Dialectics of Nature*. International Publishers, New York.

Levins, R., and R. C. Lewontin. 1985. *The Dialectical Biologist*. Harvard University Press, Cambridge.