3

Confirmation of Ecological and Evolutionary Models

1. INTRODUCTION

This paper concerns hypothesis testing and confirmation in evolutionary and ecological theory. I outline specific criteria used in evaluating evidence for theories, and demonstrate the use of each criterion through examples from various branches of evolutionary biology and ecology. The philosophical discussions in Roughgarden (1983), Strong (1983), Simberloff (1983), and Quinn and Dunham (1983), which focus on a Popperian approach to theory testing and acceptance, present some important issues in the testing of evolutionary and ecological explanations. I find that imprecision of criteria of testing and confirmation is the weakest point in these discussions. As an alternative to a Popperian approach (as defended by, e.g., Simberloff 1983), and to other approaches commonly cited by biologists (e.g., J. Platt's "strong inference"), I suggest a new description of confirmation that includes a detailed classification of the ways in which a theory may be confirmed (see, e.g., Oster and Wilson 1978 for an endorsement of Platt's 1964 paper).

Roughgarden (1983) proposes that one establishes an empirical fact in science "by building a convincing case for that fact." What counts as a convincing case depends, according to Roughgarden, on "common sense and experience" (pp. 583–584).

As Strong (1983) rightly points out, the appeal to common sense is problematic; common sense may not be "common" to all scientists concerned, and it says nothing about testing. Strong's solution to problems of confirmation and testing is not much of an improvement, though: "our

regard would be greatest for theories that have passed multiple independent, tough tests" (1983, p. 638). One is immediately inclined to ask what a "tough" test is, what constitutes "passing," and whether theories that have passed some test, but not "multiple, independent, tough tests," could be acceptable. Quinn and Dunham, in addressing the same problem, conclude, "Theories are embraced when, in part, a relatively simple explanation seems to account satisfactorily for much of a complex set of observations, and are abandoned or modified as the weight of post hoc additions becomes a burden, and other, comparably simple and appealing viewpoints are suggested" (1983, p. 613).

The chief weakness of these discussions is the lack of precision. It is unclear what it is for an explanation "to count satisfactorily" for a set of observations, or for a theory to "pass multiple, independent, tough tests," or for a set of evidence to count as "a convincing case." In what follows, I offer a view of confirmation that reveals various factors determining the support of theories by data. In doing so, I pursue a naturalistic approach to the philosophy of science, compatible with the approach recently defended by Ronald Giere (1985). I present a description of the various ways in which empirical claims about models can be confirmed; it is not assumed that all accepted models are supported in all of these possible ways. I would also like to emphasize that this schema is not intended as straightforwardly normative, that is, as a checklist for "good" or "well-confirmed" theories, but, rather, as a list of the types of support deemed significant within the disciplines of evolutionary biology and ecology. After a brief introduction to the various forms of confirmation, I discuss examples taken from evolutionary theory and ecology.

2. CONFIRMATION

Throughout this paper, I use the semantic view of scientific theories. On the semantic view, scientists present descriptions of ideal systems; a set of (logical) models can be defined relative to these structures. In presenting a scientific theory, scientists can be understood to be presenting sets of models, to be used to explain the world (for a detailed presentation of the semantic view, see Suppes 1957, 1967; van Fraassen 1970, 1972, 1980; Suppes 1972, 1977; for application of the semantic approach to evolutionary theory, see Beatty 1980, 1981, 1982; Thompson 1983, 1985; Lloyd 1983, 1984, 1986a, 1986b).

According to the state-space version of the semantic view (see van Fraassen, Suppe) models are specified by defining the variables with which the system in nature is described, the laws that describe the changes or structure of the system, and parameters, quantities in the model with a constant value.

Explanations involve claims regarding the applicability of a model to a natural system. Giere has characterized these claims, which he calls "theoretical hypotheses," as having the following general form: "The designated real system is similar to the proposed model in specified respects and to specified degrees" (1985, p. 80). For instance, a population geneticist might claim that a certain natural or laboratory population is a Mendelian system, that is, that it conforms with a Mendelian theoretical model. Certain attributes of the system – the distribution of gene frequencies, for instance – are thus explained through the homomorphism of the natural population to the theoretical model.

The activity of confirming models is more accurately described as confirming the empirical claims made about models, that is, the claims stating that a natural system (or kind of natural system) is homomorphic in certain respects to the model.

Patrick Suppes has given an analysis of the hierarchy of theories needed to link the natural system to the ideal system described by the theory (1962). In his (admittedly preliminary) study, Suppes presents three levels of models used to relate the empirical data to the theoretical model: theoretical models, models of the experiment, and models of data. The basic idea is that the logical models of the theory are too broad, because a model of the experimental data might fail to match precisely a theoretical model; for example, concepts or entities might be used in the theory that have no observable analogue in the experimental data (see Suppes 1962, p. 253). A series of gradually more specified models may be defined in order to make direct comparison possible.

The model of the experiment is the first step in specifying the theoretical model enough to enable comparison with the empirical results. The model of the experiment is a definition of all possible outcomes of a particular experiment which would satisfy the theoretical model. The next step in Suppes's hierarchy is the model of the data. In the context of a specific performance of an experiment, a portion of the total possible space of outcomes can be defined, each of which is a possible realization of the data. A possible realization of the data counts as a model of the data when it fits the model of the experiment well enough according to

statistical goodness-of-fit tests. Models of data are usually restricted to those aspects of the experiment that have variables in the theory (1962, p. 258)

My use of "empirical claim regarding the model" is essentially equivalent to Suppes's "model of the experiment," in that it is more specified, more concrete than the abstract theory itself. For our purposes, the issues regarding "fit" (Sections 2.1, 2.3) can be understood in terms of statistical tests involving models of data and models of the experiment (in Suppes's terms), although I shall not use the distinction between models of the theory and models of the experiment. The issues involving independent support for aspects of the model (Sections 2.2, 2.3) remain outside the goodness-of-fit relations discussed by Suppes. Hence, I shall not use Suppes's terminology, although I see my description of confirmation as compatible with his.

In general, empirical evidence confirms a claim if the evidence gives additional reason to accept the claim. Evaluation of confirmation involves an evaluation of the support of claims regarding the applicability of a model to evidence, that is, an evaluation of the relation between the data and the model. Past discussions of confirmation have often been too general and too vague to be of real use. I take it that the semantic approach provides a precise framework within which theoretical, methodological, and empirical issues can be discussed.

I suggest that three distinct factors bear on the confirmation of empirical claims about models through data: fit between model and data; independent testing of aspects of the model; and variety of evidence, which can itself be of three sorts (see below). Traditionally, philosophical discussions of confirmation have concentrated on the fit between data and model, and on variety of instances of fit, which is one type of variety of evidence. Detailed examples of each type of confirmation will be given after a brief summary of the types.

2.1. Fit between Model and Data

The most obvious way to support a claim of the form "this natural system is described by the model," is to demonstrate the simple matching of some part of the model with some part of the natural system being described. For instance, in a population genetics model, the solution of an equation might yield a single genotype frequency value. The genotype frequency is a variable in the model. Given a certain set of input variables (e.g., the initial genotype frequency value, in this case), the output values

of variables can be calculated using the rules or laws of the model. The output set of variables (i.e., the solution of the model equation given the input values of the variables) is the outcome of the model. Determining the fit involves testing how well the genotype frequency value calculated from the model (the outcome of the model) matches the genotype frequency measured in relevant natural populations. Fit can be evaluated by determining the fit of one curve (the model trajectory or coexistence conditions) to another (taken from the natural system); ordinary statistical techniques of evaluating curve-fitting are used for this evaluation.

2.2. Independent Support for Assumptions

Numerous assumptions are made in the construction of any ideal system. These include assumptions about which factors influence the changes in the system, what the ranges for the parameters are, and what the mathematical form of the laws is. On the basis of these assumptions, the models are given certain features. Many of these assumptions have potential empirical content. That is, although they are assumptions made about certain mathematical entities and processes during the construction of the ideal system, when empirical claims are then made about this ideal system, the assumptions may have empirical significance. For instance, the assumption might be made during the construction of a model that the population is panmictic, that is, that all genotypes interbreed at random with each other. The model outcome, in this case, is still a genotype frequency, for which ordinary curve-fitting tests can be performed on the natural population to which the model is applied. But the model can have additional empirical significance, given the empirical claim that a natural system is a system of the kind described in the model. The assumption of panmixia, as a description of the population structure of the system under question, must be considered part of the system description that is being evaluated empirically. Evidence to the effect that certain genotypes in the population breed exclusively with each other (i.e., evidence that the population is far from panmictic) would undermine empirical claims about the model as a whole, other things being equal. In other words, the assumption that genotypes are randomly redistributed in each generation is intrinsic to the Hardy-Weinberg equilibrium. Hence, although the assumption that the population is panmictic appears nowhere in the actual definition of the model type – that is, in the law formula – it is interpreted empirically, and plays an important role in determining the empirical adequacy of the claim.

By the same token, evidence that the assumptions of the model hold for the natural system being considered will increase the credibility of the claim that the model accurately describes the natural system.

In other words, it is taken that direct testing provides a stronger test than indirect testing, hence a higher degree of confirmation if the test is supported by empirical evidence. Direct empirical evidence for certain empirically interpreted aspects of the model that are not included in the state variables (and thus are confirmed only indirectly by goodness-of-fit tests) therefore provides additional support for the application of the model.

This sort of testing of assumptions involves making sure that the empirical conditions for application of the ideal description actually do hold. In order to accept an explanation constructed by applying the model, the conditions for application must be verified.

The specific values inserted as the parameters or fixed values of the model are another important aspect of empirical claims. In some models, mutation rates, and so on, appear in the equation – part of the task of confirming the application of the model involves making sure that the values inserted for the parameters are appropriate for the natural system being described.

Finally, there is a more abstract form of support available, in which some general aspect of the model, for instance, the interrelation between the two variables, or the significance of a particular parameter or variable, can be supported through evidence outside the application of the model itself.

2.3. Variety of Evidence

Variety of evidence, of which there are three kinds, is an important factor in the evaluation of empirical claims. I discuss three kinds of variety of evidence here: (1) variety of instances of fit; (2) variety of independently supported model assumptions; and (3) variety of types of support, which include fit and independent support of aspects of the model.

Often, empirical claims are made to the effect that a model is applicable over a more extended range than that actually covered by available evidence. This extrapolation of the range of a model can be performed by simply accepting or assuming the applicability of the model to the entire range in question. A more convincing way to extend applicability is to offer evidence of fit between the model and the data in the new part

of the range. Provision of a variety of fit can thus provide additional reason for accepting the empirical claim regarding the range of applicability of the model. For instance, a theory confirmed by ten instances of fit involving populations of size 1000 (where population size is a relevant parameter) is in a different situation with regard to confirmation than a theory confirmed by one instance of each of ten different population sizes ranging from 1 to 1,000,000. If the empirical claims made about these two models asserted the same broad range of applicability, the latter model is confirmed by a greater variety of instances of fit. That is, the empirical claim about the latter model is better confirmed, through successful applications (fits) over a larger section of the relevant range than the first model. Variety of instances of fit can therefore provide additional reason for accepting an empirical claim about the range of applicability of a model.

Variety of fit is only one kind of variety of evidence. An increase in the number and kind of assumptions tested independently, that is, greater variety of assumptions tested, also would provide additional reason for accepting an empirical claim about a model. This is just the sort of confirmation Thompson (1985) found lacking in many sociobiological explanations of human behavior. Thompson argues that the difference in the acceptability of sociobiological explanations of insect behavior and of human behavior is that the auxiliary theories used in applying genetic models to human beings are unsupported. For instance, there is no neat physiological model linking genes to a particular phenotype in the case of homosexuality, as there is, for example, in sickle cell anemia (see Thompson 1985, pp. 205–211). The assumptions needed to apply the genetic models to human beings are largely unsupported empirically, and on these grounds, the sociobiological explanations are rejected, argues Thompson.

The final sort of variety of evidence involves the mixture of instances of fit and instances of independently tested aspects of the model. In this case, the variety of types of evidence offered for an empirical claim about a model is an aspect of confirmation.

According to the view of confirmation sketched earlier, claims about models may be confirmed in three different ways: (1) through fit of the outcome of the model to a natural system; (2) through independent testing of assumptions of the model, including parameters and parameter ranges; (3) through a range of instances of fit over a variety of the natural systems to be accounted for by the model, through a variety of assumptions tested, and through a variety of types of evidence.

3. APPLICATIONS

3.1. Fit

3.1.1. ISLAND BIOGEOGRAPHY The area of island biogeography offers a straightforward example of confirmation of a model application through its fit with empirical findings. The first attempt at a quantitative theory of island biogeography was made by MacArthur and Wilson (1963, 1967). They presented a mathematical model for determining the equilibrium numbers of species on islands. In this model, the equilibrium number of species is represented by the point of intersection of the immigration and extinction curves of the island, which are drawn as a function of the number of species already present and the distance from the mainland.

The assumptions of the model include speculations about the equations of the curves, and about the effects of varying both island size and the distance from the island to the source of the immigrating fauna.

In the first empirical tests of the MacArthur-Wilson model, the investigators assumed that it was most important to show: (1) that there exists an equilibrium number of species, and (2) that the MacArthur-Wilson model accurately represents the relationship among the species equilibrium number and the species turnover rate (see Wilson and Simberloff 1969).

Wilson and Simberloff chose to test the two empirical claims above by removing fauna from seven very small islands (Mangrove Islands in the Florida Keys), and surveying the colonization results. The results are taken to support the empirical claims outlined earlier.

In support of the claim that an equilibrium number of species does exist, Simberloff and Wilson (1969) cite three types of evidence: first, the number of species on the control (nondefaunated) islands did change during the period of the experiment; second, untreated islands with similar area and distance from source faunas have similar equilibrium numbers of species to those arrived at on the experimental islands; and, third, there was an increase of species on the experimental islands to approximately the same number as before defaunation, and then oscillation around this number.

These types of evidence can be taken as instances of fit between the outcome of the model (the equilibrium number of species) and the empirical findings (the actual number of species, and the pattern over time). Later, independent investigations found additional support from other bodies of data for the existence of an equilibrium number of species (Diamond, 1969).

Simberloff and Wilson also examined the actual turnover rate on the islands, and compared it with the predictions resulting from the MacArthur-Wilson model. They found that the experimental results were roughly consistent with the model prediction. Again, the outcome of the model (somewhat loosely) fitted the empirical findings.

In follow-up research on these islands, the investigators offer further confirmation for the existence of an equilibrium number of species and the accuracy of the model equation through further instances of fit (Simberloff and Wilson, 1970).

3.1.2. PUNCTUATED EQUILIBRIUM Let us examine the type of support offered for the controversial theory of punctuated equilibrium, presented by Niles Eldredge and Stephen Jay Gould (Eldredge 1971; Eldredge and Gould 1972; Gould and Eldredge 1977). For all the controversy, the theory of punctuated equilibria is a relatively simple sort of model, which has two main features. First, speciation by branching of lineages (as in allopatric speciation, i.e., speciation by geographic isolation, see Mayr 1963) is the primary source of significant evolutionary change, according to the model, rather than the gradual transformation of lineages (phyletic transformation). Second, speciation occurs rapidly in geologic time, and is followed by long periods of stasis.

As Gould emphasized, the model presents a picture of the relative frequencies of gradual phyletic transformation and punctuated equilibrium (Gould, 1982). Gould and Eldredge's 1977 paper contains a section entitled "Testing Punctuated Equilibria," in which they present various approaches to the fitting of the model to the data. Because punctuated equilibrium is a model about relative frequency, the general approach to confirmation is to test the distribution of instances. Gould and Eldredge discuss two ways in which the frequency distribution can be tested. First, the model can be applied to individual cases (of evolutionary change) with the right sort of features. The authors discuss the merits of a number of cases of individual fit. In some cases, Gould and Eldredge claim that the data presented as confirmation for the gradualist model actually have a tighter fit to their punctuated equilibria model (see especially the discussion of Gingerich 1977, pp. 130–134). In all of these cases, the issue is whether the data presented conform sufficiently to the predictions or structures of the models in question.

A second type of test involves examination of quantifiable features of entire clades or communities, and comparison of these features to results expected from the model. Stanley (1975) devised a number of

this sort of high-level test for the punctuated equilibria model. In several tests, he demonstrates that, given the estimated time span and major morphological evolution, the gradualist model produces rates of evolutionary change that are too slow. Under the punctuated equilibria model, however, the evolutionary changes could conceivably have taken place within the estimated time span. The support being offered for the theory of punctuated equilibria here is that it has a better fit with the data than the gradualist model (see Stanley 1975, 1979; Gould and Eldredge 1977, pp. 120–121).

3.1.3. STATISTICAL POWER Statistical analysis is commonly used in evaluation of the fit of the empirical data to the model. A recent discussion on the power of statistical tests emphasizes that there is more to fit of model to data than maintaining a level of alpha (Type I error) under .05, where the probability of committing a Type I error is the probability of mistakenly rejecting a true null hypothesis (Toft and Shea, 1983). Toft and Shea argue that the probability of committing Type II errors (in which the investigator mistakenly fails to reject a false null hypothesis) is important, and has been neglected. The "power" of a test is the probability of not committing a Type II error.

A specific problem has arisen regarding power of tests in investigations of competition theory in ecology – the failure to demonstrate that a certain factor has an effect on a system is sometimes taken as the demonstration that the factor has no effect. Such a conclusion is unwarranted by the evidence, as well as by statistical theory, as Toft and Shea point out, and they call on ecological investigators to include power tests in their results in the future (Toft and Shea, 1983).

Toft and Shea's criticism of the investigators' method can be understood as an elaboration of the definition of a "good fit." It is not enough, they argue, to have a low probability of Type I errors in evaluating the fit of a model to a natural system; without consideration of Type II errors, tightness of fit is open to misinterpretation, and can be used to support false claims.

3.2. Independent Testing

3.2.1. PUNCTUATED EQUILIBRIUM To return to the topic of punctuated equilibria, Gould and Eldredge, in their discussion of the empirical support for the model, included a section on "indirect testing" (see 1977, pp. 137–129). Under the schema used in this paper, these tests can be

understood as independent tests of the empirical assumptions of the model type.

The situation is as follows: one primary assumption of the model of punctuated equilibrium is that a major amount of genetic change can and does occur in the speciation event itself. That is, the major genetic differences between species must be laid down during the process of speciation rather than gradually through the duration of the species' existence. Gould and Eldredge consider evidence for the concentration of genetic change in speciation events as an important possible source of confirmation of their theory. For instance, in one case (although they found much of the evidence ambiguous with regard to their model), they reported, "We are pleased that some recent molecular evidence . . . supports our model" (1977, p. 138). The case in point offers evidence for the concentration of genetic change in speciation, and thus provides independent confirmation for an important assumption of their model.

Note the difference between fit and independent testing. In the case of fit, the model outcome, which involves the relative frequencies of gradual phyletic transformation versus punctuated equilibria, is compared to the actual frequencies of these forms of speciation. In independent testing, an assumption that is important in constructing the model is evaluated separately from the model itself.

3.2.2. POPULATION GENETICS Population genetics theory consists of a large number of related models based on the Hardy-Weinberg equilibrium, which is in turn based on Mendel's basic laws of inheritance. Because the Hardy-Weinberg "law" is an equilibrium equation, any mathematical descriptions of changes in the system being described must involve parameters inserted into expanded versions of the Hardy-Weinberg equation which produce the correct changes in the models. For instance, take a model that will give you the gene frequency of A in the next generation. A large amount of change mutation from a to A will effect the outcome of the model, so the mutation rate, μ, of a to A is included as a parameter in the model. Similarly with migration and selection. In other words, models for gene frequency changes must include factors that visibly affect gene frequencies.

Population genetics models, for the most part, yield single gene frequencies or distributions of gene frequencies. These frequencies and distributions are the part of the model tested for fit. But empirical and experimental testing is also done on the parameter values and their ranges. The subject of mutation and genetic variability, for instance, has served as a

central issue in population genetics research. Although the theoretical problems will not even be mentioned here, the point is that it has been vital to the success and acceptance of population genetics models that the empirical assumptions and parameter values be tested (see Dobzhansky 1970; Lewontin 1974; Mayr 1982).

Determining the value of the mutation parameter of a particular gene is a task theoretically and methodologically distinct from testing the accuracy of the population genetics model in determining gene frequencies. For example, one can find tables of mutation rates for various genes in various organisms; these tables are the results of counting (traditionally done through inbreeding experiments), rather than being the results of calculations done using the Hardy-Weinberg equilibrium. The general idea is that the parameter is isolated and tested separately from the model in which it appears.

3.2.3. GROUP SELECTION Michael Wade, in his discussion concerning models of group selection, examines certain key assumptions common to the models. Wade argues that these (speculative) assumptions are unfavorable to group selection, that is, in these models, group selection is considered an important cause of changes in gene frequency only under a very narrow range of parameter values. Wade challenges the empirical adequacy of the model assumptions and suggests alternative assumptions derived from empirical results (Wade 1978).

The mathematical models of group selection examined by Wade involve general assumptions about extinction, dispersion, and colonization. One of the five assumptions challenged by Wade is that group selection and individual selection always operate in opposite directions. That is, it is assumed in the models that an allele that is favored by selection between groups would be selected against on an individual level. According to Williams (1966), group selection should never be called up if individual selection can be used to explain the evolution of a trait. Thus, if individual selection and group selection are in the same direction, group selection might never be appealed to. Group selection will only be discussed in cases in which it operates in opposition to individual selection.

Wade argues against this assumption, pointing out that any trait favored by group selection (i.e., any trait that increases the likelihood of successful proliferation of the population or decreases the likelihood of extinction) also could be favored by individual selection. He offers

experimental results of group and individual selection acting in the same direction (see Wade 1976, 1977, 1978).

3.3. Variety of Evidence

3.3.1. DARWIN'S EVIDENCE Darwin considered the large variety of evidence supporting his theory of evolution by natural selection to be partial grounds for accepting the theory. He wrote, "I have always looked at the doctrine of natural selection as an hypothesis, which if it explained several large classes of facts, would deserve to be ranked as a theory deserving acceptance" (1903, 1:139–140).

As far as Darwin was concerned, the theory of natural selection does account successfully for "several large classes of fact," including the principle of classification of living things, embryonic resemblance among organisms of very different taxa, and the adaptation of living beings to each other and to their environments (1919, 2:207). In other words, models based on the concept of natural selection fit the empirical observations in a wide variety of fields. This is an example of the first kind of variety of evidence, that is, variety of fit. The existence of adaptive characters in organisms is accounted for by referring to a natural selection model in which those organisms that are well adapted to their environment survive and reproduce at a proportionally higher rate than organisms without the adaptive mechanisms. Eventually, then, the adaptive mechanism would be expected to become a fixed trait in the population (given the right conditions of heredity, etc.). Similarly, natural selection models can account, for instance, for the strange fact that at a certain age, it takes a trained eye to distinguish human from chicken from fish embryos (Darwin 1964, pp. 439–440). The resemblance is understood as a result of common ancestry. Common ancestry also holds the key to taxonomic classification; classification becomes genealogy – the tracing of lineages – according to Darwin (1919, 1:485; 1964).

Thus, Darwin took it as a virtue of his theory of natural selection that it could account for a wide range of natural phenomena, that is, that it exhibited variety of evidence (see Lloyd 1983, for discussion).

3.3.2. POPULATION GENETICS Empirical support for basic population genetics model types (e.g., single-locus models based on the Hardy-Weinberg equilibrium) exhibits a different sort of variety of evidence from that claimed by Darwin for the theory of natural selection. The

mathematical models used to calculate equilibria and changes in gene frequencies have been shown to fit a wide range of natural populations (e.g., those summarized in Dobzhansky, 1970). In addition, the parameters of many population genetics models have been evaluated and tested separately, as discussed in Section 3.2.2 of this paper.

Hence, population genetics models are not confirmed by fit alone but also by independent testing of model assumptions and a pattern of fits over a range of actual populations. The empirical support for population genetics represents more than one type of evidence, that is, it exhibits the third kind of variety of evidence, namely, variety of types of support.

3.3.3. MARINE ECOLOGY Part of a recent upheaval in ocean ecology involves a debate about the appropriate range of testing for a model, that is, the variety of fit of the model being tested. Before 1979, it was thought that phytoplankton growth rates and nutrient uptake (usually ammonium uptake) rates were coupled temporally. Models constructed to represent these rates under steady-state conditions had been successful in laboratory studies. Assuming the accuracy of this model type, however, led to a puzzle about the natural, oceanic systems: studies of oceanic waters showed the level of nitrogenous nutrients (primarily ammonium and nitrate) to be undetectably low, even though the data from photosynthetic activity indicated that the phytoplankton was absorbing nitrogenous nutrients (McCarthy and Goldman 1979, p. 670).

Furthermore, research on the chemical composition of laboratory and oceanic phytoplankton showed that the chemical composition of oceanic populations was most similar to laboratory populations growing at near maximal rates. However, in order to grow at such high rates, phytoplankton need high ambient nutrient levels (Goldman, McCarthy, and Peavey 1979). Thus, their chemical composition suggested that oceanic phytoplankton must be experiencing high nutrient levels, but the open ocean nutrient levels were extremely low.

Some light was shed on the puzzle when laboratory studies showed that phytoplankton are capable of rapid nutrient uptake. This, and the ability to store nitrogen, would make it possible for them to have a maximum growth rate, even when nutrient concentrations are very low. If rapid nutrient uptake occurs, then it is not necessarily true that growth rates are tied to nutrient uptake rates, as assumed in the previously accepted models, and found in laboratory studies carried out at steady state (Goldman, McCarthy, and Peavey 1979, p. 213). McCarthy and Goldman speculated that individual phytoplankton cells might encounter minute

zones of elevated nitrogen levels, which they could absorb very quickly (1979). This would account for high rates of production despite the low observed nutrient level.

Phenomena of such a small scale are not taken into account when considering steady-state situations. Under steady-state conditions, the medium is homogeneous in space and time. That is, organisms are not subject to a feast and famine existence; therefore, rapid nutrient uptake and storage are phenomenologically invisible. Goldman et al. write: "to explore further questions [concerning nutrient dynamics] involves new approaches for studying microbial interactions on temporal and spatial scales that are far smaller than were previously assumed to be important" (1979, p. 214).

Later studies confirmed that differences in methods of measuring parameter values made significant differences to the experimental results. It had usually been assumed that the rate of nutrient uptake was linear over the course of the hours or tens of hours in an experiment. When short-term nutrient uptake responses were tested, however, they were found to be nonlinear (Goldman, Taylor, and Glibert 1981). In one experiment, the phytoplankton completed uptake of the nitrogenous nutrients during the first two hours of the experiment. If the measurements had been performed in the usual time span, for example, after 24 hours, the estimates of nitrogen turnover rates would have been an order of magnitude off. Estimates of phytoplakton growth rate based on these nitrogen turnover rates would in turn have been "in gross error" (Goldman, Taylor, and Glibert 1981, p. 146). The investigators concluded that choice of incubation period can have serious consequences in hypothesis testing (Goldman, Taylor, and Glibert 1981, p. 137).

This situation can be redescribed in the terms of our confirmation schema. When tested against steady-state laboratory systems, the model linking phytoplankton nutrient uptake rate to growth rate seemed adequate. A problem arose, however, when open ocean systems were found to contain very low nutrient levels, whereas phytoplankton were apparently growing at maximal rates. Investigators later extended the range of experimental systems over which the model was tested. In this case, the time component of the system definition was expanded significantly, to include short-range tests. When tests spanning seconds to minutes were performed, the steady-state model was found not to fit the data. This has led to the suggestion by the biologists that models representing a phytoplankton system must be tested against an incubation period based on the "time scale of physiological responses by phytoplankton" (Goldman,

Taylor, and Glibert 1981, p. 137). Such experiments have led to suggestions of new models for oceanic nutrient dynamics, using these shorter time scales.

4. CONCLUSION

The taxonomy of confirmation presented in this paper included empirical support for models in the form of: (1) fit of the model to data; (2) independent testing of various aspects of the model; and (3) variety of evidence. I presented examples of each type of confirmation, drawing from a range of evolutionary and ecological theories. Instances in which scientists criticize other investigators for lack of sufficient support of a given type are included. I have not attempted to analyze or to justify the various forms of support – that is a project for another paper. Rather, I have attempted to establish the plausibility and importance in evolutionary biology of different categories of empirical support. The greater complexity and variety in my approach, as compared to, for example, a Popperian approach, can facilitate detailed analysis and comparison of empirical claims.

ACKNOWLEDGMENTS