VOLUME 62, NO. 4 *THE QUARTERLY REVIEW OF BIOLOGY* DECEMBER 1987

HYPOTHESIS TESTING IN ECOLOGY: PSYCHOLOGICAL ASPECTS AND THE IMPORTANCE OF THEORY MATURATION

CRAIG LOEHLE*

Savannah River Ecology Laboratory,
Drawer E, Aiken, SC 29801 USA

ABSTRACT

Proper hypothesis testing is the subject of much debate in ecology. According to studies in cognitive psychology, confirmation bias (a tendency to seek confirming evidence) and theory tenacity (persistent belief in a theory in spite of contrary evidence) pervasively influence actual problem solving and hypothesis testing, often interfering with effective testing of alternative hypotheses. On the other hand, these psychological factors play a positive role in the process of theory maturation by helping to protect and nurture a new idea until it is suitable for critical evaluation. As a theory matures it increases in empirical content and its predictions become more distinct. Efficient hypothesis testing is often not possible when theories are in an immature state, as is the case in much of ecology. Problem areas in ecology are examined in light of these considerations, including failure to publish negative results, misuses of mathematical models, confusion resulting from ambiguous terms (such as "diversity" and "niche"), and biases against new ideas.

INTRODUCTION

SCIENTIFIC METHOD, statistical rigor, and hypothesis testing are being emphasized increasingly in ecology. Articles have begun to address the question of proper null, or neutral, models, particularly in community studies (Caswell, 1976a; Connor and Simberloff, 1979a, 1986; Harvey, Colwell, Silvertown, and May, 1983; Simberloff, 1983; Strong, 1980, 1983), although the null models themselves are controversial at this stage (Gilpin and Diamond, 1984; Grant and Abbott, 1980; Harvey et al., 1983). The proper role of theory and mathematics in ecology is also being assessed (Levin, 1981; May, 1981; Pielou, 1981). Advocates of hypothesis testing are calling for increased rigor (Holling, 1978; Hurlbert, 1984; Quinn and Dunham, 1983; Reckhow and Chapra, 1983; Simberloff, 1983; Strong, 1983; Waggoner, 1975). Finally, articles have begun to address the power of statistical tests for estimation of process rates (Nilsson and Nilsson, 1983) and detection of lack of differences or failure to reject the null hypothesis (Toft and Shea, 1983).

All of this activity is a sign of ecology's incipient emergence as a hard science. (Here "hard science" means a science able to establish theories and laws with predictive and/or explanatory power.) At such a juncture in the history of a science it is not uncommon to find debate about methods and philosophical issues, and ecology is no exception. A question at the heart of these issues is the primacy to be given to the classic hypothetico-deductive (H-D) method. The case for the efficacy of the H-D method seems compelling (Lakatos, 1970; Nagel, 1979; Platt, 1964; Popper, 1959, 1963; Strong, 1980), yet in ecology it is assailed by some and ignored by many. This is a puzzling problem in need of resolution. Recent

---

* Present address: Savannah River Laboratory, Bldg. 773-42A, E. I. du Pont de Nemours & Company, Aiken, South Carolina 29808 USA.

results from the psychology of science can help resolve this debate by showing that the H-D method varies in appropriateness according to the state of development of theory. Psychological studies are also useful here because problem solving is a mental activity, subject to psychological biases and shortcomings, rather than an exercise in pure logic. The main psychological principles involved are confirmation bias, theory tenacity, and difficulties with mental operations on multiple simultaneous hypotheses. The philosophical framework will be presented, followed by results from the field of psychology and finally discussion of the proper role of and limitations to hypothesis testing in ecology.

Exceptions to Hypothesis-Testing

First, however, it is necessary to show that many aspects of ecology do not necessarily lend themselves to analysis by the H-D methodology. Recent emphasis on hypothesis testing has made it fashionable to advocate the framing of every research question as a test of a hypothesis that can be falsified. But many research questions are of the "what, when, where" variety, i.e., they constitute observational or descriptive research. For example:

"What is the geographic distribution of radioactivity around a power plant?"

"What is the breeding season of species X?"

"What severity of drought will kill plant Z?"

The results of these studies are maps or dates or quantities. Although sampling designs and statistical tests are usually involved (e.g., ANOVA on levels of drought stress in the third example above), no hypotheses are being tested—data are merely being amassed for descriptive purposes; serious hypotheses, after all, have *explanatory power*: they explain phenomena in terms of history, mechanisms, or evolution. This is the distinction between statistical inference and testing scientific hypotheses. For example, we estimate the contributions of mortality factors A, B, and C for a population. A statistical inference would be that given the means and variances we can state that A is greater than B than C. This is only scientific inference if some theory *predicts* that A should be larger (or smaller, etc.). Trying to formulate observational (descriptive) studies as tests of hypotheses is awkward and misleading. Yet these studies are valuable to science, may require great ingenuity and perseverance, and may be very expensive.

During initial exploration of a phenomenon there may be no theory to test. There may be an empirical curve, such as that produced by light saturation during photosynthesis, for which we have no explanation. It may only be possible to establish the magnitude of the effect or perhaps fit a curve through the data. The process of explaining can provide theories (mechanistic, historical, or evolutionary) that may be testable (Bunge, 1967). It is at this point that the hypothesis-testing debate becomes relevant. In the following discussion, descriptive studies are not under consideration.

HYPOTHETICO-DEDUCTIVE METHOD

Sir Karl Popper, one of the great philosophers of science of the 20th Century, demonstrated the impossibility of proving a scientific theory true by means of induction, because no amount of supportive evidence assures us that contrary evidence will not be found (Popper, 1959, 1963). Inductive logic can only increase our confidence in a theory until data are available that prove the theory to be wholly or partly wrong. His significant principle of demarcation states that theories are scientific only if they are potentially falsifiable.

Falsification is based on the logical asymmetry between proof and disproof. Consider the hypothesis "All numbers are evenly divisible by 2." Many examples ("2, 4, 6, 8 . . .") lend support, but only one negative instance, "3," is required for refutation. This asymmetry of proof is strongest in logical subjects, such as mathematics, and becomes weaker as phenomena become more complex or more influenced by stochastic factors.

Popper realized that, in practice, experimental error renders hypothesis testing inexact. This is why he emphasized testing "risky" predictions. Risky predictions concern phenomena that are highly unlikely to occur by chance, are often novel, and are contrary to competing theories or are not predicted by them. Testing risky predictions is efficient. If the prediction is borne out, the hypothesis is strongly corroborated. If not, it is strongly refuted. The fact that the prediction is highly unlikely by chance means that we are less dependent on small statistical differences.

Strong Inference

In an extension of Popper's work, Platt (1964) proposed the concept of strong inference. Although Platt was not the first to propose the use of multiple hypotheses (see Chamberlin, 1965, reprinted from 1890), he noted that certain fields, such as biochemistry, are characterized by very rapid progress that is the result of a common research methodology. Single experiments in these fields are often designed to distinguish between at least two or even four or more hypotheses. Such an approach is very efficient and yields rapid progress. Strong inference applies the following steps to every problem in science formally, explicitly, and regularly:

(1) Devise alternative hypotheses;

(2) Devise a crucial experiment (or several of them), with alternative possible outcomes, each of which will as nearly as possible exclude one or more of the hypotheses;

(3) Carry out the experiment so as to obtain a clean result;

(4) Recycle the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain.

Using these steps we work our way through a "conditional inductive tree" or "logical tree" in which our experiments, by exclusion, determine our choice at each decision point. This is an efficient sequence of steps and, according to Platt, the fastest method for exploring the unknown so as to reach firm conclusions.

In ecology, Step 3 is not usually easy. The difficulties caused by such factors as randomness, complexity, and unobservable variables are too well known to belabor here. However, experimental difficulties do not provide a complete explanation. In addition to these experimental difficulties, I believe that psychological barriers inhibit devising and testing alternative ecological hypotheses. The vagueness of much theory in ecology makes it difficult to derive explicit hypotheses or predictions. Understanding these factors should lead to faster progress in ecology.

COGNITIVE PSYCHOLOGY AND SCIENTIFIC METHOD

It is not surprising that there is a difference between hypothesis-testing criteria proposed by philosophers and those used by scientists. Actual data are, after all, rarely so clear-cut as to allow formal logical deductions. In addition, humans are not perfect problem-solving machines. The history of science is full of creative but wrong ideas, many of which persisted for unusually long periods. We have had animal magnetism, N-rays, and Lamarckism, for example. Clearly, science is not a mechanical process for cranking out the truth. Which characteristics of the way scientists think and solve problems are consistent with both successes and failures? Studies of problem solving by scientists can help answer this question.

Early work in the psychology of science was largely anecdotal or concerned with overly simplistic laboratory problem-solving tasks (see discussion in Tweney, Doherty, and Mynatt, 1981). Recently, progress has been made using the tools of cognitive psychology, including interview techniques, self-reporting of subjects during problem-solving tasks, and analysis of scientific diaries and notebooks (e.g., Goodfield, 1981; Gruber and Barrett, 1974).

Confirmation Bias

One of the most prominent findings in the psychology of science is a class of phenomena called "confirmation bias," which is a tendency to try to confirm one's theory, or to not seek out or use disconfirming evidence. This can occur even when disconfirming evidence is clearly presented and glaringly contradicts the hypothesis. This type of bias was noted as early as 1890 (Chamberlin, 1965), and has since been experimentally demonstrated (Tweney, Doherty, and Mynatt, 1981). Experimental subjects often have difficulty when instructed to find the example that will prove a rule to be false, particularly when the rule is complex. This is an asymmetry in hypothesis generation.

It also can be difficult to compare two hypotheses. It is easier to compare two sets of data against one theory than two theories against one set of data, according to Tweney, Doherty, and Mynatt (1981). Additionally, unconscious confirmation bias may creep into the scientific process via exploratory data analysis of large data sets and via the tendency to publish only significant or confirming results (Selvin and Stuart, 1966). Caswell (1976b) has pointed out that there is a tendency for modelers to show only how their model fits the data, rather than to subject it to any strong tests. In experiments, scientists have shown far too much confidence in the reliability of small samples (Tversky and Kahneman, 1971). Small sample sizes produce not just spurious positive results that are confusing to everyone, but also a high probability (around 50%) of not detecting an effect that in fact exists (Tversky and Kahneman, 1971). This problem seems particularly acute in the social sciences, but ecology is not exempt. All of this weakens the efficacy of hypothesis testing, even when it *is* attempted.

Theory Tenacity

Theory tenacity has also been found to be a significant factor in problem-solving behavior (Lakatos, 1970; Mitroff, 1974; Tweney, Doherty, and Mynatt, 1981). It differs from confirmation bias in that it reflects a commitment to basic assumptions; confirmation bias can occur in cases with no emotional bias merely because it is easier to think of confirming than disconfirming evidence. Mitroff (1974) found evidence that those scientists in the space program most responsible for major progress and new ideas were also often unwilling to give up their ideas in the face of contradiction—i.e., they had high tenacity. The most well-known researchers in a field are often the ones that most strongly denounce a major new idea. In the extreme, some great scientists have drifted into disrepute by clinging to a personal hypothesis or outdated ideas. Emotional investment and personal involvement in ideas lead to theory tenacity. A classic example in ecology is Raymond Pearl (Kingsland, 1985), who loudly championed the logistic equation as a law of population growth (discussed further below). If more than one theory is extant, passionate debate may ensue, and sometimes the more vigorous debaters prevail until, and sometimes beyond, the point at which decisive experiments demonstrate the errors in their arguments. Sometimes only one theory (called the ruling theory by Chamberlin, 1965) is available; in this instance theory tenacity may be particularly strong because rejecting it will leave nothing. A bad theory is thought to be better than none. It is to Darwin's credit that on several occasions he gave up a hypothesis without another to replace it, though this caused him great anxiety (Gruber and Barrett, 1974).

Theory Maturation

Bunge (1968) and Tweney, Doherty, and Mynatt (1981) have proposed that theory and data quality are key concepts for evaluating hypothesis testing methods. In the early stages of theory development it is often unclear exactly what predictions the theory makes. As Lakatos (1970) points out, early versions of a theory may only refer to ideal situations, and elaboration may be required to derive testable predictions. Insistence on testing predictions too early is dogmatic falsificationism (Lakatos, 1970), which may result in premature rejection of a partially correct theory. Additionally, evidence against an immature theory may not be very convincing because of the theory's inherent vagueness. When different individuals derive contradictory predictions A and B from theory Y, disproof of prediction A does not necessarily induce us to give up Y. Psychological research backs up this philosophical analysis with the finding that premature attempts to falsify hypotheses interfere with problem solving (Gorman and Gorman, 1984).

Theory maturation is necessary before strong inference can be effective. (See Bunge, 1968, for philosophical treatment and examples of the maturation process.) Maturation changes a theory from vague and qualitative to precise and predictive; its empirical content increases (Lakatos, 1970). It is often accompanied by elaboration of experimental, analytical, or mathematical methodologies. During this process a theory will be faced with many contrary or unassimilated facts, as well as with competing theories. Theory tenacity then becomes a positive force (if the theory is a good one), which keeps the idea alive in the face of negative evidence. Darwin, Einstein, and Bohr tenaciously clung to theories they felt to be right even when key elements could not be proven or some data were contradictory.

Confirmation bias (conscious or not) can also be a positive force that results in the accumulation of instances (facts, case histories, experiments) that support the theory. This may allow better evaluation of competing theories or refinement of the details of a given theory. Lack of theory maturation leads to stagnation. This occurs when ruling theories are never subjected to rigorous testing, and results in a large mass of unassimilated exceptions. Ecology has been rife with ruling theories that are not effectively tested for long periods (McIntosh, 1980, 1985). Examples include Huchinson's suggested 1.3 size ratio for sympatric competitors and the dictum that diversity begets stability.

Since it is difficult to perform a critical test of an immature theory, effort in many areas of ecology may be better focused on increasing theory maturity rather than trying to prove or disprove the theory. There are many steps that can be taken to increase the maturity of a theory. One of the first steps involves the attempt to convert a verbal statement of theory into a more precise form. A general, plausible theory such as "animals gather food in an optimal way" can lead to a variety of models that attempt to embody the theory (in this case the plethora of optimal foraging models). Plurality of thought is useful at this stage. Comparing the models can show if their predictions differ, and if so in what ways. In this way, Schoener (1983) showed that several models of optimal feeding-territory size are special cases of a more general one. Comparisons may also be purely mathematical and may result in a general model or derivation of transformations that can convert one model into another (e.g., relations between various sigmoidal growth curves, Grosenbaugh, 1965).

Defining the scope of the theory increases theory maturity by enhancing testability and precision. Examination of domains of application clarifies restrictive assumptions and indicates the universality of predictions. For example, in exploring relations between body size and metabolic rates or life-history parameters, comparison of different taxa (e.g., birds vs. mammals) is useful. Theory scope can also be clarified by enumerating classes of behavior to be predicted or allowed by the theory. For example, by listing the possible ideal types of population behavior around equilibrium (e.g., stable, limit cycle), May (1973) clarified the scope of analytic population theory. Boundary conditions are also important for defining theory scope. They include such factors as environmental variables, spatial relations, and age class structure.

A check on the level of maturity of a theory is provided by the criterion of robustness (Levins, 1966). A theory is robust if two or more lines of argument or types of experiment lead to the same conclusion. If a theory is robust to particular assumptions or data sets, then it is more likely that attempts to test the theory will be successful because test results will be less often compromised by secondary considerations. We may examine robustness by asking: Do different representations (models) of the theory make the same predictions? Do different measures of entities in the theory (e.g., different diversity measures) lead to the same results? Is the mathematical model sensitive to parameter variation, aggregation scheme, or variables left out of the model?

The importance of operational definitions was stressed by Nagel (1979). For example, quantifying diversity (defining it operationally) showed that there are several different types of diversity. The mere attempt to define phenomena operationally can dramatically increase theory maturity. A fundamental step in Galileo's studies, for example, was distinguishing speed, velocity, and acceleration from each other and defining how to measure them. I believe that the lack of operational definitions for system, ecosystem, and community (among others), causes constant problems in ecology.

In order to go from theory to experiment, it is often necessary to make deductions (predictions) from a theory. Formally stating the logical structure of a theory, perhaps in axiomatic or mathematical form (e.g., Lewis, 1982; Reiners, 1986), makes a theory more well formed and makes deductions easier. Even in the case of a well-formed, mathematically stated theory, the consequences or predictions of the theory are not always obvious. Deriving and clarifying predictions from a theory is an important but difficult step. From Newton's initial inverse square relation for gravity it was not at all self-evident that planets should follow elliptical orbits. Elaboration of the mathematical structure and consequences of a theory can be extensive; mathematicians are still working out the mathematics associated with the theory of relativity 60 years after Einstein. There has been a significant elaboration of the mathematics associated with population genetics and population dynamics in ecology. Other aspects of ecology, such as ecosystem studies, have encountered systems that are analytically intractable so that mathematical elaboration is untenable and simulation is the principal useful tool. Cumulative progress is slower when simulation is necessary than when analytical solutions are possible.

Once predictions have been derived from a theory, experimental methods often become crucial to the theory-maturation process. In ecology new techniques have opened up fields previously blocked and allowed theories to be tested. New techniques include radio tracking of animal movements, and the use of radioactive tracers, genetic markers, physiological monitoring equipment, and aquatic microcosms.

Also key to maturation is the development of new statistical techniques. Estimation of population size relies heavily on mark-recapture methodologies developed specifically for ecology. Multivariate methods, such as detrended correspondence analysis and discriminant analysis have allowed community ecology to enter a more quantitative stage of development. A hotly debated issue concerns the development of appropriate null models, particularly for studying community organization and determining the role of competition. Consideration of statistical power and use of proper tests can help prevent incorrect rejection of a valid theory or spurious confirmation of an invalid one (e.g., Garsd, 1984).

Such operations can help bring a theory to a level of maturity at which more rigorous testing is possible. At the same time the utility of the theory will increase because of greater explicitness. When theory quality (maturity) increases, the distinctions between competing hypotheses become clear enough for unambiguous experimental testing. Then strong inference can come into play, as Platt (1964) documents for some fields of science.

APPLYING THE RESULTS TO ECOLOGY

Opposition to H-D Methods in Ecology

In spite of the apparently convincing case made for falsificationism and strong inference, many oppose this program. Roughgarden (1983) complains, in essence, that that is not how we actually do science and that we do not need formal logic. May (1981) objects to philosophers telling scientists how to do science. May cites Darwin's notebooks (Gruber and Barrett, 1974), which describe how he actually arrived at his theories, as evidence against the hypothetico-deductive model, noting that Darwin did not propose explicit hypotheses and test them. The H-D method and strong inference, however, are valid no matter how theories are obtained. Dreams, crystal balls, or scribbled notebooks are all allowed. In fact, induction may be used to create empirical relations which then become candidates for hypothesis testing even though induction cannot be used to prove anything. Popper was most concerned about the public test of a theory: e.g., now that Mr. Darwin has published his treatise on evolution, what are we to make of it? How do we judge if it is wrong? How do we decide if a new version of the theory is better?

Telling evidence on the status of falsifiability and strong inference in ecology is the rarity and weakness of this type of treatment, as pointed out by Strong (1980) and Connor and Simberloff (1986), and the disdain with which such philosophical issues are treated (e.g., MacFadyen, 1975; Bartholomew, 1982). The vigorous and inconclusive debates about the role of competition (Connor and Simberloff, 1986), the literature on model evaluation (Caswell, 1976b; Loehle, 1983) and the vigorous debate on the role and appropriateness of null models (Caswell, 1976a; Connor and Simberloff, 1986; Gilpin and Diamond, 1984; Grant and Abbott, 1980; Harvey et al., 1983) all illustrate that proper hypothesis testing in ecology is an unresolved issue. Historical summaries of theory in ecology by McIntosh (1980, 1985) further illustrate the lack of strong corroboration or disproof of theories and the inertia of established theories in ecology. Theories often muddle along in a plausible but unconfirmed state, often side by side with several alternatives.

A counter-argument is made by Fretwell (1975), that by way of the influence of Robert MacArthur the H-D method gained acceptability in ecology, resulting in a dramatic increase in hypothesis testing in the ecological literature from the 1950s to the 1970s. Although this is probably correct (see Kingsland, 1985), much of the testing, including MacArthur's and Fretwell's work, was, to quote Fretwell, based on "weak data." The tests were often inconclusive or just mildly confirming (McIntosh, 1980, 1985; Pulliam, 1980). Generation of hypotheses (theories) is important but must be followed by theory maturation and finally by strong tests.

This, then, is the paradox: An apparently compelling logical argument and successful application of the strong inference method in other fields is viewed with hostility, ignored, or weakly applied in ecology. Applying the concept of theory maturity and results from cognitive psychology (elaborated above) enables us to resolve the paradox. In particular, we can address behaviors that interfere with efficient hypothesis testing and lead to inconclusive debates.

Bias in Hypothesis Testing

Disconfirming evidence may be suppressed when it conflicts with a ruling theory or encounters journal reviewers exhibiting confirmation bias or theory tenacity. In a study of the review process in psychology, Mahoney (1977) submitted two sets of contrived research papers, differing only in their results. He found that disconfirmatory papers (challenging "accepted" findings) were less likely to be accepted for publication and less likely to be rated as methologically sound than were confirmatory papers. Van Valen and Pitelka (1974) cite instances of this kind in ecology. On occasion it may be very difficult to publish carefully done research in major ecological journals if the author is "merely" refuting someone else's theory without proposing an alternative (e.g., Heck, 1976). The publication of disconfirming evidence is crucial to scientific progress because it allows us to refine or reject theories (Connor and Simberloff, 1979a).

Theory tenacity among the established scientific community can result in strong bias against new theories (Honig, 1982; Horrobin, 1982). The extent of this bias may be such that new theories cannot find an outlet. Lindeman's classic paper on trophic structure (Lindeman, 1942) is an outstanding example of this problem; he could not get it published until Hutchinson backed him. Many theoretical journals have been founded for the purpose of giving new ideas an outlet—e.g., *Quarterly Review of Biology*, *Speculations in Science and Technology*, *Medical Hypotheses*, and *Theoretical Population Biology*, founded by Raymond Pearl, William Honig, David Horrobin, and Robert MacArthur, respectively. Reviewers may also show a tendency toward dogmatic falsificationism (Lakatos, 1970) by demanding that authors support a new theory with conclusive evidence. It is rarely possible to propose and prove a theory all in one blow, however, because a period of theory maturation and data gathering is necessary first, and because few scientists are both theoreticians and experimentalists. It is, of course, valid to object to theories that are not even, in principle, testable.

Publishing Only Positive Results

Confirmation bias is exhibited by the failure to report the non-occurrence of something, since only occurrences are considered data. For example, in studies of tree root systems the occurrence of root grafts is considered an anomaly and is sometimes reported as such. Lack of grafts has generally not been reported because this was considered for many years to be the normal condition. In my own work, my theory (unpubl.) predicts that certain trees should *not* form root grafts, but the data for testing this are the very data *not* consistently reported.

In a more general sense, an experiment that does not show a relationship of some sort may be perceived as a failure. Results of such studies are rarely published. This is documented in psychology by Mahoney and DeMonbreun (1978). Of course, in ecology, confounding factors may make it difficult to establish a relationship, so that a careless study will almost certainly turn up nothing. But the converse, that all studies that turn up nothing must have been careless, is not true. A carefully designed and executed study with sufficient statistical power should be considered for publication even if it shows no treatment effect. This is because a lack of effect may be informative when compared with those studies in which an effect was found.

Bias Owing to Delays

Another factor contributing to theory tenacity and confirmation bias is the rigidity associated with long time-scales. The time between writing the first draft of a new proposal and starting field work may easily be three or four years, with an additional few years before the final data are in hand. Such a long delay between hypothesis formation and data analysis may so increase commitment to the hypothesis that inherent confirmation bias and theory tenacity tendencies become even stronger.

McIntosh (1980, 1985) suggests that in ecology speculations have often been elevated to the status of laws merely by the passage of time, particularly when the author is prominent. This is likely due to the long delays between publishing a theory and effectively testing it. This may be sharply contrasted to biochemistry, for example, where, as Platt (1964) points out, it is sometimes possible to line out alternative hypotheses at the blackboard in the morning and begin work on the problem in the afternoon, with results available in a few days or weeks. While not all problems are so quickly solved, rapid hypothesis testing makes it far less likely for a researcher to become committed to an idea personally or in public before testing it in the laboratory. It also makes it feasible for a single person or team to tackle all of the alternative hypotheses, in contrast to ecology where it is seldom possible for a single researcher to test more than one hypothesis. [Tweney (pers. commun.), however, points out that delays can also have a salutary effect by allowing second thoughts or theory maturation before a grant is obtained or data analyzed.]

Inconclusive Debates

The lack of maturity of much theory in ecology explains much of the inconclusive wrangling that takes place; theory tenacity makes each side in the debate hold to its position, but theory immaturity prevents a conclusive outcome and may even prevent the disputants from discussing the same problem. An example is the debate over the role of competition in organizing communities (Connor and Simberloff, 1979b, 1986; Grant and Abbott, 1980; Roughgarden, 1983; Simberloff, 1983; Strong, 1983). The long debates and confusion about the niche concept are also typical (Hurlbert, 1981). This type of ambiguity is symptomatic of immature theory. A theory must be unambiguous before rigorous tests can be made.

The basic terms (primitives, objects) of a theory are frequently a source of problems in ecology because they are so often ambiguous and not operationally defined. For example, in the debate about determinants of ecosystem stability, it is often not clear what type of stability is being referred to, nor are criteria clear for measuring it (Connell and Sousa, 1983). In the early theories stating that climax communities had the highest diversities, the two primitives in the theory, "diversity" and "climax," had several meanings. Diversity was used to mean total number of species (land area unspecified), landscape diversity, number of species per number of individuals, and information theoretic measures, of which there are several not necessarily commensurate indices. "Vegetation climax" was understood as the upland mesic community type toward which landscapes evolved, as vegetation in the prolonged absence of major disturbance (e.g., fire), or as vegetation in the context of natural disturbances (e.g., natural prairie fires). An unambiguous test of this theory is not possible until these terms are consistently defined.

In addition to being precise, primitives must be operationally defined (Nagel, 1979). For example, the words "system" and "ecosystem" are usually vague and not operationally defined. We can rarely be sure that a particular group of organisms is a "system" or an "ecosystem" because there are no criteria for deciding. Two fir trees 100 miles apart have no measurable interactions yet we may casually refer to them as members of the boreal coniferous "ecosystem." This is a typological classification, not a functional one. Without clear definitions, debates about system properties, dynamics, and organization are guaranteed to be inconclusive.

Lack of Progress

Some fairly strong predictive theories in ecology are now reaching a stage of moderate maturity. These include optimal foraging theory (Pyke, 1984), optimal feeding-territory size (Schoener, 1983), adaptive significance of leaf shape (Givnish and Vermeij, 1976), relationships between body size and physiological functions (Hofman, 1983; Lindstedt and Calder, 1981), and the intermediate disturbance hypothesis explanation of diversity (Connell, 1978). For each of these theories, early statements of the theory were incomplete, vague, without criteria for testing, or nonquantitative, but greater precision was gradually attained. In other instances, however, progress is apparently blocked. The long and bitter debate over density-dependent versus density-independent population regulation is an example. It is enlightening to examine, in terms of psychological factors and theory maturity, areas where progress is blocked.

I will begin by examining the story of Raymond Pearl and the logistic model. (Historical data are from Kingsland, 1985.) In the 1920s, Pearl became one of the first biologists to advocate the use of mathematical models for population studies. He focused on the logistic model and quickly came to view it as a "law" of population growth. He was very energetic, and by applying the logistic equation to a wide array of problems and publishing these studies in many journals, he succeeded (where others before him had failed) in popularizing the logistic model and showing that mathematical models could be useful for population work. At the same time he inspired a heated controversy, largely because he oversold the logistic model and committed some methodological errors. His first error may be described in terms of confirmation bias. To illustrate the usefulness of the logistic equation, he fit it to census data for several countries. In most of them, data were only available for half of the curve or less. He showed a strong bias in interpreting the fit of these data as "proving" the logistic "law," when they were generally inadequate tests. Pearl exhibited extreme theory tenacity, forgetting objections he himself had raised against earlier authors (such as that fit to data does not prove explanation), ignoring many criticisms, and creating many ad hoc adjustments to his "law" when it was challenged by contrary data. Perhaps his greatest error was in not understanding his theory's level of maturity. A newly proposed theory will almost always need major improvements; it is immature. The logistic model was no exception. Raising it to the status of a law meant that any deviations from it were due to particular circumstances requiring ad hoc clauses. It was thus rather immune to criticism. Such an approach is bound to be considered arrogant and arouse opposition. In Pearl's case confirmation bias and theory tenacity were so extreme that they no longer performed their usual beneficial theory-building function but led to rigidity and fruitless controversy.

In recent years, progress has been slow in community ecology because theory maturity and data quality are often low. This is largely due to the great complexity of communities and the difficulty in controlling or manipulating them. Long time-scales also interfere with execution of informative experiments. Certain debates (e.g., the role of competition in structuring communities) are correspondingly full of much heat and little light (e.g., Grant and Abbott, 1980, vs. Connor and Simberloff, 1979b); the experiments are conclusive only to the particular camp conducting them (if then). Theory tenacity tends to make the opposing camps dig their trenches deeper.

This problem is particularly severe in the mathematical treatment of community dynamics and stability. It is necessary to develop mathematical methodologies and explicit predictions, but model assumptions should be related to biology and predictions should be related to data. In many cases, unfortunately, neither of these developments occur. For example, many mathematicians have written papers on N-species Lotka-Volterra systems, generally concerning stability properties. Such simple models can be useful for studying factors in isolation or as ideal cases (e.g., a system consisting solely of two competitors, an ideal case that can actually be studied in the laboratory). But problems arise from failure to be sure that the factors left out do not affect the property under consideration (i.e., stability). On the basis of nonlinear models, it has been postulated that multiple equilibria are likely to occur (see Holling, 1978; Connell and Sousa, 1983). Using linear models, or eliminating competition or refuges from consideration, can change this result completely. Such simplifying assumptions are thus part of the subject being studied rather than something to be glossed over. We can see that this problem is significant by noting common assumptions that affect the property being examined but that are untested or nonbiological: the number of species is even (Roerdink and Weyland, 1981); no switching of predator diets occurs; mutualism and omnivory do not occur; the competition matrix is anti-symmetric (Porati, Porati, and Vecli, 1978); the sum of compartments equals a constant (Hirata, 1980); competition coefficients are constant; or the system is linear and donor controlled. Competition is actually excluded from much of network theory that is based on static food web models.

Such work does not serve to increase the maturity of community theory, because there is almost no attempt to state or test assumptions, relate the models to theory, or derive or test predictions. Consequently progress in quantifying community theory is slow. What is needed is progress in the sense of Lakatos (1970), where there is an increase in empirical content. Such progress would (1) constrain the mathematics, (2) examine the mathematical assumptions, (3) operationalize the variables in the equations (how do we measure it?), and (4) derive predictions for testing. Little of this has taken place. Assumptions are usually vague or unspecified, predictions of the models are not testable, and many of those doing field tests of community theory do not utilize the mathematics.

Disconfirming evidence has been virtually ignored by mathematicians examining N-species persistence. For example, it is often assumed that an N-species equilibrium point exists or conditions are elucidated for it to exist. But field work shows the importance of spatial heterogeneity and disturbance in regulating herbivory and competition and allowing many species to persist in the community (Connell, 1978; Whitham, Williams, and Robinson, 1984). A critical assumption of the models, that interaction coefficients (e.g., competition and predation) between any pair of species are independent of the species composing the remainder of the community, has been shown experimentally to be drastically wrong (Hairston, 1968; Neill, 1974; Wilbur, 1972). Further, the use of constant competition and predation coefficients ignores the very real importance of refuges in preserving species when their populations are low. These and other severe criticisms have been ignored; thus the Lotka-Volterra analyses are often more in the realm of pure mathematics than of ecological theory. Posed as pure mathematics, they are immune to disconfirmation of a scientific kind. They thus easily acquire an aura of absolute truth, when actually no attempt has been made to apply them to real data. Similar comments may apply to other areas of ecology, such as mathematical evolutionary theory (Van Valen and Pitelka, 1974).

Platt's (1964) view can help restore perspective on the role of mathematics. He contends that mathematics is a weak, fine-mesh box that will only hold the phenomenon if it is first caught in a strong, but coarse, logical box. We establish this logical box by the process of strong inference, which establishes the qualitative features of the phenomenon. Returning the mathematics to the service of scientific discovery requires closer attention to assumptions, experiments, and theory in relation to the mathematics in question. We must agree that experiments can refute a mathematical result before we admit that the mathematics embodies theory.

CONCLUSIONS

We can now draw some conclusions about the rate at which theory maturation is likely to occur in a particular field and why certain fields of ecology have made more progress than others. When process time-scales are short, as they are, for example, in plant physiology, progress is likely to be rapid, particularly if experimentation is possible. Introduction of mathematical models is likely to speed up this process because it allows sharper predictions that are more easily tested. Thus, there has been rapid progress in physiological ecology (plant and animal), using methods that more closely resemble strong inference.

When longer time-scales are involved, theory maturation may be rapid if a history of the phenomenon is available. Thus, geology and modeling of tree growth benefit from detailed records of the past (stratigraphy and tree rings, respectively). If time-scales are long, if little history of the system is left, and if the system is complex, then progress is likely to be slow because it will be difficult to define alternative hypotheses clearly so as to test them unambiguously. Confirmation bias and theory tenacity are then likely to interfere with scientific progress. Studies of community organization (including the roles of competition and predation), ecosystem functioning, demographics, and evolution are susceptible to these problems. Such studies rely on non-experimental evidence (Connor and Simberloff, 1986); thus, the proper role of hypothesis testing, the effect of psychological factors, and the impact of theory maturation are particularly relevant.

Long time-scales and noisy complex systems should not, of course, cause despair. Significant progress can be made by following Popper and making risky predictions rather than attempting to use only goodness of fit or rejection of null hypotheses (Connor and Simberloff, 1986).

Thus, cognitive psychology and philosophy can make significant contributions to a field like ecology. The symptoms that indicate trouble in ecology include terminological muddles (e.g., niche theory), vociferous but inconclusive debates (e.g., competition theory), lack of progress, and accepted dogmas that have never been corroborated or seriously challenged (McIntosh, 1980, 1985). Many areas of ecology currently exhibit one or more of these symptoms. Understanding the psychological factors of confirmation bias and theory tenacity can help clarify the reasons for fruitless debates and irrational arguments in ecology. Understanding the concept and process of theory maturation can help prevent dogmatic falsificationism and increase the ease with which ecological theories are made more precise and testable.

ACKNOWLEDGMENTS

I would like to thank anonymous reviewers, Paul Bertsch, Peter T. Haug, Stuart Hurlbert, David Kovacic, Lyndon Lee, Gary Meffe, John Pinder, David Scott, Richard Seigel, Peter Stangel, and particularly Ryan D. Tweney for helpful reviews of manuscript drafts. Sue Novak and Becky Wolf provided editorial assistance and Cynthia Turnipseed and Cathy Houck typed the manuscript. This work was supported by Contract DE-AC09-76SR00819 between the U.S. Department of Energy and the University of Georgia's Savannah River Ecology Laboratory.

REFERENCES

BARTHOLOMEW, G. A. 1982. Scientific innovation and creativity: A zoologist's point of view. *Am. Zool.*, 22:227-235.

BUNGE, M. 1967. *The Search for Truth: Studies in the Foundations, Methodology and Philosophy of Science. Vol. 3 (II): Scientific Research.* M. Bunge (ed.). Springer-Verlag, New York.

——. 1968. The maturation of science. In I. Lakatos and A. Musgrave (eds.), *Problems in the Philosophy of Science* #3, p. 120-147. North-Holland, Amsterdam.

CASWELL, H. 1976a. Community structure: A neutral model analysis. *Ecol. Monogr.* 46:327-354.

——. 1976b. The validation problem. In B. C. Patten (ed.), *Systems Analysis and Simulation in Ecology*, Vol. 4, p. 313-325. Academic Press, New York.

CHAMBERLIN, T. C. 1965. The method of multiple working hypotheses. *Science*, 148:754-759. [Reprinted from 1890.]

CONNELL, J. H. 1978. Diversity in tropical rain forests and coral reefs. *Science*, 199:1302-1310.

CONNELL, J. H., and W. P. SOUSA. 1983. On the evidence needed to judge ecological stability or persistence. *Am. Nat.*, 121:789-824.

CONNOR, E. F., and D. SIMBERLOFF. 1979a. You can't falsify ecological hypotheses without data. *Bull. Ecol. Soc. Am.*, 60:154-155.

——, and ——. 1979b. The assembly of species communities: Chance or competition? *Ecology*, 60:1132-1140.

——, and ——. 1986. Competition, scientific method, and null models in ecology. *Am. Sci.*, 74:155-162.

FRETWELL, S. D. 1975. The impact of Robert MacArthur on ecology. *Annu. Rev. Ecol. Syst.*, 6:1-13.

GARSD, A. 1984. Spurious correlation in ecological modeling. *Ecol. Model.*, 23:191-201.

GILPIN, M. E., and J. M. DIAMOND. 1984. Are species co-occurrences on islands non-random, and are null hypotheses useful in community ecology? In D. R. Strong, Jr., D. Simberloff, L. G. Abele, and A. B. Thistle (eds.), *Ecological Communities: Conceptual Issues and the Evidence*, p. 297-315. Princeton Univ. Press, Princeton.

GIVNISH, T. J., and G. J. VERMEIJ. 1976. Sizes and shapes of liane leaves. *Am. Nat.*, 110:743-778.

GOODFIELD, J. 1981. *Imagined World: A Story of Scientific Discovery.* Harper & Row, New York.

GORMAN, M. E., and M. E. GORMAN. 1984. A comparison of disconfirmatory, confirmatory, and a control strategy on Wason's 2-4-6 task. *Q. J. Exper. Psych.*, 36A:629-648.

GRANT, P. R., and I. ABBOTT. 1980. Interspecific competition, island biogeography and null hypotheses. *Evolution*, 34:332-341.

GROSENBAUGH, L. R. 1965. Generalization and reparameterization of some sigmoid and other nonlinear functions. *Biometrics*, 21:708-714.

GRUBER, H. E., and P. H. BARRETT. 1974. *Darwin On Man: A Psychological Study of Scientific Creativity.* E. P. Dutton, New York.

HAIRSTON, N. G. 1968. The relationship between species diversity and stability: An experimental approach with protozoa and bacteria. *Ecology*, 49:1091-1101.

HARVEY, P. H., R. K. COLWELL, J. W. SILVER-

TOWN, and R. M. MAY. 1983. Null models in ecology. *Annu. Rev. Ecol. Syst.*, 14:189-211.

HECK, K. L., JR. 1976. Some critical considerations of the theory of species packing. *Evol. Theory*, 1:247-258.

HIRATA, H. 1980. A model of hierarchical ecosystems with utility efficiency of mass and its stability. *Int. J. Syst. Sci.*, 11: 487-493.

HOFMAN, M. A. 1983. Energy metabolism, brain size and longevity in mammals. *Q. Rev. Biol.*, 58:495-512.

HOLLING, C. S. 1978. *Adaptive Environmental Assessment and Management.* John Wiley & Sons, New York.

HONIG, W. M. 1982. Peer review in the physical sciences: An editor's view. *Behav. Brain Sci.*, 5:216-217.

HORROBIN, D. F. 1982. Peer review: a philosophically faulty concept which is proving disastrous for science. *Behav. Brain Sci.*, 5:217-218.

HURLBERT, S. H. 1981. A gentle depilation of the niche: Dicean resource sets in resource hyperspace. *Evol. Theory*, 5:177-184.

——. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.*, 54:187-211.

KINGSLAND, S. E. 1985. *Modeling Nature.* Univ. Chicago Press, Chicago.

LAKATOS, I. 1970. Falsification and the methodology of scientific research programs. In I. Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge*, p. 91-195. Cambridge Univ. Press, Cambridge.

LEVIN, S. A. 1981. The role of theoretical ecology in the description and understanding of populations in heterogeneous environments. *Am. Zool.*, 21:865-875.

LEVINS, R. 1966. The strategy of model building in population biology. *Am. Sci.*, 54:421-431.

LEWIS, R. W. 1982. Theories, structure, teaching, and learning. *BioScience*, 32:734-737.

LINDEMAN, R. L. 1942. The trophic-dynamic aspect of ecology. *Ecology*, 23:399-418.

LINDSTEDT, S. L., and W. A. CALDER III. 1981. Body size, physiological time, and longevity of homeothermic animals. *Q. Rev. Biol.*, 56:1-16.

LOEHLE, C. 1983. Evaluation of theories and calculation tools in ecology. *Ecol. Model.*, 19:239-247.

MACFADYEN, A. 1975. Some thoughts on the behavior of ecologists. *J. Anim. Ecol.*, 44:351-363.

MAHONEY, M. J. 1977. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognit. Ther. Res.*, 1:161-175.

MAHONEY, M. J., and B. G. DEMONBREUN. 1978. Psychology of the scientist: An analysis of problem-solving bias. *Cognit. Ther. Res.*, 1:229-238.

MAY, R. M. 1973. *Stability and Complexity in Model Ecosystems.* Princeton Univ. Press, Princeton.

——. 1981. The role of theory in ecology. *Am. Zool.*, 21: 903-910.

MCINTOSH, R. P. 1980. The background and some current problems of theoretical ecology. *Synthese*, 43:195-255.

——. 1985. *The Background of Ecology.* Cambridge Univ. Press, Cambridge.

MITROFF, I. I. 1974. Norms and counter-norms in a select group of Apollo moon scientists: A case study of the ambivalence of scientists. *Am. Soc. Rev.*, 39:579-595.

NAGEL, E. 1979. *Teleology Revisited and Other Essays in the Philosophy and History of Science.* Columbia Univ. Press, New York.

NEILL, W. E. 1974. The community matrix and interdependence of the competition coefficients. *Am. Nat.*, 108:399-408.

NILSSON, S. G., and I. N. NILSSON. 1983. Are estimated species turnover rates on islands largely sampling errors? *Am. Nat.*, 121: 595-597.

PIELOU, E. C. 1981. The usefulness of ecological models: A stocktaking. *Q. Rev. Biol.*, 56:17-31.

PLATT, J. R. 1964. Strong inference. *Science*, 146:347-353.

POPPER, K. R. 1959. *The Logic of Scientific Discovery.* Hutchinson, London.

——. 1963. *Conjectures and Refutations: The Growth of Scientific Knowledge.* Harper & Row, New York.

PORATI, M. I. G., A. PORATI, and A. VECLI. 1978. Analytical conditions for the conservative form of the ecological equations. *Bull. Math. Biol.*, 40:257-264.

PULLIAM, H. R. 1980. On digesting a theory. *Auk*, 97:418-420.

PYKE, G. H. 1984. Optimal foraging theory: A critical review. *Annu. Rev. Ecol. Syst.*, 15:523-575.

QUINN, J. F., and A. E. DUNHAM. 1983. On hypothesis testing in ecology and evolution. *Am. Nat.*, 122:602-617.

RECKHOW, K. H., and S. C. CHAPRA. 1983. Confirmation of water quality models. *Ecol. Model.*, 20:113-133.

REINERS, W. A. 1986. Complementary models for ecosystems. *Am. Nat.*, 127:59-73.

ROERDINK, J., and A. WEYLAND. 1981. A generalized Fokker-Plank equation in the case of the Volterra model. *Bull. Math. Biol.*, 43: 69-79.

ROUGHGARDEN, J. 1983. Competition and theory in community ecology. *Am. Nat.*, 122:583-601.

SCHOENER, T. W. 1983. Simple models of optimal feeding-territory size: A reconciliation. *Am. Nat.*, 121:608-629.

SELVIN, H. C., and A. STUART. 1966. Data-dredging procedures in survey analysis. *Am. Stat.*, 20: 20-23.

SIMBERLOFF, D. 1983. Competition theory, hypothesis testing, and other community ecological buzzwords. *Am. Nat.*, 122:626-635.

STRONG, D. R., JR. 1980. Null hypotheses in ecology. *Synthese*, 43:271-285.

——. 1983. Natural variability and the manifold mechanisms of ecological communities. *Am. Nat.*, 122:636-660.

TOFT, C. A., and P. J. SHEA. 1983. Detecting community-wide patterns: Estimating power strengthens statistical inference. *Am. Nat.*, 122: 618-625.

TVERSKY, A., and D. KAHNEMAN. 1971. Belief in the law of small numbers. *Psychol. Bull.*, 76: 105-110.

TWENEY, R. D., M. E. DOHERTY, and C. R. MYNATT. 1981. *On Scientific Thinking.* Columbia Univ. Press, New York.

VAN VALEN, L., and F. A. PITELKA. 1974. Intellectual censorship in ecology. *Ecology*, 55:925-926.

WAGGONER, P. E. 1975. Idols of the model or bringing home the Bacon. In F. J. Vernberg (ed.), *Physiological Adaptation to the Environment*, p. 547-557. Intext Ed. Publ., New York.

WHITHAM, T. G., A. G. WILLIAMS, and A. M. ROBINSON. 1984. The variation principle: Individual plants as temporal and spatial mosaics of resistance to rapidly evolving pests. In P. W. Price, C. N. Slobodchikoff and W. S. Gaud (eds.), *A New Ecology*, p. 15-51. John Wiley & Sons, New York.

WILBUR, H. M. 1972. Competition, predation, and the structure of the *Ambystoma-Rana sylvatica* community. *Ecology*, 53:3-21.