On Hypothesis Testing in Ecology and Evolution

James F. Quinn; Arthur E. Dunham

ON HYPOTHESIS TESTING IN ECOLOGY AND EVOLUTION

JAMES F. QUINN AND ARTHUR E. DUNHAM

Division of Environmental Studies and Department of Zoology, University of California, Davis,
California 95616; Department of Biology, University of Pennsylvania,
Philadelphia, Pennsylvania 19104

The use of explicit hypothesis testing in ecological and evolutionary studies has been the subject of much recent discussion. In the past several years, a number of workers have been severely criticized for a failure to pose and test alternative explanations for patterns observed in nature (e.g., Abbott et al. 1977; Connor and Simberloff 1978; Grant and Abbott 1980; Simberloff 1978; Strong et al. 1979; Simberloff and Connor 1981) and several philosophical overviews have recently been published (Peters 1976; Levins and Lewontin 1980; Simberloff 1980; Strong 1980). The intellectual basis for this discussion is not new. In the early seventeenth century, Francis Bacon (particularly the *Novum Organum*, 1620) discussed the role of proposing alternative explanations and conducting explicit tests to distinguish between them as the most direct route to scientific understanding (Eiseley 1973). Popper had outlined his formulation of proper scientific method by 1934. In an influential paper, Platt (1964) characterized explicit formal hypothesis testing in science as "strong inference" and argued that it is a hallmark of virtually all scientific progress:

Strong inference consists of applying the following steps to every problem in science, formally, explicitly, and regularly:
  1) Devising alternative hypotheses;
  2) Devising 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) Carrying out the experiment so as to get a clean result;
  1') Recycling the procedure, making subhypotheses or sequential hypotheses to define the possibilities that remain; and so on (p. 347).

. . . For exploring the unknown, there is no faster method; this is the minimum sequence of steps (p. 347).

It seems to me that the method of most rapid progress in . . . complex areas, the most effective way to use our brains, is going to be to set down explicitly at each step just what the question is, and what all of the alternatives are, and then to set up some crucial experiments to try to disprove some (p. 352).

In this formalism, correct explanations cannot be proven deductively, except by eliminating all possible alternatives, but incorrect explanations may be disproven by contradictory experiments or observations (Popper 1959, 1972, 1983a).

Popper (1959, but see 1983a, 1983b) argued explicitly that estimates of probability are unfalsifiable, and thus not subject to scientific test under criteria of disproof. In general, propositions not subject to rejection by contrary observations are denied status as scientific theories in the Popperian model. Useful empirical generalizations with known exceptions, including most biological "laws" seem to fall in this category. Perhaps the best known consequence of this viewpoint is Popper's celebrated denial that Darwin's theory of evolution by natural selection is a valid scientific theory (1972, 1976; see also Peters 1976). Clearly this argument in its simplest form poses major problems for ecologists and evolutionary biologists. Postulated ecological causes or relationships can rarely be strictly disproven, although they may often be shown to be unimportant or improbable. Tests of historical causality and experiments with evolutionary constraints are often impossible in principle to perform.

In practice, the logic of ecological and evolutionary research differs from the Popperian model in being largely inductive. Biologists usually rely on a methodology much akin to statistical hypothesis testing, in which potential causal processes are identified, and their probable contributions are evaluated, weighted, and tentatively generalized to other situations. Generally, the object of investigation is the proportion of observed variation that may be explained through the use of one or more predictors (e.g., food limitation, predators, soil nitrogen, etc.), and the proportion that is to be ascribed to "chance." The predictors of "statistical hypotheses" are chosen in part for reasons of simplicity, measurability, and tractability, and represent deliberately oversimplified caricatures of the assumed underlying processes. "Chance" refers less to true physical randomness than to the contributions of the large number of deterministic effects not included in the model. It is insufficiently appreciated that hypothesis testing in this sense is an inductive or even descriptive procedure, and does not correspond especially closely to the deductive logic of "strong inference." The use of the words "hypothesis testing" for both certainly has helped obscure the differences. The logic of statistical inference has been developed at length by Hacking (1965) and many others, and will not be reviewed here.

From the formal deductive model, Platt (1964) and others derive a prescription for proper and efficient scientific methodology: that potential explanations should be explicitly listed and incorrect ones systematically eliminated, leaving an ever dwindling number of possibilities within which the truth presumably must lie. The simplicity of this prescription and its apparent success in some physical sciences and experimental areas in biology belie the difficulty of its application to complex systems of multiple causality, such as those usually studied in ecology and evolution. Implicit in "strong inference" is an assumption that the competing hypotheses to explain observed phenomena are general, mutually exclusive, and, to some extent, exhaustive. Falsification by observation or experiment occurs only to the extent that a hypothesized cause of an observed phenomenon can be shown not to operate at all, and is informative only to the extent that the disproof may be generalized to other situations. However if many causes contribute to an observed pattern, none will be eliminated from consideration by a properly designed experiment. Generally, no single cause can be shown to account for all

of the observed variation in patterns and processes in natural communities. The objective of investigation in cases of this sort is not to determine the single cause of a pattern, as no such cause exists, but rather to assign relative importances to the contributions of, and interactions between, a number of processes, all known or reasonably suspected of operating to some degree.

A number of epistemological and historical critiques of the usefulness of a strong inference model of scientific method now exist. The limitations of purely deductive logic and the doctrine of disproof of specific hypotheses as a prerequisite to progress have been developed at length by Lakatos (1970, 1974) and others. Some philosophers dispute both the possibility of objective criteria for scientific truth in many kinds of investigation (Feyerabend 1975) and the desirability of complete objectivity on the part of individual scientists (Campbell 1979). It is certainly true that the history and sociology of actual scientific advances often correspond poorly to the process envisioned in the hypothetico-deductive model (Kuhn 1970; Brush 1974). Nevertheless, many ecologists and evolutionists appear to accept "strong inference" as the proper model for investigation of complex biological phenomena.

In the ensuing discussion, we will be less concerned with formal method than with examples of specific applications to ecological research. We accept that formal methodologies, even if imperfect, are often useful in clarifying logical stuctures of arguments and in identifying avenues of thought that might otherwise be missed. However we believe that attempts to force study of highly overlapping mechanisms of ecological and evolutionary change into a rigid hypothetico-deductive mold have the potential to detract from understanding. We see three major classes of problems.

1. Formal hypotheses generally cannot usefully be posed in a way that allows meaningful disproof of a finite number of discrete possibilities. Possible contributing causes are not "hypotheses" of hypothetico-deductive reasoning, because in patterns with multiple causes, it is not possible in principle to perform "critical tests" to distinguish between the "truth" of processes occurring simultaneously.

2. Treating possible contributing causes as distinguishable hypotheses leads to univariate critical tests. However, the behavior of a multivariate process may not be safely inferred from any combination of univariate tests if there are strong interactions among contributing causes.

3. In the hypothetico-deductive formalism, understanding is only increased when a hypothesis is rejected. Thus it may be presumed that acceptance of ecological or evolutionary causality only conveys information when a converse "null hypothesis" of nonexistence of the cause has been rejected (e.g., Strong 1980); however propositions about causality in natural communities rarely have single or simply stated converses. In practice, reliable null hypotheses may often be impossible to construct, as we generally cannot deduce the nature of the expected patterns that would evolve in the absence of any given biological process. In most cases, even if a "null hypothesis" can be posed, it has no probability of being strictly correct, and a sufficiently sensitive critical test will necessarily lead to rejection. Thus testing a "null hypothesis" would appear to have no value in formal deductive logic. In practice, null hypotheses represent

reference points for measurement (an inductive procedure) rather than constructs of deductive logic. In this respect, they are similar to other biological models, such as optimality or Leslie matrix formulations, which also may provide useful approximations to some natural processes, but have no probability of being strictly true. The usefulness of any model in this regard depends upon the reliability of its formulation.

The first point seems self-evident. A moment's reflection will reveal that testable statements about causal relationships in ecology and evolution are virtually never posed in a way that makes "alternative hypotheses" mutually exclusive. This is particularly true when processes may vary over a continuum (e.g., the relative unpalatability of a mimic in Batesian vs. Müllerian mimicry). It may often be useful to classify continuously varying phenomena into a finite number of discrete categories. It does not seem reasonable, in a quest for methodological purity, to require the use of predefined categories in order to permit falsification of discrete statements. Attempts to force observations into categories erected for other purposes or circumstances can easily lead to overlooking unique characteristics of a particular situation. As an illustration of this problem, consider the proposition that competitive processes structure natural communities. The hypothesis that competition is the exclusive determinant of species' ranges and abundances can of course be rejected a priori, as can the "null hypothesis" that competition has exactly no effect. In practice, we hope to measure the relative impacts of competition and perhaps predation or soil structure, but it is not at all clear what kind of discrete hypothesis such an endeavor could reject. (We may, of course, conclude that the contribution of competition is too small to measure, but that is a matter of probable strength, not existence, of competition as a possible contributor to the observed pattern.)

The second point is closely related to the first. Effects of simultaneous processes often do not combine additively. Thus attempts to perform a series of univariate "tests" of individual causal factors may misestimate their actual contributions. Yet under a formal hypothetico-deductive scheme, interactions between mutually exclusive causes are not possible. Here again, we argue that appropriate biological methodology is often more analogous to standard statistical hypothesis testing than to "strong inference." Testing for nonadditive interactions is of course a standard part of multivariate experimental design and statistical analysis.

The third class of problems is related specifically to the use of "null hypotheses" in the study of natural communities. A number of workers have proposed the routine application of noninteractive "random" models as alternatives to models involving interspecific interactions (Caswell 1976; Connor and Simberloff 1978; Lawlor 1980; Strong et al. 1979; Simberloff 1978). Plausible "random" models involve a variety of biological assumptions about the nature of the species involved, their vagility, colonization processes, population growth, the carrying capacity of the environment for individuals and species, and related phenomena (Grant and Abbott 1980; Colwell and Winkler 1983). As with any other contributing causes, it is infinitely unlikely that these processes explain all of the observed variation in nature, and thus the "null hypotheses" merely represent some of the many feasible determinants of community structure whose relative contributions

might be assessed. The value of "null models" is certainly as a construct from which departures may be measured to estimate the impact of processses (e.g., competition) not embodied in the model. The reliability of such estimates, however, depends upon being able to state the model explicitly and estimate its parameters more accurately than those of the process being evaluated. Null hypotheses in ecology are often unsatisfactory because they are virtually impossible to specify completely, or require knowledge unavailable directly and difficult to estimate independently of the pattern being studied.

We will develop these themes further using examples from published studies revolving around a theme of hypothesis testing. Our object is not to criticize the particular findings of the studies chosen, which as a whole represent laudable attempts to systematize and organize an often diffuse and nonrigorous literature, but rather to illustrate that rigid adherence to a logical formalism of testing and rejecting (or failing to reject) supposedly alternative explanations can easily lead to rigorously fallacious conclusions.

NON-ALTERNATIVE HYPOTHESES—MECHANISMS OF PLANT SUCCESSION

Plant succession was originally viewed as a process in which earlier colonists change the physical conditions in a newly available or disturbed habitat, thereby rendering it suitable for later colonists who could not have survived the earlier conditions. In the process these later colonists change conditions so that the early ones cannot persist. In its extreme form, succession was viewed as analogous to ontogeny, with earlier stages (species) being necessary to pave the way for later stages (Clements 1916, 1928, 1936). From the beginning this view was challenged by Gleason (1917, 1926, 1927) and others, who felt that the paucity of late successional species in recently disturbed sites represented slow dispersal and growth, rather than any general need to have the habitat modified by earlier colonists. (See Drury and Nisbet [1973] for a review.)

More recently, Connell and Slatyer (1977) proposed that successional patterns may be characterized as belonging to one of three alternative types, which they term "facilitation," "tolerance," and "inhibition." Facilitation represents the classical view of succession in which early colonists are required for, or increase the rate of establishment of later species. Inhibition occurs when early species retard the arrival or establishment of later species. Tolerance is more akin to a Gleasonian view of succession, in which all species can invade immediately following a disturbance. Late successional and climax species come to predominate because they persist longer than early species, and slowly replace them as they die or are removed by local disturbance.

The three successional models are presented as alternative in the sense that properly designed experiments can in principle distinguish which one actually describes any particular situation (Connell and Slatyer 1977). A sufficient test then is whether late successional species become established more rapidly, less rapidly, or equally rapidly, relative to a control, in a plot in which an early species has been thinned or removed (Sousa 1979). Clearly if this experiment were performed in an undisturbed community consisting of one successional and one climax

species, the possible outcomes are that removal of the successional species will slow (facilitation), accelerate (inhibition), or possibly have exactly no effect (tolerance) on the later species. In this case, the three models are mutually exclusive and exhaustive. Tolerance, however, appears to be a null hypothesis, of measure zero, to the other two, unless some arbitrary set of weak interactions is assumed equivalent to no interaction, as in applied inferential statistics.

In almost any more complex case, experimental results that do not allow one to distinguish between the "alternative" processes are possible, or even probable. Performing the test with two species requires knowledge of which species is "climax" and which "successional." If both species persist over long periods of time, one species accelerates establishment of the second, and the second inhibits establishment of the first, facilitation and inhibition could be said to occur simultaneously. Alternatively, at low densities, one species might facilitate the establishment of the second (e.g., by fixing atmospheric nitrogen), but at high densities inhibit it (perhaps by shading). The problems become further magnified when more than two species are considered. In multispecies successions it is generally impossible to rank species exactly in competitive ability or order of appearance (Quinn 1979, 1982, in prep.); thus ambiguities must be present in the model and experimental procedure. Cases of intransitive competitive relationships (A displaces B, B displaces C, but C displaces A; Buss and Jackson 1979) seem particularly difficult to incorporate into Connell and Slatyer's formalism. Problems of density-dependent interactions are magnified by the possibility of interaction (e.g., A and B individually facilitate C, but a mixed stand of A and B inhibits the establishment of C). Finally, it seems likely in any complex succession, that some early colonists (e.g., nitrogen fixers) will facilitate later invaders, whereas other early colonists, if established, will slow the arrival of later invaders. In this case, assignment of the appropriate "alternative" hypothesis will depend upon the successional species chosen for experimentation.

The inability of three simple characterizations to encompass the full diversity of successional processes should be neither surprising nor disturbing, and in some circumstances these simple models may be quite useful (Connell and Slatyer 1977; Usher 1979; Quinn 1979; Sousa 1979). Models of succession, however, illustrate the general problem that nontrivial simple hypotheses about complex systems are never exactly correct. In general, as detailed knowledge of natural history increases, cases that do not fit any of the hypotheses directly, and contain major elements of several, are sure to arise. In such cases, understanding may be better served by direct communication of results of observations and experiments designed to measure underlying processes, rather than by forcing observations into the form of a set of artificially distinct "hypotheses."

INTERACTION BETWEEN HYPOTHESES—DISTRIBUTIONS OF INTERTIDAL ORGANISMS

Ecology textbooks are replete with studies of the causes of distributional limits of species (Krebs 1978; Whittaker 1975). Some of the classical studies come from intertidal communities, in which abrupt limits in vertical distribution of species are often observed (Lewis 1964; Stephenson and Stephenson 1972; Ricketts and

Calvin 1968; Carefoot 1977). Connell (1961*a*, 1961*b*, 1970) demonstrated that the distributions of barnacles on rocky shores could be experimentally modified by removing competitors or predators. His suggestion that upper limits to distributions are limited by physical tolerances (e.g., desiccation, freezing, or heat) and lower limits by biological interactions (e.g., predation and competition) is now well established. For any given species, this suggests a series of critical experiments to distinguish between major hypotheses for the causes of distributional limits. Frequently considered hypotheses include: (1) The limits represent physiological limits for survival. (2) The species does not recruit into the area. (3) Limits are set by the action of a competitor. (4) Limits are set by predation.

The hypothesis that a species is not found outside its adult range because of failure to disperse or settle into the area may be rejected by showing that recruits settle into the area (Connell 1961*b*). This could be done by providing free substrate, e.g., a cleared caged plot or settling plate, and by finding natural recruits. Physical limitation could be eliminated by showing that individuals survive after being transplanted outside their normal range and given appropriate protection from other species (perhaps by caging). Limitation by a "biological enemy," such as a competitor or predator, would be indicated if removal of the second species led to an increase in range, and rejected if it led to a decrease or did not change.

Clearly the hypotheses posed above are not of necessity mutually exclusive. It would be quite possible to have distributional limits affected simultaneously by increasing physical stress, competition, and predation. A more subtle problem emerges when actual experimental results are examined. In most studies, physical stress does not appear to determine the lower limits of intertidal distributions (reviewed by Connell 1972; Carefoot 1977). Yet experiments frequently produce a pattern that would appear to lead to rejection of a hypothesis of limitation by a second species. When a generalized predator is removed, distributions of many prey species typically contract as they are displaced by one or a few dominant competitors. This has been termed a "keystone predator" effect (Paine 1969), and has been repeatedly observed in the field and produced in controlled experiments (Paine 1966, 1971; Harper 1969; Paine and Vadas 1969; Connell 1971, 1978; Dayton 1971; Lubchenco 1978; among others). Even though these observations would appear to reject biological limitation hypotheses, the experimental outcome of the "critical experiment" on predators clearly results from an interaction of predation and competition, and produces an effect opposite to that predicted from the action of either alone (e.g., removal of a "biological enemy" results in contraction, rather than expansion of range).

Analogous results have been found for interactions between physical stress and competition (Dayton 1971; Levin and Paine 1974; Connell 1978; Sousa 1979; Paine and Levin 1981). It seems a common characteristic of marine benthic communities that physical factors causing chronic mortality actually contribute to the persistence of many species by simultaneously removing their principal competitors. Paine (1979) has found that three-way interactions between competition, physical disturbance, and dispersal ability are needed to explain the persistence of a common seaweed. Similarly, physical disturbance in the form of wave shear often appears to decrease predation rates, diminishing the keystone predator effect and increasing the probability of competitive exclusion (Quinn 1979).

In all of these cases, much of the convincing evidence for the roles of, and interactions between, causative factors comes from controlled experiments. Although frequently undertaken with the intent of falsifying the hypothesis that a predator causes prey distributions to be as they are, the principal value of a predator removal experiment seems to be as a relatively direct measure of the effect of the predator on prey distributions. Experiments to clarify ecological causality would seem to be more appropriately directed toward measuring the impact of, and interaction with, potential influences, rather than somehow eliminating all but one by experiment.

"NULL" HYPOTHESES

In order to safely estimate the impact of a particular causative factor, it is necessary to examine two situations in which the strength of the putative cause differs. This is frequently accomplished by comparing situations in which the factor operates with ones in which it is absent. In experiments, controls allow this comparison, but in systems not amenable to experimental manipulation, appropriately constructed null models are used to describe the system in the absence of the action of the postulated causal process.

Perhaps the most troublesome applications of null model techniques in ecology involve attempts to study the importance of interspecific interactions, particularly competition, in determining species' distributions and abundances. Effects of interspecific interaction are notoriously difficult to measure directly without recourse to experimental techniques, which in many cases are infeasible. Natural communities without interspecific interaction do not exist, and thus may not be called upon as reference points for comparison. "Null models" are therefore used to mimic the behavior of hypothetical noninteractive communities.

Null models are in no sense uniquely defined for any natural setting. There seems little possibility species numbers and distributions can generally be deduced from first principles (but see Caswell 1976). More biological reality may be incorporated in the form of species to be considered, physical and spatial limitations on species movement, survival, population growth, and interspecific differences in mortality, resource use, and habitat preference. The choice of such factors considered may drastically alter the predicted nature of the "null" community. The precision and accuracy with which comparisons can be made depends on the reliability of estimates of biological parameters of the "null" models. In many models, estimates cannot be made independently of the actual distributions to which predictions of the model will be compared, and statistical inferences about the predictive power of "null" models may be extremely problematical.

As an example of proposed applications of null models, we will examine studies of the distribution of animals on islands, but many of the difficulties we will discuss apply more generally to the problems of estimating the behavior of any null hypothesis of no interaction in nature.

One approach to studying community dynamics has been to take advantage of "natural experiments," or partially isolated and replicated communities, such as biotas of islands in archipelagoes. Observed patterns are compared with those

predicted by particular models or theories. This technique has been used extensively to explore the role of interspecific competition in structuring island vertebrate communities (Diamond 1970, 1978; Grant 1966, 1968, 1969; Schoener 1974, 1975). Agreement between prediction and observation has been taken as support for the applicability of theories of exploitative competition and niche differentiation. This approach has been criticized severely for failure to first test the "null hypothesis" of a noncompetitive community randomly assembled from available colonists to the islands. The critics insist that no interaction may be inferred from distributional data unless the noninteractive case can be rejected (Connor and Simberloff 1979; Strong et al. 1979).

In almost all cases, the hypothesis that island biotas represent random subsamples of the potential colonists that arrived on the islands in the past is posed as an alternative to an important role for competition. (It is rarely specified whether the arrivals of interest are individuals or species.) Unfortunately the actual arrivals are not known in any cases of interest and must be inferred from present distributional data of some kind. In the case of island biotas, reliable estimation of the expected distribution of colonists in the absence of interspecific interaction on the islands requires a number of biological assumptions that seem no more compelling than that of competition.

1. The species pool sampled must be taxonomically appropriate. Grant and Abbott (1980) have observed that the power to detect interspecific interaction declines as more distantly related taxa, presumably less likely to have a major impact on one another, are included in the species pool. Support for this contention comes from Connor and Simberloff's (1979) analysis of the occurrence of pairs and trios of bird or bat species in the West Indies. Exclusive pairs or trios of species within the same family occur 37%–56% more frequently than predicted by the "null hypothesis," whereas exclusive pairs and trios chosen without regard to family never show a deviation of more than 7%. In all of the West Indies comparisons presented, exclusive groupings appear more frequently than predicted by Connor and Simberloff's particular model of "chance." This is consistent with the qualitative predictions of competitive structuring, but the apparent strength of the competitive effect would appear to be stronger within a family. Processes other than competition may, of course, also be consistent with these observations (Simberloff and Connor 1981).

2. The geographic source of potential colonists must be specified to estimate the source pool. How this is done will depend on the particular model. In cases where colonists are assumed to come from a mainland source (MacArthur and Wilson 1967; Simberloff 1974), the number of species, and their relative abundances and probabilities of reaching an island will all depend upon the extent of the presumed source area. Larger areas will place more species in the estimated colonist pool, but more distant individuals or species will have lower expected arrival rates. In actuality, the potential source areas are likely to vary from species to species, not necessarily independently of the competitive processes to be tested against the "non-competitive" colonization model. The source estimation is further confounded if colonization occurs between islands, as the source may not be estimated independently of the distribution of species used in the test.

3. The probability of arrival and establishment must be known for each species.

The implicit assumption in many simple models (e.g., MacArthur and Wilson 1967) that each species in the pool is equally likely to invade is certainly not strictly true, and for many purposes does not provide even a useful approximation. However, the actual probability of invasion and persistence, even in the absence of interspecific interaction, is a function of the probability, or rate, of arrival, the probability of increase in numbers once present, and the probability of extinction following establishment. These probabilities will vary with abundance and distribution within the source area, vagility, distance from the source, birth and death rates under the particular (presumably variable) physical conditions and resource levels on each island, and in the case of many active dispersers, such as birds, individual choice. Under most conditions, few of these parameters are likely to be known much more reliably than the intensity of competition. A methodology which requires their use to establish the reality of interspecific interaction seems destined to failure.

One way of dealing with some of these estimation problems has been to estimate the probabilities of arrival and establishment of species from the proportion of islands in any size class occupied by a particular species (the "incidence function" of Diamond [1975]; Connor and Simberloff 1979; Simberloff and Connor 1981). This technique, however, yields predictions of species distributions derived from the distributions used to test the predictions, so the test is in no way independent. In particular, post-colonization competitive exclusion will be incorporated into the estimate of noninteractive colonization rates, artificially improving the fit to the noninteractive model (Colwell and Winkler 1983; Diamond and Gilpin 1982).

4. Parameters of the noninteractive model must not be estimated from an interactive biota. In some cases, such as the examination of bill-size ratios in insular birds, the species interaction hypothesis is that the patterns observed are those allowed by the predicted competitive interactions, e.g., some limiting similarity and even spacing of sizes (Schoener 1965, 1974; Grant 1968; Abbott et al. 1977). The corresponding null hypothesis is that the distribution of the character is that which would arise in the absence of interaction. There seems no a priori basis for choosing such a distribution. Clearly not all bill sizes are equally probable, and those observed may be the result of a variety of unknown historical events, including competition. One estimation procedure that has been attempted has been to choose species randomly from the species list of a presumed source area and to use the character distributions from those samples as the "null" distribution (e.g., Strong et al. 1979; Schoener 1983). However the degree to which these samples mimic a noninteractive pattern surely depends upon an assumption that the source pool is essentially noninteractive. Subsampling from a fauna highly structured by competition no doubt yields samples showing considerable competitive structure. For example, if there were a "limiting similarity" principle operating in the source pool, there could be no species more similar than the limiting amount in even the most noncompetitive derived fauna (see also Colwell and Winkler 1983). Yet it is clearly unreasonable to require that the role of competition in a complex source fauna be understood in order to permit study of competition in a simple island fauna.

The bottom line is that, no matter how heuristically desirable it may seem,

measuring the impact of biological interaction against the reference point of a noninteractive null hypothesis is often not a realistically achievable goal. The characteristics of a noninteractive biota are not known from first principles, and cannot be empirically measured. Estimation of its characteristics depends upon knowledge of other biological parameters, such as distribution and abundance patterns in the source area, dispersal distances, detailed colonization processes, etc., that are no better established than the consequences of interspecific interaction. In practice, the purported null hypotheses are better viewed as dispersal models, alternative causes in the non-mutually-exclusive sense discussed above. Viewed this way, "random" or "null" models have no "logical primacy" over other possible causative factors in the sense claimed by Strong (1980). On the other hand, we agree with Simberloff and Connor (1981) that the ease of constructing noncompetitive models which predict patterns similar to those ostensibly resulting from competition makes the measurement of competitive effects from distributional data unreliable, and a conclusion of strong competitive effects uncompelling in most cases. When feasible, direct experimentation is desirable (see also Connell 1975; Dunham 1980; Grant 1972; Hairston 1981; Paine 1966, 1971). We find Simberloff and Connor's claim that a random colonization model is more "parsimonious" than a competition model to be distinctly a matter of taste.

DISCUSSION

Throughout this paper, we have argued that a strict application of a formal "strong inference" methodology to elucidating potential causes of patterns in nature is frequently infeasible. Putative causes generally cannot be stated in a way that they are either mutually exclusive or potentially global in their application. Critical experiments to distinguish between the truth of "alternative" causes cannot be performed in principle, and the criteria for, and logical import of, falsification of a potential cause is unclear. We believe that rigid insistence on the "hypothesis testing" formalism has the potential to distract from understanding in several ways. Treating useful generalizations, such as models of succession, as well-defined alternatives denies the possible richness of a continuous range of possible outcomes and suggests inappropriate experiments to distinguish between single points in that range, each infinitely improbable as a description of the actual truth. Non-alternative causes may interact and influence patterns observed in a way that will not be detected by the kind of univariate critical experiment that would be used to attempt to reject proper Popperian mutually exclusive hypotheses. As illustrated by the discussion of a simple interaction in the intertidal zone, such experiments can even lead to erroneous rejection of an important causal process.

Processes contributing to pattern in natural communities do not often lend themselves to easy statement as hypotheses of hypothetico-deductive formalism, since relative contribution and possible interaction are the objects of investigation, not truth or falsity of the process. Ostensible "critical tests" often have value in measuring these contributions. "Rejection" sets the statistical limits of detection as probable upper bounds to the estimate of the process's relative

impact. Thus the role of the null hypothesis is as a reference point for measurement of unknown departures, in the spirit of statistical hypothesis testing, rather than as an alternative hypothesis with some probability of being strictly true, in the sense of "strong inference." "Null" models are not the only possible reference points. For example, optimality models may serve a similar role, also with no probability of representing strict truth.

In other cases, logical reference points may not exist. For example, in a study of temperature effects, a null temperature analogous to a null model of no competition is difficult to define. Most investigators would proceed, if possible, to manipulate temperature experimentally, and measure its effect by regression. By the same logic, there seems no logical imperative that a null model be considered (or as we have argued, even be usefully definable) in an ecological or evolutionary investigation.

We do not intend, in any of our discussion, to downplay either the value of considering multiple approaches to biological problems, or the dangers of attempting to demonstrate preconceived causality without adequate consideration of other potentially contributing processes (Chamberlin 1897; but see Campbell 1979). We believe, as do most of the outspoken advocates of formal hypothesis testing, that controlled experimentation, when feasible, is a powerful technique for removing uncertainty from our understanding of the natural world. We suggest, however, that ecology and evolution are not blessed with clearcut criteria for acceptance of theories, much less methodological prescriptions or requirements for successful science. We view the changes of understanding in these fields as perhaps more akin to Kuhn's model of the establishment of paradigms in science than to Platt's model of strong inference, although we see no reason to distinguish between large scale changes in viewpoint (if there have been any since Darwin) and more modest theoretical advances. 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. Consideration of alternatives and careful experimentation obviously contribute to this process. Formal method is a guide to innovation, however, not a requirement, and healthy skepticism toward a single methodological model seems thoroughly as appropriate as toward any other claim of scientific truth.

SUMMARY

Theories of causality in ecology and evolution rarely lend themselves to analysis by the formal method of "hypothesis testing" envisioned by champions of a "strong inference" model of scientific method. The objective of biological research typically is to assess the relative contributions of a number of potential causal agents operating simultaneously. Sensibly stated hypotheses in the methodology of most field investigations are similar to hypotheses of applied statistics. They are not intended to be mutually exclusive, in any sense exhaustive, or global in their application. It is not possible in principle to perform a

"critical test" or experiment to distinguish between the truth of "alternative hypotheses" if the proposed causal processes they caricature occur simultaneously.

We consider several examples in which a rigid hypothetico-deductive methodolcgy applied to nonalternative ecological "hypotheses" could lead to fallacious conclusions. It has been proposed that processes of ecological succession may be separated into alternative modes of "facilitation," "inhibition," and "tolerance." Yet attempts to experimentally reject one or more of the supposedly distinct hypotheses cannot, in principle, distinguish between them in a variety of biologically interesting cases. In studies of the limits of distributions of intertidal organisms, reasonable univariate experimental tests of possible causes would lead to rejection of "biological enemy" hypotheses when a "keystone predator effect" occurs because the interaction between competition and predation reverses the direction of the effect on some prey populations expected from either process in isolation.

Particular problems arise when "null models" in ecology are treated as hypotheses of "strong inference." Models of ecological or evolutionary causality rarely have single or easily stated "null" converses. Tractable null models have no probability of being strictly true, and thus may be rejected a priori as hypothetico-deductive constructs. In practice, their role is as a reference point for measurement of departures. Their usefulness in this regard depends upon the reliability with which the characteristics of biology without interaction can be estimated. Applied to studies of interspecific competition through the use of species distributions, purported null hypotheses make different biological assumptions than those of the interactive models. They seem neither especially more reliable nor in any way more fundamental. We see no reason to accept the recent claims that "null hypotheses," as applied in ecology and evolution, have any logical primacy or greater parsimony than other approaches to partitioning the variation observed in natural communities among the contributions of many observable causes.

Careful consideration of possible explanations and controlled experimentation contribute a great deal to ecological and evolutionary knowledge. However, we believe that the hypothetico-deductive model of scientific method can provide misleading prescriptions for efficient investigation and acceptance of evidence in phenomena with multiple causes, and should be applied with appropriate skepticism.

LITERATURE CITED

Abbott, I., L. K. Abbott, and P. R. Grant. 1977. Comparative ecology of Galapagos ground finches: evaluation of the importance of floristic diversity and interspecific competition. Ecol. Monogr. 47:151–184.

Bacon, F. 1620. Novum organum. Part 2. Instauratio magna. J. Billium, London.

Brush, S. 1974. Should the history of science be X-rated? Science 183:1164–1172.

Buss, L. W., and J. B. C. Jackson. 1979. Competitive networks: nontransitive competitive relationships in cryptic coral reef environments. Am. Nat. 113:223–234.

Campbell, D. T. 1979. A tribal model of the social system vehicle carrying scientific knowledge. Knowledge: Creation, Diffusion, Utilization 1:181–201.

Carefoot, T. 1977. Pacific seashores. University of Washington Press, Seattle.

Caswell, H. 1976. Community structure: a neutral model analysis. Ecol. Monogr. 46:327–354.

Chamberlin, T. C. 1897. The method of multiple working hypotheses. J. Geol. 5:837–848.

Clements, F. E. 1916. Plant succession. Carnegie Inst. Wash. Publ. 242.

———. 1928. Plant succession and indicators. H. W. Wilson, New York.

———. 1936. Nature and structure of the climax. J. Ecol. 24:252–284.

Colwell, R. K., and D. W. Winkler. 1983. A null model for null models in biogeography. In D. S. Strong, Jr., D. S. Simberloff, L. G. Abele, and A. B. Thistle, eds. Ecological communities: conceptual issues and the evidence. Princeton University Press, Princeton, N.J. (in press).

Connell, J. H. 1961a. Effects of competition, predation by Thais lapillus and other factors on populations of the barnacle Balanus balanoides. Ecol. Monogr. 31:61–104.

———. 1961b. The influence of interspecific competition and other factors on the distribution of the barnacle Chthamalus stellatus. Ecology 42:710–723.

———. 1970. A predator-prey system in the marine intertidal region. I. Balanus glandula and several predatory species of Thais. Ecol. Monogr. 40:49–78.

———. 1971. On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. Pages 298–312 in P. J. den Boer and G. Gradwell, eds. Dynamics of numbers in populations. Proceedings of the Advanced Study Institute on dynamics of numbers in populations, Osterbeek, 1970. Centre for Agricultural Publishing and Documentation, Wageningen.

———. 1972. Community interactions on marine rocky intertidal shores. Annu. Rev. Ecol. Syst. 3:169–192.

———. 1975. Some mechanisms producing structure in natural communities: a model and evidence from field experiments. Pages 460–490 in M. Cody and J. Diamond, eds. Ecology and evolution of communities. Harvard University Press, Cambridge, Mass.

———. 1978. Diversity in tropical rain forests and coral reefs. Science 199:1302–1310.

Connell, J. H., and R. O. Slatyer. 1977. Mechanisms of succession in natural communities and their role in community stability and organization. Am. Nat. 111:1119–1144.

Connor, E. F., and D. Simberloff. 1978. Species number and compositional similarity of the Galapagos flora and avifauna. Ecol. Monogr. 48:219–248.

———. 1979. Assembly of species communities: chance or competition? Ecology 60:1132–1140.

Dayton, P. K. 1971. Competition, disturbance, and community organization: the provision and subsequent utilization of space in a rocky intertidal community. Ecol. Monogr. 41:351–389.

Diamond, J. M. 1970. Ecological consequences of island colonization by southwest Pacific birds 1: types of niche shifts. Proc. Natl. Acad. Sci. USA 67:529–536.

———. 1975. Assembly of species communities. Pages 342–444 in M. L. Cody and J. M. Diamond, eds. Ecology and evolution of communities. Harvard University Press, Cambridge, Mass.

———. 1978. Niche shifts and the rediscovery of interspecific competition. Am. Sci. 66:322–331.

Diamond, J. M., and M. Gilpin. 1982. Examination of the "null" model of Connor and Simberloff for species co-occurrences on islands. Oecologia 52:64–74.

Drury, W. H., and I. C. T. Nisbet. 1973. Succession. J. Arnold Arbor. 54:331–368.

Dunham, A. E. 1980. An experimental study of interspecific competition between the iguanid lizards Sceloporus merriami and Urosaurus ornatus. Ecol. Monogr. 50:309–330.

Eiseley, L. 1973. The man who saw through time. Scribner's, New York.

Feyerabend, P. K. 1975. Against method. Humanities Press, London.

Gleason, H. A. 1917. The structure and development of the plant association. Bull. Torrey Bot. Club 44:463–481.

———. 1926. The individualistic concept of the plant association. Bull. Torrey Bot. Club 53:7–26.

———. 1927. Further views on the succession concept. Ecology 8:299–326.

Grant, P. R. 1966. Ecological incompatibility of bird species on islands. Am. Nat. 100:451–462.
———. 1968. Bill size, body size, and the ecological adaptations of bird species to competitive situations on islands. Syst. Zool. 17:319–333.
———. 1969. Colonization of islands by ecologically dissimilar species of birds. Can. J. Zool. 47:41–43.
———. 1972. Convergent and divergent character displacement. J. Linn. Soc. 4:39–68.
Grant, P. R., and I. Abbott. 1980. Interspecific competition, island biogeography, and null hypotheses. Evolution 34:332–343.
Hacking, I. 1965. The logic of statistical inference. Cambridge University Press, Cambridge.
Hairston, N. J. 1981. An experimental test of a guild: salamander competition. Ecology 62:65–72.
Harper, J. L. 1969. The role of predation in vegetational diversity. Brookhaven Symp. Biol. 22:48–61.
Krebs, C. J. 1978. Ecology: the experimental analysis of distribution and abundance. 2d ed. Harper & Row, New York.
Kuhn, T. S. 1970. The structure of scientific revolutions. University of Chicago Press, Chicago.
Lakatos, I. 1970. Falsification and the methodology of research programmes. Pages 91–195 in I. Lakatos and A. Musgrave, eds. Criticism and the growth of knowledge. Cambridge University Press, Cambridge.
———. 1974. Popper on demarkation and induction. Pages 241–273 in P. Schilpp, ed. The philosophy of Karl Popper. Open Court, LaSalle, Ill.
Lawlor, L. R. 1980. Structure and stability in natural and randomly constructed competitive communities. Am. Nat. 116:394–408.
Levin, S. A., and R. T. Paine. 1974. Disturbance, patch formation and community structure. Proc. Natl. Acad. Sci. USA 71:2744–2747.
Levins, R., and R. Lewontin. 1980. Dialectics and reductionism in ecology. Synthèse 43:47–78.
Lewis, J. R. 1964. The ecology of rocky shores. English University Press, London.
Lubchenco, J. 1978. Plant species diversity in a marine intertidal community: importance of herbivore food preference and algal competitive abilities. Am. Nat. 112:23–39.
MacArthur R. H., and E. O. Wilson. 1967. The theory of island biogeography. Princeton University Press, Princeton, N.J.
Paine, R. T. 1966. Food web complexity and species diversity. Am. Nat. 100:65–75.
———. 1969. A note on trophic complexity and community stability. Am. Nat. 103:91–93.
———. 1971. A short-term experimental investigation of resource partitioning in a New Zealand rocky intertidal habitat. Ecology 52:1096–1106.
———. 1979. Disaster, catastrophe, and local persistence of the sea palm Postelsia palmaeformis. Science 205:685–687.
Paine, R. T., and S. A. Levin. 1981. Intertidal landscapes: disturbance and the dynamics of pattern. Ecol. Monogr. 51:145–178.
Paine, R. T., and R. L. Vadas. 1969. The effect of grazing by sea urchins Strongylocentrotus spp. on benthic algal populations. Limnol. Oceanogr. 14:710–719.
Peters, R. H. 1976. Tautology in evolution and ecology. Am. Nat. 110:1–12.
Platt, J. R. 1964. Strong inference. Science 146:347–353.
Popper, K. R. 1959. The logic of scientific discovery. Basic Books, New York.
———. 1972. Objective knowledge: an evolutionary approach. Clarendon, Oxford.
———. 1976. Unended quest: an intellectual autobiography. Open Court, La Salle, Ill.
———. 1983a. A proof of the impossibility of inductive probability. Nature 302:687–688.
———. 1983b. Realism and the aim of science. Vol. I of the postscript to the logic of scientific discovery. Hutchinson/Rowan & Littlefield, London.
Quinn, J. F. 1979. Disturbance, predation and diversity in the rocky intertidal zone. Ph. D. diss. University of Washington, Seattle.
———. 1982. Competitive hierarchies in marine benthic communities. Oecologia 54:129–135.
Ricketts, E. F., and J. Calvin. 1968. Between Pacific tides. 4th ed. Stanford University Press, Stanford, Calif.
Schoener, T. W. 1965. The evolution of bill size differences among congeneric species of birds. Evolution 19:189–213.
———. 1974. Resource partitioning in ecological communities. Science 185:27–39.

———. 1975. Presence and absence of habitat shift in some widespread lizard species. Ecol. Monogr. 233–258.

———. 1983. Size differences among sympatric, bird-eating hawks: a worldwide survey. *In* D. S. Strong, Jr., D. S. Simberloff, L. G. Abele, and A. B. Thistle, eds. Ecological communities; conceptual issues and the evidence. Princeton University Press, Princeton, N.J. (in press).

Simberloff, D. S. 1974. Equilibrium theory of island biogeography and ecology. Annu. Rev. Ecol. Syst. 5:161–182.

———. 1978. Using biogeographic distributions to determine if colonization is stochastic. Am. Nat. 112:723–726.

———. 1980. A succession of paradigms in ecology: essentialism to materialism and probabilism. Synthèse 43:3–39.

Simberloff, D. S., and E. F. Connor. 1981. Missing species combinations. Am. Nat. 118:215–239.

Stephenson, T. A., and A. Stephenson. 1972. Life between tidemarks on rocky shores. Freeman, San Francisco.

Strong, D. R., Jr. 1980. Null hypotheses in ecology. Synthèse 43:271–285.

Strong, D. R., Jr., L. A. Szyska, and D. S. Simberloff. 1979. Tests of community-wide character displacement against null hypotheses. Evolution 33:897–913.

Sousa, W. P. 1979. Experimental investigations of disturbance and ecological succession in a rocky intertidal community. Ecol. Monogr. 49:227–254.

Usher, M. B. 1979. Markovian approaches to ecological succession. J. Anim. Ecol. 48:413–426.

Whittaker, R. H. 1975. Communities and ecosystems. Macmillan, New York.