Wiens JA (1989) Determining pattern and process: the logical structure of community ecology. Pp. 18-37 en: *The ecology of bird communities*. Cambridge University Press. Cambridge

2

Determining pattern and process: the logical structure of community ecology

Communities are interesting to ecologists, as MacArthur said, because of their patterns. Once detected, a pattern is something that to most ecologists cries out for an explanation – what process or processes caused the pattern? To document the existence of a particular pattern and determine the associated processes, however, is not so simple as many ecologists have thought. Patterns are derived from observations, which are our closest approximation of 'facts'. If the observations are made using flawed procedures, these 'facts' are suspect. If the derivation of patterns from the observations or the development of process explanations do not follow proper logical pathways, neither the patterns nor their interpretations can be trusted. Unless we are certain that the patterns we detect are accurate representations of nature, there is really nothing to be explained. Elaborating process explanations of such patterns simply compounds the uncertainty. Detecting community patterns and understanding their relationships to processes therefore requires a clear understanding of the logical structure of scientific investigation, which is my focus in this chapter. It also requires attentiveness to the operational structure of gathering and analyzing observations, which I will consider in the following chapter.

Pattern and process

In ecology, a *pattern* is a statement about relationships among several observations of nature. It thus connotes a particular configuration of properties of the system under examination. Neither observations nor patterns, however, are free of biases. Sale (1984) has noted that we screen or filter observations for patterns at a most elemental level, as our sensory abilities have evolved for detecting patterns. As a consequence, what we see or hear is a simplified, more ordered approximation of reality. As Grant (1977: 298) has observed, 'pattern, like beauty, is to some extent in the eye of the beholder'.

Superimposed in this sensory filtering of observations is an intellectual filtering that also influences pattern detection. What we perceive is determined by our mental preconceptions about what there is, and these preconceptions are closely bound to the language we use to express and organize our perceptions (Haila 1986). The specialized jargon used in a scientific discipline simply enhances the way language categorizes and constrains our preconceptions of nature. The patterns we see are strongly influenced by our culture.

Culture also affects pattern detection through its worldview, its preconceptions about how nature is. Western science is conducted within the context of a worldview that seeks *order* in nature (Wiens 1984a, Simberloff 1982, Dayton and Oliver 1980, Greene 1981). In accordance with this preconception, we find most interesting those patterns that express some regularity, and 'pattern' is often equated with the expression of an aspect of the apparent orderliness of life (e.g. Eldredge and Cracraft 1980). Because they are repeatable, ordered relationships also permit predictions to be made, and there is thus a pragmatic emphasis on such patterns as well.

Relationships among observations need not express regularity, however, in order to represent patterns. A scatter of points on a graph is a pattern, even though it may display no statistically evident regularity or order and may offer no predictions. The statement that there is no statistically significant pattern among such observations is not a statement of no relationship among the points but only a conclusion that the relationship is not *ordered* in some specific fashion. Admittedly, regular, repeatable relationships are more interesting than less regular patterns, and they should probably remain the primary focus of our investigations. To emphasize only those patterns that display apparent order and neglect those that do not, however, may foster unwarranted confidence in the orderliness of nature and leave many important patterns unexplored. Our challenge is to detect patterns in observations and to understand them, whether or not regularity in the patterns is immediately apparent.

A pattern, once established, represents something to be explained. Explanation rests on an understanding of *processes*, of what causes the pattern to be as it is. A process is thus synonymous with a cause, the operation of some factor or factors that produce a particular relationship among observations. A process is not simply something happening. If a community increases in diversity through time, for example, the diversity change or species enrichment is not a process but rather a change in pattern. Processes such as immigration or extinction are what brought about the diversity increase. Pattern and process are often confused. Thus, Thomson (1980:

720) considered numerical responses and niche shifts to be responses to competition, but he referred to them as processes that produced community patterns: 'both processes have equivalent end results, i.e. both affect community structure, species diversity, etc.' Numerical responses and niche shifts are population-level patterns that may contribute to community-level patterns, but the process causing them, in Thomson's view, is competition. The distinction between a pattern and a process is not always clear-cut, as all patterns contain the effects of processes. Moreover, what we call a process at one level (e.g. extinction) may be recognized as a pattern at some other level (e.g. the change in abundance from > 1 to 0). Despite the conceptual difficulty of dichotomizing patterns and process, it is important to distinguish between them operationally, for a failure to recognize pattern detection and process explanation as different phases of scientific activity has led to fuzzy thinking and incorrect conclusions, as we shall see.

There is, by definition, an inevitability in the sequence: cause (process) → effect (pattern). Because of this, there is a time lag between the operation of a process and the appearance of or change in a pattern. Current patterns are consequences of past or ongoing processes. In attempting to define cause–effect relationships, we usually assume that this time lag is sufficiently short that it can be ignored or that past processes operating in the same way that they do now (the assumption of uniformitarianism). If processes operate in an episodic fashion with considerable time lags, however, drawing links between process and pattern may be difficult. This difficulty is especially apparent in attempts to test evolutionary or adaptive arguments (e.g. Gould and Lewontin 1979, Brady 1982, Sober 1984), but it besets a good deal of community ecology as well (Connell 1980).

The philosophical foundation: hypothesis testing

How does one structure an investigation in order to detect pattern or determine process? Part of the task involves observing and measuring accurately, according to a proper design (see Chapter 3). Investigations must also have a proper logical structure. Insofar as the philosophy of science provides guidance in this matter, it deserves careful consideration. There exists a wide array of philosphical positions on how science should be or has been done, ranging from the rigid inductive framework of Logical Positivism (e.g. Carnap 1923; see Rosenberg 1985) or the strict falsification program of Popper (1959, 1962) through the more flexible positions of Lakatos (1978) and Laudan (1977), to the relaxed, 'anything goes' approach advocated by Feyerabend (1975). As a consequence, an ecologist can probably find a philosophical position to legitimatize whatever approach he takes. Roughgarden (1983: 583) suggests that science is rarely done following any 'formal rules'; rather, we develop knowledge by making a 'convincing case' for a position on the basis of experience and 'common sense'. What is convincing or common sense to one person may not be to another, of course. More importantly, common sense is what accords with preconceptions that are widely shared in a discipline, its 'conventional wisdom' (Strong 1983). Adopting an approach based on common sense increases the likelihood that only evidence that accords with this 'wisdom' will be considered.

It is not my intention here to review philosophies of science (see Hull 1974, Suppe 1977, Sarkar 1983, Rosenberg 1985) or to advocate any specific position as being 'best'. Nor do I wish to develop any tightly structured, formal philosophical framework that must be followed in order to practice 'good' science. Some degree of formalism in the logical operations of science nonetheless seems essential.

Most ecologists now purport to be testing hypotheses when they conduct their research, so I will emphasize the logical structure of hypothesis testing. This does not mean, however, that testing hypotheses is the only way to do science and gain knowledge. In order to develop enough basic understanding of an ecological system that tests are focused on realistic and sensible hypotheses, the system must generally be investigated in a less formalized fashion first. This may be through the formulation of preliminary models or 'working hypotheses', educated guesses or speculations that serve to guide research but are not directly testable (Thornhill and Alcock 1983). Exploratory data analysis, in which data are subjected to various forms of pattern analysis to see what emerges rather than to formal statistical hypothesis tests (Tukey 1977), may also be an important activity (James and McCulloch 1985). Such analysis may help to determine which of several working hypotheses merits formal testing. These less formalized approaches are *not* atheoretical – in fact, the observations are usually gathered with a particular theory in mind, even if explicit tests of the predictions of that theory may not be immediately forthcoming. To claim that research is not scientific unless it involves testing of hypotheses is absurd. Indeed, hypothesis testing in the absence of the necessary background provided by natural-history studies is likely to be a sterile and meaningless activity.

Deriving hypotheses and predictions

At some point, knowledge of a system is sufficient that provisional models or working hypotheses can be restated as formal hypotheses. Hypotheses are statements that take the general form 'if A, B, C, and D hold, then X follows'. They represent conjectures about nature, questions or propositions that, for a specific situation, have not yet been answered. The value of hypotheses lies in their capacity to provide clear answers to specific questions, but to do so the hypotheses must be testable and the tests must be fair. I have outlined the generalized logical procedures for testing hypotheses in Fig. 2.1: after discussing this framework, I will illustrate it with a specific example.

Hypotheses may develop by two pathways: inductively, from observations and patterns, or deductively, from theoretical constructs. In practice, this distinction is far from absolute, for there is a continual interplay between observations/patterns and theory in science. Although it is true that theory at times springs from bold, intuitive leaps, most often its development is prompted by observations of patterns or problems posed by them. On the other hand, theory often determines what sorts of observations are made or what kinds of patterns are sought (e.g. Haila and Järvinen 1982). Grant (1977) has suggested that, after Lack embraced the theory that competition produces ecological segregation among species, he forced observations to fit that framework.

In order to be testable, hypotheses must offer clear predictions. Often neither the general theory nor a more restricted hypothesis is very specific or unambiguous, but the predictions (which are what we test) must be precise and must logically follow from the hypothesis. There is more behind a prediction, however, than a logical connection with a hypothesis. In order to generate specific, restricted predictions, assumptions must be made; more often than not, these are left unstated. Many of these assumptions are contained in other theories and are usually accepted without challenge (Kuhn 1970a, Lakatos 1978, Dunbar 1982). Brown (1981), for example, noted how predictions of the general theory of competition in ecology are not only dependent upon the assumptions of that theory but also upon the premises of other background theories, such as the Lotka–Volterra model of population interactions. Changing these assumptions in biologically realistic ways may lead to totally different predictions (e.g. Abrams 1976, 1983). Moreover, any prediction contains a *ceteris paribus* (all other things being equal) clause, which effectively permits one to ignore a host of other variables or effects (Brady 1982). The more that such detail is contained in the predictions instead of sheltered in the background assumptions, however, the clearer tests of the predictions will be.

It is at this stage that one of several violations of logic may appear. Suppose that one has gathered a number of observations and from them detected a pattern. A hypothesis is developed to explain the pattern.

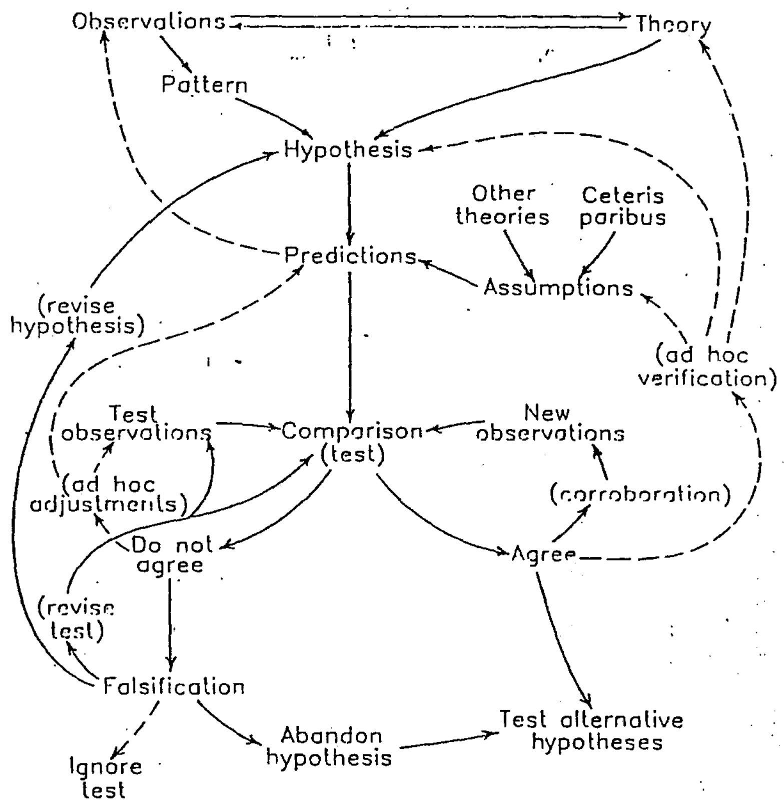

Fig. 2.1. Flow chart of the steps involved in hypothesis-testing. Solid lines show acceptable paths; dashed lines indicate logically flawed operations.

Predictions are generated and then compared with the original observations. They agree, and the original observations and pattern are then considered 'evidence' that 'proves' the explanation (hypothesis). Because the same observations were used to generate and to test the hypothesis, this agreement is not surprising. The 'test' depends on circular logic, however, and is totally invalid.

Testing and falsification

What, then, is a test of the predictions? Testing rests on a comparison of the predictions with some set of new observations. If the hypothesis is constructed properly, two results are possible: the observations agree with the predictions or they do not [a third outcome, neither 'yes' nor 'no', is eloquently discussed by Pirsig (1974: 288)]. For a hypothesis to be testable, both outcomes must be possible. The hypothesis must be open to question and its predictions susceptible of something other than only support or only rejection. A theory that is guaranteed by its internal logical structure to agree with all conceivable observations is more characteristic of religion than of science (Lewontin 1972, Dayton 1979).

Falsification, or the potential of falsification, is thus a key element in scientific hypothesis-testing. An emphasis on falsification is central to Popper's (1959) philosophy of science. Popper argued that, because one can have greater confidence in the truth of a falsification than of a verification of a hypothesis, falsification contributes more to scientific progress than does verification. Popper also stressed the value of 'risky' predictions: those that have the greatest *a priori* probability of falsification are the best predictions, for a failure to falsify them is a strong corroboration of the predictions and the hypothesis. This does not mean, of course, that one should intentionally construct and test hypotheses that are blatantly absurd on the outside chance that they will not be falsified and we will all be surprised. Intellectual honesty is also a necessary ingredient in hypothesis formulation.

Popper's original formulation of this falsificationist philosophy was rigid: a single falsifying test could lead to the rejection of a hypothesis. Both philosophers (e.g. Lakatos 1978, Laudan 1977) and practicing ecologists (e.g. Simberloff 1982, Quinn and Dunham 1983, Hilborn and Stearns 1982, Haila 1982) have argued (correctly) that rigid adherence to such a scheme is unrealistic and likely to detract from understanding. They suggest that one needs to balance the frequency and strength of falsifications of an hypothesis against the frequency and strength of prior corroboration of its predictions. Popper himself relaxed his initial position in later writings (1962, 1983). Some philosophers, however, abandoned Popper's falsificationist emphasis entirely (see Suppe 1977, Bronowski 1977), leading ecologists such as Diamond (1986a) to question the value of the approach in scientific investigations. Although it is true that a rigid adherence to a strict falsificationist philosophy can lead one to discard useful hypotheses on the basis of trivial exceptions, a more balanced approach to falsification provides a valuable logical structure for scientific investigations.

One outcome of a test, then, is falsification: the observations do not agree with those predicted. What can one conclude from this? The most obvious conclusion, of course, is that the predictions were incorrect and the hypothesis is therefore false. Several other explanations of such a result are possible, however (see Clutton-Brock and Harvey 1979, Brady 1982, Jarman 1982, Hilborn and Stearns 1982):

1. The observations are faulty. They may have been gathered in a biased manner, such that negative evidence was somehow favored and confirmatory evidence disregarded. The observations may have been incomplete or insufficient, perhaps because some key variables were not measured. This leads to an incorrect documentation of pattern and a failure of the comparison. A test based on such observations is of course unfair, but this may not become evident until the test is actually made (if then). Such problems with observations can apply either to the test observations or to the original observations on which the hypothesis was founded, or to both. If the test observations are faulty, the test must be revised and conducted again; if the problems are with the original observations, the hypothesis itself must be revised (Fig. 2.1).
2. The observations are accurate but the result is a statistical artifact. There is always a possibility that the results may not agree with the predictions as a consequence of chance and sampling variation. This is likely to lead to an unjustified rejection of the hypothesis (a Type I error), especially if there has been little replication in the test. There are other ways in which the behavior of samples or sets of numbers can produce apparent patterns that are mathematical or statistical artifacts; May (1975) has considered some of these in detail. In either instance, revisions of the test are necessary.
3. Patterns in the results may be affected by other variables or causes. Interacting variables that were not controlled for in the test comparison or inappropriate variables that were included in the observations may act as 'noise', obscuring the relationship being tested. The pattern examined by the hypothesis may have more than one cause and, if the hypothesis and its predictions focused on only one of these, the expected pattern might fail to appear. These problems stem from violations of underlying assumptions of the hypothesis. Revisions of both the hypothesis and the test are required.
4. The comparison itself is flawed. Perhaps the predictions and the test observations are both logically and empirically correct, but they are compared in an inappropriate manner (e.g. an incorrect application of multivarate statistics), such that the expected pattern simply cannot be demonstrated. Perhaps the predicted pattern is tested with a heterogeneous data set that contains observations

that are not comparable but are nonetheless compared (Andrewartha and Birch 1984). This calls for a revision of the test.

5. The assumptions may be violated. When a test of predictions is conducted, everything that is behind those predictions is part of the test, including assumptions derived from other theories and the *ceteris paribus* clause. If one tests a hypothesis that assumes ecological equilibrium by using observations from a system that is clearly nonequilibrial, for example, the failure to match the predicted patterns may reflect problems with the underlying assumptions and inappropriateness of the test rather than a failure of the hypothesis itself. Revision of the test is needed, but a hypothesis so heavily laden with assumptions that clear tests are difficult also merits reconsideration.

   The *ceteris paribus* assumption is often especially suspect, and this leads to another sort of logical transgression that is frequent in community ecology. Faced with an apparent falsification of a cherished hypothesis (a failure to document ecological differences or 'niche partitioning' among similar, syntopic species, for example), one may argue that the test and the falsification are invalid, as other things were not equal (e.g. the species may partition niches on some other, unmeasured dimension). As the domain of possible unmeasured variables is quite large, such *ad hoc* practices effectively make the hypothesis untestable by sheltering it from falsifications.

6. The logic is faulty. Each of the steps between the original observations or theory and the comparison of test observations with predictions in Fig. 2.1 involves a logical operation. If one has erred, for example, in deducing the hypothesis from a general theory or in deducing predictions from the hypothesis, falsification is a likely consequence. Revision of the hypothesis and review of the logic used in each step are required.

Clearly, a falsification may result from much more than a simple failure of the hypothesis being tested. These additional reasons for falsification all reflect poor design of either the hypothesis or the test (or both). If a test is carefully structured, the influences of confounding variables will be minimized, observations will be accurate and complete, the comparison will be valid, all assumptions will be clearly stated and matched in the test, and so on. Few tests in ecology are so carefully designed, however, and errors or biases may affect the test even if one *is* careful. Unless one considers these

other possible reasons for a falsification, there is a real danger of being misled by such a result.

There is a third abuse of logical procedures that may occur at this stage (Fig. 2.1). Finding that one's observations do not agree with the predictions, one may be tempted to alter either the observations or the predictions in an *ad hoc* manner to bring them into closer agreement. This may be done, for example, by adding another term to an equation, by changing a model from a linear to a nonlinear form, or by removing 'outlier' data points to change a nonsignificant relationship into a significant one. Sometimes such procedures are justified, and part of the process of testing hypotheses includes changing the hypothesis or predictions to accommodate apparent anomalies. Such revised hypotheses then require fresh tests, however. To modify the data or the predictions and then to treat the modified set as a 'test' and 'verification' of the hypothesis and theory in the absence of additional, independent tests is logically invalid. Nonetheless, it is done (see Wiens 1983 for examples).

Corroboration

What if the observations of a test agree with the predictions? This is usually taken as evidence that the hypothesis is correct or, at the very least, corroborated. Our confidence in the relationships expressed in the hypothesis is bolstered, especially if the test was a 'risky' one. Such seemingly confirmatory results, however, are subject to the same sorts of alternative interpretations that apply to falsifications (items 1–6 above). Thus, just as faulty observations can mislead one into falsification, so also can they produce apparent agreement with expectations. Inadequate sampling can contribute to this, by fostering Type II statistical errors. The comparison used in the test or the logic used in deriving the hypothesis or predictions may be flawed in such a way that the likelihood of confirmation is enhanced. The predicted results may be due to other totally different causes, and our conclusions regarding the hypothesis under test will thus be misguided. Elements of untestability hidden among the underlying assumptions may also lead only to confirmatory results. As a consequence of these factors, corroboration may not be as compelling as it might appear. Further tests, structured so as to deal with such confounding factors, are required.

There is yet another logical error that often accompanies apparent corroborative tests (Fig. 2.1). Provided with observations that match the predictions of some hypothesis, one may conclude not only that the predictions have been verified, but that the hypothesis, its underlying assumptions, or the general theory as a whole have been proven correct as well. Finding ecological differences among coexisting species, for example, one might conclude that such differences are a result of competition, which created the niche partitioning that permits the species to coexist. Such conclusions are not warranted by the test but are inferences and assertions only. The problem is that the observations (differences among the species) do match the predicted pattern, but the role of the postulated process (competition) in producing the pattern has itself not been tested. Burdon-Sanderson drew attention to this problem in his inaugural address to the British Association for the Advancement of Science in 1893: 'To assert that the link between *a* and *b* is mechanical, for no better reason than that *b* always follows *a*, is an error of statement, which is apt to lead the incautious reader or hearer to imagine that the relation between *a* and *b* is understood, when in fact its nature may be wholly unknown' (1893: 467). To a large extent, this logical trap is a consequence of an emphasis on verification rather than falsification of hypotheses. Andrewartha and Birch (1984) term this the 'error of the misplaced premise', noting that the explanation or theory seems so plausible that its truth may be accepted with absolute certainty.

Testing alternatives

One way to reduce the likelihood of errors in falsification or corroboration is to make the hypothesis and predictions clear and concise, minimize their reliance on underlying assumptions, and conduct the test following rigorous procedures. It is also important to consider alternative hypotheses that address the same general question but do so using different sets of assumptions or premises and thus lead to different predictions. This method of 'strong inference' (Chamberlin 1965, Platt 1964) may be particularly appropriate when one suspects that a pattern is the result of multiple causes or the interaction of several variables (but see Haila 1982, Hilborn and Stearns 1982, Quinn and Dunham 1983). This approach is represented in tests of statistical hypotheses by contrasting a null hypothesis of no effects of a treatment or relationship among variables with one or more alternative hypotheses that offer specific predictions of patterns. Ideally, these tests are constructed so that falsification of the null hypothesis leads logically to provisional acceptance of the alternative hypothesis. In ecology, however, it is often difficult to devise null and alternative hypotheses that are logically complementary to each other (Connor and Simberloff 1986) and the null and alternative hypotheses must therefore be tested separately.

Ecologists have also given particular attention to conceptual null hypotheses or neutral models, which are conjectures that a particular process of interest has *not* contributed causally to the observations under study (e.g. Caswell 1976, Strong 1982, Connor and Simberloff 1983, 1986, Simberloff 1983a, Strong *et al.* 1979, Harvey *et al.* 1983). Such conceptual null hypotheses contrast with the approach of examining data for consistency with the predictions of hypotheses that contain specific causal processes in the absence of any knowledge of what values such data would take were those processes not operating (Connor and Simberloff 1986). Roughgarden (1983) has argued in favor of this approach, noting that null hypotheses are 'empirically empty' because they contain no biological processes that produce the predicted patterns. We learn nothing by falsifying such hypotheses, he claims, and they therefore do not represent viable alternatives.

At the opposite pole, Strong *et al.* (1979) have argued that, because the null hypothesis portrays what might be expected in the absence of a particular process, it has logical priority over other hypotheses. Whether or not the null hypothesis has logical priority or is simply a good starting point for investigations (Strong 1983, Simberloff 1983a), it seems clear that some sort of hypothesis based on the absence of key process assumptions should be among the alternatives considered when evaluating a specific hypothesis. Contrary to Roughgarden's view, conceptual null hypotheses *are* formulated relative to some specific causal explanation; consequently, they are not empirically empty, so long as they are correctly formulated. Both the value of null hypotheses in ecology and the procedures used in testing them have generated considerable controversy (e.g. Connor and Simberloff 1984a, Gilpin and Diamond 1984a), which I consider further in Chapter 4.

Concluding testing

How does one know when to stop cycling through the loops of Fig. 2.1 and to conclude that a hypothesis has been satisfactorily falsified or corroborated? When this occurs is a matter of judgment, based on 'convincing evidence' and 'common sense' (Roughgarden 1983). The danger, as Kuhn (1970a) so clearly shows, is in stopping prematurely in allowing one's preconceptions to equate 'convincing evidence' with the 'right' results. In a sense, testing of a hypothesis is never really completed. It is appropriate to move on to other questions, however, when, on the basis of sound logic, good design, appropriate methods, and 'risky' tests, the predictions have been clearly falsified, repeatedly corroborated, or have failed in a way that leads to a revision of the hypothesis.

Hypothesis testing and normal science

If Kuhn (1970a) is correct in his view of how science actually operates, it is important to know if the form or use of hypothesis testing differs during different phases of activity in a discipline. Do individuals engaged in 'normal science' test hypotheses in the same fashion as those embroiled in controversy or crisis? Both Kuhn (1970c) and Popper (1970) have addressed this issue; not surprisingly, they reach different conclusions. Popper does not dispute that some evidence of 'normal science' exists. He regards an individual engaged in such activity, however, as 'poorly taught': 'he has been taught in a dogmatic spirit; he is a victim of indoctrination. He has learned a technique which can be applied without asking for the reason why.' Popper suggests that, because one can always undertake critical comparisons of competing theories, using the methods of bold conjectures and criticisms that he regards as central to hypothetico-deductive science, *good* scientists do not suffer from the symptoms that Kuhn describes.

Kuhn, on the other hand, suggests that during periods of 'normal science' most individuals are engaged in 'puzzle solving', that is, in answering questions that are posed within the context of the prevailing paradigm. Hypothesis testing in a falsificationist spirit is usually not part of this endeavor (contra Fretwell 1972) because the emphasis is not on strong or 'risky' tests or falsification but rather on verification of variations of the paradigm. Because they take current theory for granted, individuals 'are freed to explore nature to an esoteric depth and detail otherwise unimaginable. Because that exploration will ultimately isolate severe trouble spots, they can be confident that the pursuit of normal science will inform them when and where they can most usefully become Popperian critics' (Kuhn 1970b: 247). Kuhn views formal hypothesis testing as 'the strategy appropriate to those occasions when something goes wrong with normal science, when the discipline encounters crisis'.

Neither of these views is entirely correct. Whether labelled as such or not, many scientists do undertake a form of hypothesis testing in their normal activities, whether or not controversy exists. There is no doubt, however, that interest among ecologists in philosophy and concern over the appropriate logical structure of scientific investigation were expressed much more strongly beginning in the late 1970s than they were during the previous two decades, when symptoms of Kuhnian 'normal science' were more apparent in community ecology.

Testing hypotheses: an example

To explore some aspects of how hypotheses of pattern and process are actually developed and tested in community ecology, it is helpful to consider a specific example. Using the format of Fig. 2.1, I have indicated some of the major features of what might be termed the 'size-ratio hypothesis' in Fig 2.2 (see Wiens 1982, Simberloff and Boecklen 1981, Roth 1981, Simberloff 1983b). This idea is evaluated further in Chapter 7.

We begin with Hutchinson's (1959) original observation that bill lengths within several sets of congeneric, sympatric bird species differed by an average ratio of 1.3. Hutchinson's ratio value was derived inductively, but the observation was made theoretically explicit by Hutchinson and MacArthur (1959). The theory was subsequently enriched through general competition theory and, especially, limiting similarity theory (e.g. MacArthur and Levins 1967, May and MacArthur 1972). This work generates the hypothesis that competition imposes a limit to the similarity of sizes of coexisting, ecologically similar bird species. A specific, testable prediction of this hypothesis is that such species will differ in bill lengths by an average ratio of 1.3; this also implies that several coexisting species will be spaced along a bill-length spectrum by constant ratios of 1.3.

Several assumptions are contained in this prediction. Because it is derived in part from general competition theory, we assume that food resources are limiting to the species and that the species use similar resources (i.e. they are indeed potential competitors). Implicitly, it is also assumed that the community is fully packed with species. More specifically, one must assume that the food-size resource spectrum is continuous and continuously available to the species (MacArthur 1972) and that food-resource use is a direct function of bill length. Finally, there is, of course, the *ceteris paribus* clause.

The predictions have been tested by measuring the bill lengths within sets of similar coexisting species and calculating the ratios. Additional tests have involved determinations of the frequency distributions of size ratios over a large number of sets of species (e.g. Schoener 1965, 1984, Roth 1981, Simberloff and Boecklen 1981). It turns out that for many of these tests the prediction has not been upheld: the size ratios neither average 1.3 nor are they constant. Does this represent a clear falsification of the hypothesis? Perhaps, but other explanations come to mind. The tests might be flawed in various ways, for example. Perhaps the comparison has included species that are not really ecologically similar, obscuring any pattern that might exist among the subset of similar species (e.g. Bowers and Brown 1982, Grant and Abbott 1980). Bill length was measured because it was consid-

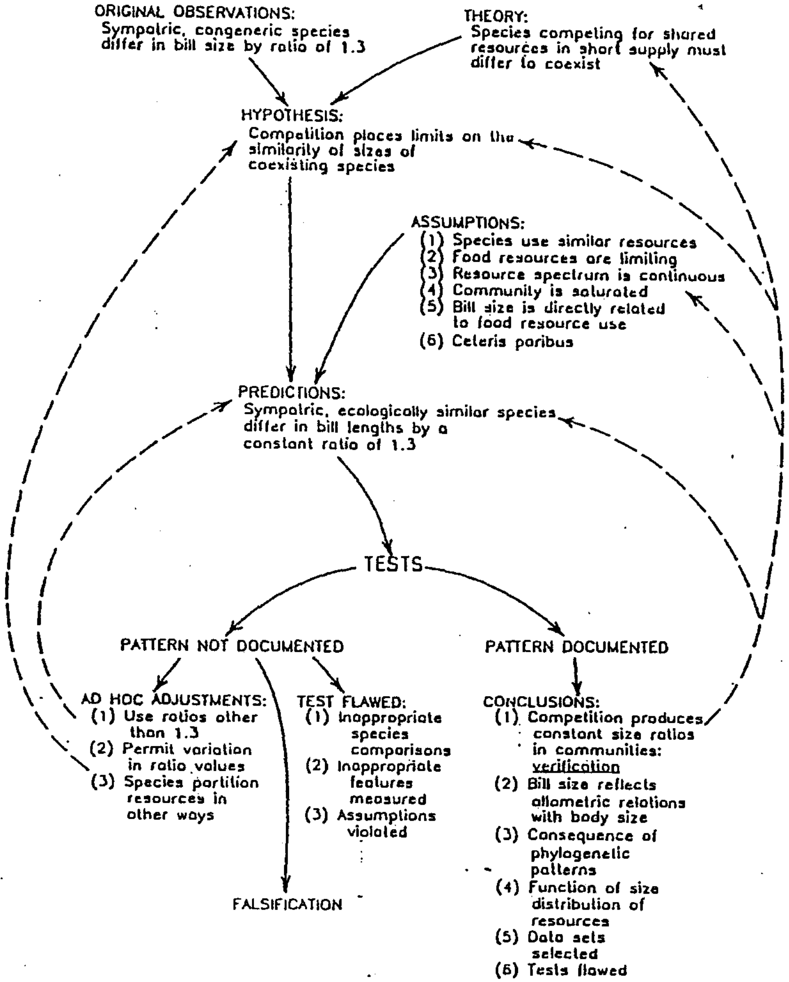

Fig. 2.2. An example of the operations involved in testing the 'size–ratio' hypothesis. The background of this hypothesis and the tests are developed more fully in Chapter 7. The arrangement of the diagram follows that of Fig. 2.1.

ered the variable most relevant to capturing prey of different sizes, but perhaps bill shape is more important than length alone (e.g. James 1982) or perhaps allometric relationships with body size impose constraints on variation in bill lengths (Clutton-Brock and Harvey 1979). Perhaps one or several of the underlying assumptions has not been met, and the tests are therefore invalid. These possibilities all reflect inappropriate tests of the hypothesis, but they may nonetheless affect the results that have been obtained.

Clearly, there may be valid reasons for concluding that the lack of agreement between the prediction and the test observations is not a strict falsification of the hypothesis. A less logical route to preserving the hypothesis involves *ad hoc* adjustments of various sorts. Thus, size–ratio values other than 1.3 (e.g. 1.05, 1.1, 1.2, 1.4, 2.0) have been considered to be consistent with the hypothesis by various investigators (Simberloff 1983b). Maiorana (1978) reconciled observations of substantial variation in ratio values within a sequence with the hypothesis by arguing that larger species in a sequence should be expected to exhibit greater variance in resource use than smaller species and thus require wider spacing in order to coexist. Others have suggested that the apparent falsification simply means that the species partition resources in ways other than by food size.

In some studies, reasonable agreement of observations with the prediction has been found. What conclusions can be drawn from this corroboration? The most obvious is that the prediction is upheld in this instance: the expected pattern, in fact, was found. But other possibilities exist. If bill length varies allometrically with body size, the pattern may reflect something having to do with body size rather than bill length alone. If the comparison has been restricted to a single phylogenetic lineage, the size patterns may be related to the phylogenetic history of the lineage rather than to the proximate ecology of the species (Wiens 1982). The regular spacing of species along the size gradient may be a consequence of the distribution of nodes and gaps along a discontinuous resource–size spectrum. In some instances, the pattern may apply only to a selected data set within a larger group of suitable data (Simberloff 1983b). Finally, the test itself may be flawed by the lack of any statistical analysis to determine whether the observations actually fit the predictions or just seem to.

If the results of the test agree with the predictions, however, such alternatives usually are not considered. The findings are regarded as verifying the hypothesis and may be used to support the correctness of the underlying assumptions as well (e.g. bill length *is* directly related to food-resource use). Frequently, the consistency of the observations with the predictions leads to the conclusion that the process contained in the general theory, competition, is indeed the cause of the observed pattern and has been demonstrated to occur. The pattern is 'explained'. This conclusion is not only logically unwarranted, but it may stifle further, more specific tests (e.g. coupling body-size differences to differences in foraging efficiency and thence resource use: Werner 1984). Such inferential 'proof' of a process underlying an observed pattern rests on the presumption that nothing else is likely to have produced that pattern, but this may not be so. The pattern *may* result from competition, to be sure, but it may also arise from other quite different processes (Wiens 1982, Simberloff 1983b). Given this, no conclusions about process from the documentation of a particular pattern are warranted.

Pattern hypotheses and process hypotheses

The temptation to explain patterns by inferences about processes, as in the above example, arises because the hypothesis contains statements about *both* pattern and process. Thus, even though the specific prediction relates only to pattern, it is a small step to accept the close coupling of process with pattern contained in hypothesis. In fact, most ecological hypotheses contain the proposition that a pattern (X) has been produced by a process (Y). They state (either explicitly or implicitly) that if Y occurs, then X will follow. Such causal 'if–then' statements contain two parts, which really represent *separate* hypotheses to be tested: (a) does X occur? and (b) if so, was it a consequence of Y? A failure to separate the process and pattern portions of a hypothesis in tests generally leaves the process buried within the underlying assumptions and available to 'testing' only by assertion and inference.

In pattern hypotheses we ask whether a particular relationship holds among a number of observations. Patterns are usually detected by comparisons, and the so-called 'comparative method' (Clutton-Brock and Harvey 1979, Jarman 1982, Thornhill and Alcock 1983) provides a framework for testing pattern hypotheses. The actual tests of such hypotheses, however, are often statistical. We may hypothesize, for example, that there is a linear relationship between two variables and then use linear regression and correlation statistics to test this hypothesis. James and McCulloch (1985) and Connor and Simberloff (1986) have emphasized the distinction between these statistical or 'empiric' hypotheses and what they call 'scientific', 'research', or 'theoretic' hypotheses, which are statements about processes. Statistical (pattern) hypotheses are tested by inductive inference, by sampling from clearly defined populations to evaluate pattern predictions. The

emphasis is on developing reliable *predictions*, and such tests address theory only to the extent that a theory generates unique and unambiguous predictions. Popper (1959, 1962, 1983; Popper and Miller 1983) explicitly disavowed a legitimate role for inductivism in scientific investigations, and statistical hypotheses, which are founded firmly on inductive tests, are therefore not 'proper' Popperian hypotheses, even though the emphasis is still upon falsification. This, of course, does not diminish their usefulness in scientific investigations.

The documentation of a regular, repeatable pattern reveals something requiring explanation: what *caused* the pattern? Hypotheses about processes are considerably more difficult than pattern hypotheses to frame and test. This is because a process must be isolated in the test to reveal its effects upon pattern; *ceteris paribus* must be assured. Some workers (Alexander 1979, Mayr 1982, Thornhill and Alcock 1983, Thornhill 1984) have argued that, with appropriate comparisons, the comparative method can be used to examine causal forces as well as to detect patterns. Appropriate comparisons are those that, because of their number and diversity, are likely to randomize and therefore control the influence of other variables on the result (Thornhill 1984). If it were possible to achieve complete randomization of other variables in a comparison, this might be true, but intercorrelations among variables make such randomization of effects unlikely.

Increasingly, experiments are suggested as the best way to test process hypotheses and determine causality. In experiments, critical confounding variables are controlled rather than presumed to be randomized and the variable of interest is manipulated. For example, if one hypothesizes that the habitat niche breadth of species A is restricted by the presence of a competing species B, the expansion of A's habitat niche when B is experimentally removed should be clear evidence of B's competitive effects when it is present. Such experiments and their interpretations, however, are not without problems (Hilborn and Stearns 1982, Quinn and Dunham 1983, Hurlbert 1984, Bender *et al.* 1984, Underwood and Denley 1984, Underwood 1986b), some of which are considered in Chapter 3. Moreover, in many instances logistic constraints or other considerations (e.g. endangered species or habitats) make experimental manipulations impractical. In these situations, one must rely on using carefully planned, nonmanipulative comparisons to test process hypotheses (Diamond 1986a).

In the end, of course, experiments are simply another form of comparison, in which the investigator establishes the domain of the comparison rather than leaving this up to Nature. The results of experiments are patterns, not processes. Thus, although experiments may isolate the *effects* of processes, they rarely test process explanations directly. One must still rely on inference to reach conclusions about process. Experiments may foster a greater confidence in such causal inferences than does the comparative method, but they do not assure the truth of such inferences. As Andrewartha and Birch (1984: 222) have observed, 'there is no epistemological magic by which a precise probability (far less a certainty) can be attributed to a causal explanation'.

Conclusions

The philosophical and logical foundation of posing questions or testing hypotheses is important here because it provides a necessary framework for evaluating and designing investigations of bird communities. Unfortunately, not all investigators have been attentive to these matters, and this casts doubt upon the conclusions they have reached. Ecologists should recognize that the detection of pattern and the documentation of underlying process are conceptually separate and generally sequential phases of scientific investigation. Both pattern detection and process documentation are best done by following a flexible hypothetico-deductive procedure in which hypotheses that are derived from observations and/or theory are used to generate specific, testable predictions. 'Testable' means that the predictions, and thus the hypothesis, must be susceptible to falsification or corroboration. Interpreting the outcome of such tests, however, is not a simple matter, for a variety of factors other than the apparent truth or falsity of the hypothesis can contribute to the outcome. Caution and insight are required in reaching the conclusion that a particular test falsifies or corroborates the hypothesis. Pattern hypotheses may be tested by direct, nonmanipulative comparisons, but this 'comparative method' is of limited value in testing process hypotheses, for which experiments may be more appropriate. Conclusions about processes drawn from experimental tests, however, still rely on inference. A well designed experiment may increase our confidence in such conclusions, but it cannot guarantee certainty.

Opportunities for committing logical errors exist at several points in this hypothetico-deductive framework. Most apparent are those in which (a) *ad hoc* changes are made in the hypothesis or its predictions in order to fit the observations, which are then used as 'proof' of the hypothesis, or (b) the agreement between observations and predictions is used as a basis for inferring that the underlying assumptions or the process explanation have been confirmed. Neither is an acceptable practice.