

Ecological problems and how they are approached

Introduction

Much of ecology consists in making observations and then devising plausible explanations for the observations. Because alternative explanations of the observed phenomena frequently are available, the process by which the conclusions have been reached is known as "weak inference." It is not that the conclusions are necessarily wrong; the problem is that there is little assurance that they are right, and the widespread use of the approach has led to severe criticism, from both outside and inside the field.

In principle, manipulative experiments provide a preferable alternative, because their planning requires at least an implied prediction of the outcome, and making predictions is an integral part of science. The after-the-fact explanations mentioned earlier are sometimes called "predictions" by their promulgators, but there is no assurance that the information was not available before the prediction was made. There is no implication of dishonesty. The danger is as follows: All workers in ecology have a lot of factual information about the real world, and assuming that they are interested in understanding nature, they think about how things work. It is virtually impossible to separate known facts from the thought process. The knowledge is, perhaps unconsciously, taken into account in formulating hypotheses about how nature works. Thus, when a "test" is proposed, the prior knowledge is used, and the outcome is known in advance. Such a confirmation is, of course, spurious. I have selected an example from one of my own publications (Hairston 1964). In a consideration of the relationship between species abundance distributions and trophic level, I predicted that there should be a greater variance among the abundances of terrestrial carnivores than among the abundances of the coexisting herbivores. The reasoning was that the carnivores should be affected more by competitive interactions, which would be expected to be stable and hence would lead to large differences in abundance. Herbivores, on the basis of the arguments in Hairston, Smith, and Slobodkin (1960),

should be affected by their respective intrinsic rates of increase, which should vary irregularly with weather conditions, involving frequent reversals. Thus, the herbivores would have less chance for long-continued trends in their respective abundances, and they were predicted to show less variance in abundance. After a tenuous series of assumptions and calculations, I presented evidence that confirmed the predictions for soil arthropods. It must be pointed out that my graduate students and I had carried out research on those organisms irregularly for thirteen years (Hairston & Byers 1954; Hairston 1959; Engelmann 1961), and much of the relevant information must have been in my mind, at least subconsciously, at the time that I made the predictions. Such lapses of rigor were common at the time, and remained largely unchallenged until the current controversy over hypothesis testing in ecology. [See the papers in Strong et al. (1984b), for example.]

Appeals are sometimes made to the complexity of nature, apparently in the hope of avoiding the rigorous thought that is required to formulate clear hypotheses, and to avoid devising proper tests to permit choices among competing hypotheses. The process of formulating clear hypotheses and devising acceptable tests, called "strong inference," was clearly described and vigorously promoted two decades ago in a famous paper by Platt (1964). An integral part of the process is planning experiments that will yield answers to the questions that are asked about nature. Some ecological questions are posed on such a broad scale and are of such complexity that satisfactory experiments have not yet been devised to answer them. Nevertheless, successful completion of many field experiments over the past twenty-five years has shown that some problems that once were considered unapproachable by experimentation have been attacked and solved.

Before considering the ecological problems that remain to be solved, we may as well admit that one person's solution is another's challenge. That is because ecologists are not in full agreement as to what is acceptable as "proof." There are some who regard at least one of the following questions as having been answered, and probably there are some who do not agree that these questions are important.

Ecological questions amenable to experimentation

All of the questions that follow have been attacked through field experiments, but few ecologists would argue that they have been

answered fully. In later chapters there are descriptions and critiques of a number of these experiments. The experiments are arranged according to the various habitats in which ecologists have worked, rather than the kind of questions that follow. The reason for categorizing experiments by habitat is the hope that the arrangement will encourage an understanding of the ecological processes in each kind of setting, rather than trying to generalize over all settings. The field has not advanced to the stage where we even know if such generalizing is possible in principle.

What determines the abundance of species in nature?

This question continues to vex ecologists, despite much work on the subject since 1957, the date of the Cold Spring Harbor Symposium at which the proponents of various views vigorously aired their opinions. The stimulus for the symposium was the debate over density dependence versus density independence. One of its unfortunate aspects was that the protagonists never agreed on the terms of the debate. Both groups conceded the accuracy of the observations of the other. It was the relevance of the observations that was at stake. On the one hand, there were the plain correlations of abundance and population changes with aspects of weather (Davidson & Andrewartha 1948a,b), or the determination of the intrinsic rate of natural increase by laboratory-controlled temperature and moisture (Birch 1953). On the other hand, there were many laboratory populations showing a reasonably good fit to the logistic (the most density-dependent) model, long-term field studies showing fluctuations around a fairly constant mean, and the logical and mathematical argument that without density dependence, populations would fluctuate with increasing amplitude, eventually either going extinct or reaching completely unrealistic numbers. After the symposium, the debate was not so much settled as exhausted, and except for the exchange between Smith (1961, 1963a) and Andrewartha (1963), most ecologists simply turned to other interests.

The dichotomy has been shown to be false, in that unexpected effects have been found. For example, it had been more or less assumed that the limitation of a population by its food supply automatically meant that the population was under density-dependent control. Yet Wise (1975) showed that spiders were food-limited, without their abundance

being able to affect the supply of food, and Juliano (1986) reported a similar finding for scavenging carabid beetles. Filter-feeding sessile organisms in large bodies of water must frequently be in the same kind of situation.

As a result of work in the last three decades, the question is now posed in a more sophisticated manner, and we are concerned with whether density-dependent factors operate continuously or are interrupted by stochastic disturbances. Some ecologists are arguing that the disturbance itself is necessary for the stability of some ecosystems (Lewin 1986). If that is the case, we need to know the frequency of the disturbance, and especially the environmental factor that constitutes the disturbance. The dispersal phase in the life cycle seems to be especially sensitive to density-independent influences, especially when the dispersers are numerous and small and are carried passively by wind or water. Their success or failure in reaching a particular habitat has been known to determine the outcome of carefully executed experiments. For example, Paine's famous demonstration that a predatory starfish determined the composition of the community of sessile invertebrates (Paine 1966, 1974) could not be confirmed by his student Dayton (1971) because the mussels, which were the competitive dominant, failed to settle in large numbers during his study (Underwood & Denley 1984).

How important are interactions between species?

Ecological field experiments have established beyond question the fact that interspecific competition, predation, and mutualism are not rare in nature. What has not been settled is whether or not any of them, especially competition, is sufficiently pervasive to have an important effect on the dynamics and structure of assemblages of coexisting species. In two surveys of the literature that appeared almost simultaneously, Schoener (1983) and Connell (1983) reached different conclusions about the prevalence of interspecific competition, as shown by field experiments. The difference was only partly due to Connell's more restricted set of journals and years. Interpretation was more important. Connell was more rigorous in what he accepted as a satisfactory experiment, but he omitted from his analysis experiments involving an effect on one species at one time and place. As might be expected, he found interspecific competition less common than Schoener did. Schoener (1985) has reviewed the differences between the

analyses, and it seems clear to me that the difference is largely related to the prior mind-sets of the two authors, Connell (1975, 1980) having moved away from an initial partiality toward the frequent occurrences of competition, and Schoener (1974, 1982) arguing strongly for their importance in ecological systems.

Mathematically derived community theory relies heavily on the assumption that interspecific competition is the primary negative interaction among species. Therefore, the question debated by Connell and Schoener is vitally important to our understanding of the sets of coexisting species that we call guilds. The term "community" is here used in its original sense to mean all organisms in an area – the living part of an ecosystem. Thus, the application of theory really applies almost exclusively to guilds, as restricted from the original proposal of Root (1967), to include by usage, if not by formal agreement, a group of taxonomically related species that presumably are competing for a common resource. The critical word here is "presumably," because even in cases where competition has been conclusively demonstrated by satisfactory experimentation, the finding is nearly always confined to fewer species than a prior choice of guild membership would dictate (Hairston 1984). Schoener (1986) has discussed the usages and defended the restrictions. I refrain from quoting his preferred term, and continue to use "guild" in the restricted sense. In no example is guild membership based on demonstrated interspecific competition among all necessary combinations of the proposed members. The existence of ecological or morphological differences has been used to justify the assumption that guilds are structured by competition among the members. But that leads to the circularity that the differences arose because competition was the force selecting for the development of the differences. Noncircular identification of guilds based on interspecific competition would be difficult; there has been no successful attempt in that direction.

Given this unsatisfactory state of affairs, our recourse must be to conduct the complete set of experiments necessary to test the postulated composition of a guild. Some authors accept the assumption that competition in the past was responsible for the differences and that therefore competition is no longer detectable. They ignore the fact that prevailing theory does not require the elimination of competition, only that it be reduced in intensity to the point where coexistence is possible. The experiments in which competition has been demonstrated have shown that it could still be detected. There have been rela-

tively few examples in which even an approximately complete set of experiments has been carried out on all members of a proposed guild.

When predation is a relevant interaction, either with or without competition, the term "food web" is satisfactory; unlike "community," it is a term that has carried the same meaning for many years. A theory of food webs is being developed (Cohen 1978; Pimm 1982), but there is serious question that it will provide predictions of the outcomes of experiments, for a number of reasons explained by Paine (1988). It is true that both predation and competition have been demonstrated experimentally in some associations of species, but those phenomena were looked for independent of each other and independent of any unifying theory. One problem with current food-web theory is that it does not provide for competition at the lowest trophic level – an obvious shortcoming in most terrestrial ecosystems. Quantification of food-web relationships is a need that has caused difficulties in the construction of theory and has forced theoreticians to use qualitative relationships. The difficulty should be overcome by properly designed experiments.

Parasites constitute a different category, but one that is overlapping with predators. Mathematical models of their populations and those of their hosts have produced a variety of results. In some models, very regular cycles are predicted, a prediction that has been verified for some human viral diseases (Anderson & May 1979). Other kinds of parasites, with life cycles differing from those of viruses, carry quite different predictions (May 1985; Toft 1986). The rarity with which populations fluctuate cyclically in nature suggests either that they are unaffected by their parasites or that the important parasites of natural populations have some of the complex properties discussed by Toft (1986) or by Holmes and Price (1980). One example of an experimental test of the influence of a metazoan parasite on the population of its host is provided by the red grouse, *Lagopus lagopus scoticus*, and its nematode parasite *Trichostrongylus tenuis*. The parasite has been implicated in the long-term cycle of its host (A. Dobson, pers. commun.).

The natural history of mutualistic relationships has always fascinated ecologists. Only recently has mutualism been investigated quantitatively [see Addicott (1986) for a review]; its consequences for the structure and dynamics of guilds and communities are largely unknown (Schoener 1986). I predict that experimentation with mutualists will prove to be more difficult than has experimentation with

competitors, whereas predator-prey systems will be intermediate in difficulty.

How important are competition, predation, and mutualism?

This question cannot be answered at present because of the uneven attention that has been paid historically to the three kinds of interaction. Competition has received the most attention, probably because of its central position in community theory. Sih et al. (1985) attempted to address this question, but their analysis misinterpreted the experiments, as far as the hypotheses of Hairston et al. (1960) are concerned, and the result is unsatisfactory. As already stated, mutualism lags far behind as a subject of experimentation.

Progress will be made in studying the relative importance of the different kinds of interaction only when experiments are planned to answer the specific question. My experiments on the salamander genus *Desmognathus* show that it is possible to obtain results in systems in which both predation and competition are operating (Hairston 1986), and the earlier work of Paine (1966, 1974) and Menge (1972) had done the same for the rocky intertidal habitat.

Do interspecific interactions affect larger ecological units than guilds?

Guilds are defined in such a way that competition is all-important within them, and although it is possible to imagine competition affecting communities of a variety of taxonomic groups, predation will be the interaction of interest where more than one trophic level is involved. Experiments have been conducted in the rocky intertidal, where it is difficult to confine the definition of a guild, a difficulty that appears to be due to the nature of the limiting resource: space on a solid substrate. The papers by Yodzis (1986) and Roughgarden (1986) analyze this situation. In terrestrial assemblages larger than guilds, there have been no attempts to separate the interspecific interactions discussed earlier from what are called "ecosystem processes" – production, consumption, and decomposition. Freshwater habitats, in contrast, have yielded important results on the interactions within entire communities. The experiments of Neill (1981, 1984), Neill and Peacock (1980), Schindler (1974), Hall, Cooper, and Werner (1970), and others are described in Chapter 7.

Do interspecific interactions determine the composition of multispecies associations, or do these mostly reflect adaptations of individual species?

This question asks whether guilds and communities are "real" or are the products of our imagination. A negative answer to the former question, or a positive answer to the latter does not necessarily imply that species interactions are of trivial importance in ecology. That would mean that we would have no a priori means of estimating what array of species would have detectable interactions. Thus, a guild would be defined as a group of species among which competition has been demonstrated. There would be no means by which we could choose a set of species before the necessary experiments were conducted. Thus, sophisticated theories could be constructed to explain the mechanisms by which such guilds function and the abundances of the species involved, but it would not be possible to predict anything about guilds not yet identified. Community ecology would be in the position it has long occupied, that of writing theory to explain what is already known. The composition of many groups "of taxonomically related species that presumably are competing for a common resource" would be at least partly determined not by the strength of interactions among them but by the adaptations of the individual species to the physical environment.

What is the expected effect of any of these ecological forces in determining the direction of natural selection?

It is frequently stated that evolutionary questions are not approachable by field experiments, despite the fact that such experiments have been carried out. It is true that it is necessary to grant a basic assumption before accepting the result of such an experiment. That assumption is that the current ecological conditions will continue to provide the selective pressures that presumably led to adaptive differences between related populations. If the differences relate to the physical environment, physiological experiments will suffice to confirm adaptive differences. Where interspecific interactions are involved, it is much more difficult to meet the necessary conditions than it is where the physical environment provides the selective pressure, because proper experiments are a prerequisite for demonstrating the existence of the interactions. The whole question of demonstrating

natural selection under field conditions has recently been treated at length by Endler (1986), and it should be noted that one experiment, described in less than a page of text, constitutes his treatment of field experiments.

How would the effects of these forces differ?

The answer to this question is likely to be highly specific to the individual situation. Theoretical work in both population biology and evolutionary ecology on the effects of competition and predation (Roughgarden 1979) has given fairly straightforward distinctions of the expected differences, but Dayton (1973) showed how interpretation could change when observations were followed by experimental analysis, and my experiments with *Desmognathus* showed that what had been interpreted as the result of competition was largely the result of predation (Hairston 1986). The topic is in need of much more critical work before any generalizations can be made.

A discussion of the ways proposed to solve the problems

The questions posed earlier are much the same as those posed throughout the history of ecology. It is true that the relative emphasis placed on the different questions has shifted over the years, with recent work concentrating on interspecific interactions. There have been dramatic changes in the methods used to attack the problems. During the past decade, the use of field experiments has increased exponentially (Schoener 1983), and a longer period beginning in the mid-1960s saw the dominance of mathematical models, including the introduction of at least two new journals to handle the flood of papers (Simberloff 1983). A review of the recent literature shows that all of the methods to be discussed next are actively used at present. There is, of course, variation in emphasis among ecological journals, depending on editorial policies and the tastes of contributors.

Field observations

Ecological problems are recognized through observations made in the field, and therefore such observations are vital to the science. To become preoccupied with theory or with methodology to the exclusion of looking at nature is surely to lose contact with the world

that we are trying to understand. Such preoccupation always leads to oversimplification and thus to a limited vision of how nature works. It is true, of course, that some simplification is involved in any explanation of a natural phenomenon, and if we do not seek explanations, we can scarcely be said to have a science. Therefore, most publications of ecological observations are accompanied by interpretations. The emphasis in this book is to question the extent to which natural situations can be used to interpret the processes that brought them about. Interpretations of natural situations may be correct, but frequently they have been shown to be incorrect on subsequent experimentation. Despite the popularity of direct interpretation of ecological observations throughout the history of ecology and its advocacy by prominent ecologists, there are strong theoretical objections to this use of weak inference (Peters 1976); consequently, ecological experiments have become popular, even though they are "labor intensive, time-consuming and expensive" (Rosenzweig et al. 1985, p.194).

The members of one subset of direct observations have been called "natural experiments," and their importance in interpreting ecological processes has been defended vigorously by Diamond (1986). I argue that to label natural events experiments evades the issue of why manipulative experiments are conducted, which is first to test the validity of a specific idea, and second to avoid the charge of a posteriori reasoning.

There is no question that some situations and events in nature are especially favorable for constructing hypotheses about how the compositions of certain communities were determined and about how they function. This is especially true when a significant spontaneous event occurs after the start of a study, providing the very important initial conditions: those existing before the "experiment" began. Such a phenomenon provided the basis for the size-selective predation hypothesis of Brooks and Dodson (1965). After observations had been made on the plankton community in a small lake in Connecticut, it was colonized by a planktivorous fish, *Alosa pseudoharengus*. The subsequent difference in species composition among the planktonic crustaceans stimulated Brooks and Dodson to propose their hypothesis. Even though such events increase the likelihood that their interpretations are correct, that does stretch the concept of an experiment, and it exaggerates their significance to give them that name. Nature has no stake in being understood by us, and the so-called natural experiments universally lack important elements of experimental design, topics deferred to Chapter 2.

Mathematically derived theory

Direct interpretations of nature have been bolstered in recent decades by the application of mathematics, thus giving the appearance of scientific rigor to what in principle is a more sophisticated version of the same process of explaining what has been observed. The approach has been to think how nature might work and to formulate a mathematical model of the process. This has the merit of requiring specific hypotheses and has the potential of making precise predictions. Models force their creators and users to think clearly. They also contain two inherent weaknesses. The first is that mentioned earlier: explaining what is already known. The second is their incorporation of assumptions that frequently are unjustified. This point has been made repeatedly, especially with respect to the most widely used ecological model, the logistic equation [see discussions by Smith (1952) and Andrewartha & Birch (1954)]. It should be appreciated that much subsequent mathematical theory, such as the Lotka-Volterra competition equations, the community matrix (Levins 1968) and niche partitioning theory (MacArthur & Levins 1967; Roughgarden, 1976), is based on the same assumptions contained in the logistic equation. We should not be surprised if experiments have falsified various parts of these theories. The example of Dayton (1973) has already been cited. What are falsified usually are the assumptions on which the theory is based. Deliberate verification of the assumptions has been attempted in laboratory experiments, such as those of Gause (1934), Crombie (1945), and Vandermeer (1969), but nearly all field experiments involve conditions that are too daunting for more than an overall result. For example, my experiments on interspecific competition in the salamander genus *Plethodon* (Hairston 1980a) revealed that mutual negative influences were involved, but the assumptions on which the calculations of the coefficients of competition were based were not verified.

An even more invidious aspect of mathematical models is that unacceptable assumptions can be contained within the mathematics, as opposed to those made in constructing the theory. Those made in constructing the model usually can be identified, with sufficient effort, by nonmathematical ecologists, but the assumptions buried within the mathematics require the attention of other mathematical ecologists. The best-known case is one involving the mathematically derived conclusion of May and MacArthur (1972) and May (1973) that the original form of the MacArthur-Levins (1967) approach to limiting similarity

was robust to a considerable amount of stochastic variability. That apparently got rid of a most restrictive assumption in the original model, namely, that abundances in the community are completely determined by the coefficients of competition. Relaxation of that assumption yielded important implications about overlap in resource use and the sizes of carrying capacities of competing species that permitted coexistence of competitors. Unfortunately for this apparently encouraging development, Turelli (1978) found that some assumptions implicit in May's mathematical treatment imposed very restrictive conditions on the form of the model. That discovery severely limits the applicability of May's conclusion. The message for all ecologists, whether mathematically inclined or not, is to insist on clear, well-informed statements of the assumptions behind any theoretical models of how natural communities work.

Planned experiments

Experiments are conducted for the purpose of answering specific questions about nature. These questions ordinarily are stated as hypotheses, which are statements about how someone thinks nature works. In other words, they contain implied predictions, and confirmation of those predictions is the most powerful means available to demonstrate the accuracy of our understanding of the world around us. In the foregoing sections of this chapter, I have stated my reservations about other approaches to understanding ecological processes. Experimentation must come under the same kind of scrutiny. How does it propose to test hypotheses?

Laboratory experiments. Experimentation is most easily repeatable, and gives the most convincing results, when all the variables are under the control of the investigator. Then, one of the variables is manipulated in a systematic way in order to answer a specific question, such as this:

What is the effect of temperature on the intrinsic rate of increase of this species?

In conducting such an experiment, one should minimize variations in such factors as the duration and intensity of periods of light and the amount of moisture available, or else incorporate them in the experimental design. The ability to hold all such conditions constant, or to vary them in a systematic way, requires that the experiment be under-

taken in the laboratory. In subsequent chapters, some illuminating laboratory experiments will be described. With sufficient care, laboratory experiments should, in principle, be repeatable – a hallmark of a convincing scientific test. Even a rigorous ecological experiment in the laboratory is accompanied by the unsettling question of its applicability in nature, and controversies on this point have been severe (see the discussions of the papers by Nicholson, Birch, and Andrewartha in the Cold Spring Harbor Symposium on Quantitative Biology, 1958).

Field experiments. Under field conditions, it is not possible to exercise control over most variables. Usually, the variable of interest is manipulated in a predetermined way, and the remaining variables fluctuate independent of the experiment. Occasionally, more than one variable can be manipulated, but the number is severely limited. The questions asked in field experiments tend to be less precise than those asked in the laboratory:

Do these two species compete in nature, or, more properly for most such experiments, do they affect each other adversely?

With good luck, one can go one step further and ask this:

Under which of these conditions do the two species compete more strongly?

It is assumed that the naturally varying factors affect all experimental treatments equally, or at least randomly. The assumption introduces an element of uncertainty into the interpretation of the outcome, and reduces the assurance of being able to repeat the result. An example from exceptionally well-designed experiments is the work of Morin (1983a), Wilbur, Morin, and Harris (1983), and Morin, Wilbur, and Harris (1983). Those experiments were conducted out-of-doors under conditions that can be called seminatural or semilaboratory, depending on one's point of view, because the experimenters had control over many aspects of the conditions in the cattle tanks where the experiments took place. That fact makes the example especially pertinent. Morin conducted experiments on the effects of different levels of predation by the newt *Notophthalmus viridescens* on competitive communities of frog tadpoles of six species. The competitive dominant, *Scaphiopus holbrooki*, was also a favored food of the newt, and its relative abundance decreased with increased predation. In a similar but more elaborate experiment by Wilbur et al. (1983) and Morin et al. (1983) with four species of frog tadpoles, three of them the same as those used by Morin (1983a), *Scaphiopus* failed to dominate the com-

munities without predatory newts. The reason was an unexpected bloom of filamentous green algae that appeared because a spring drought held back the breeding of some of the species of frogs, leading to a change in the conditions in the tanks. The filamentous algae were shown experimentally to be unsuitable as food for *Scaphiopus*. Thus, the intrusion of an uncontrolled environmental factor prevented the original result from being repeated.

Some well-known ecological hypotheses and their origins

Anyone who has any curiosity is likely to have ideas about the processes going on in nature. These ideas come from having a sufficient amount of factual information, plus, especially, having the ability to visualize a problem that requires a solution. The solution of any ecological problem should begin with the formulation of one or more hypotheses about how nature works. Only rarely do authors call their own thoughts hypotheses. More frequently, they call them theories, models, conclusions, or even proofs. The appellation "hypothesis" is given by other, more skeptical ecologists.

This section describes three famous (or infamous, depending on one's point of view) hypotheses. The descriptions include the basis, factual and logical, of each hypothesis.

Santa Rosalia

The first hypothesis we shall consider is contained in G. E. Hutchinson's well-known presidential address to the American Society of Naturalists (Hutchinson 1959): "Homage to Santa Rosalia or Why are there so many kinds of animals?" Imaginative perception was required to see that that was an interesting problem — one for which the solution would reveal a great deal about ecological processes.

Hutchinson's reasoning is as follows: There are as many as 10 million species of animals on earth, and yet as far as the physicochemical environment is concerned, the number of different sets of conditions to which species can be adapted is fairly limited, especially when we consider the range of conditions that most ordinary species are able to tolerate. His next possible candidate to account for diversity is food chains. Energy considerations will limit the number of links, because the efficiency of conversion of energy from one link to the next higher link is less than 0.2, and because predators (he assumes) are, on the

average, twice the mass of their prey. That will mean that after five links, the population of the fifth link will be 10^{-4} that of the first, and after fifty links, 10^{-9} . Even starting with algal cells at the first link, and at a density of 10^6 per milliliter, the volume needed to hold one individual of the fiftieth link would be much greater than all of the oceans combined. From this approach, the limit in total species is small compared with the millions being sought. Hutchinson considers whether or not there are enough different food chains, at five links each. He points out that many herbivore species are selective in their food plants, and hence each plant species theoretically could support a food chain. With 200,000 species of terrestrial producers, a large number of animal species can be accounted for. This ignores the fact that predators are not selective, and there would be much overlap between food chains. But to proceed with the reasoning: Why are there so many species of plants? He suggests that the same principles apply as to animals. The following principles are then enunciated:

1. Diversity enhances stability. This is stated to be intuitively satisfying. In support of the statement, Hutchinson first cites prominent ecologists, such as Elton, and then claims that there is formal proof by MacArthur (1955).
2. Diversity enhances stability, because when species are added to a community, that involves taking over part of the niche of a species already present. That reduces fluctuations in the population of the latter and keeps it from "being underrepresented to a dangerous degree" (Hutchinson 1959, p. 150). No evidence is given that fluctuations are less dangerous for small populations. To the contrary, small populations are vulnerable to practically all hazards, including especially fluctuations, which can carry a small population to extinction. Niche partitioning, the means whereby diversification is achieved, depends on the hypothesis that by sharing overlapping niches, the average population is reduced, and *therefore* fluctuations are reduced.

The cornerstone of the reasoning is an appeal to "modern ecological theory," which claims that communities of many species persist better than those containing fewer, less diversified organisms. The theory, supplied by MacArthur's paper, was a popular one that appealed to the intuitive desire for a reason for preserving diversity in the face of habitat destruction and monoculture. MacArthur's paper itself, however, points out that the model is based on several unrealistic simplifica-

tions, some of which are unrealistic enough to matter (Hairston et al. 1968). There is, for example, the implicit circularity that both diversity and stability depend ultimately on the number of species present, and there is the possibility, now generally accepted, that cause and effect are reversed. Stability enhances diversity. Subsequently, May (1972) showed that interactions among a large number of species would tend to decrease the stability of the system, but it would, of course, be unfair to expect Hutchinson to anticipate that result.

Having stated the basis for his hypothesis, Hutchinson considers the limits to which the proposed process can go. There is a "problem as to how much difference between two species at the same level is needed to prevent them from occupying the same niche" (Hutchinson 1959, p. 152). He suggests that this difference can be measured by size, and he gives some data supporting a ratio of about 1.3 in a linear dimension of feeding structures as being sufficient to permit coexistence. Community composition was thus firmly based on the adaptation of the species to competitive relationships among them.

It would be difficult to exaggerate the influence of Hutchinson's paper. It stimulated a large number of papers showing confirmations of the ratio and the existence of other regular ecological differences among coexisting species [see the references in Roth (1981) and Simberloff & Boecklen (1981)]; it also provided the theoretical basis for most of the theory of interspecific relationships during the next fifteen years, as seen in the work of MacArthur (1960, 1962, 1968, 1969, 1970), MacArthur and Levins (1967), Levins (1968), Levins and Culver (1971), May (1973), May and MacArthur (1972), May and Leonard (1975), Roughgarden (1972, 1974, 1976), and Schoener (1974, 1976). There was a great desire for coherent and precise theory among ecologists at the time, and Hutchinson and his intellectual companions and descendants supplied it. Small wonder that most of the tests of the hypothesis were actually searches for confirmatory evidence.

It was some time before weaknesses in the original hypothesis and in the theoretical edifice built on it were appreciated. Eventually, the importance of predation became apparent, following the experiments of Paine (1966, 1974), Connell (1970, 1975) and others, and an increasingly skeptical attitude (Simberloff 1970, 1974, 1980, 1983; Simberloff & Connor 1981; Connor & Simberloff 1978, 1979) has led to a situation where a rethinking of much ecological theory is needed.

Hairston, Smith, and Slobodkin (HSS)

The second hypothesis to be considered has been controversial for several reasons. It was published shortly after Hutchinson's paper, and the authors were not aware of that publication. Hairston, Smith, and Slobodkin (1960) originally submitted their paper to the editor of *Ecology* in May 1959. Its rejection and subsequent submission to another journal caused a year's delay in its publication. The background for the paper was the debate over density dependence, and the paper originated in an argument among the authors about the evidence favoring competition for resources. There was agreement that "food not consumed by herbivores or carnivores is ultimately used by reducers . . . and that in any balanced state, such as we assume communities to be, the food is completely consumed by some species or other" (Hairston 1959, p. 414). We thus believed that there was good evidence for density-dependent resource limitation among the decomposers, and hence competition between the species of bacteria and fungi.

The vexing situation, as far as we were concerned, was the following: Inasmuch as nearly all of terrestrial primary production falls to the ground uneaten, it is difficult to visualize how there can be any competition for it, either intraspecifically or interspecifically. That was a point that L. C. Birch had made forcefully; yet we could not accept his explanation, which was that weather was severe often enough to keep the populations of terrestrial herbivores well below the level set by their food supply. The solution to the dilemma came when we started to consider exceptions to the observation that only a small proportion of green plant production is consumed by herbivores. The most frequent exceptions are outbreaks of introduced species that, though less clearly adapted to the local climate than native species, manage to become numerous enough to defoliate the vegetation. The explanation at which we arrived was that the introduced species had left a large number of their predators and parasites behind, and from that we concluded that the natural state of affairs is for terrestrial herbivores as a group to be preyed on at a sufficiently high rate that their numbers never reach a level where they consume much of the dominant vegetation.

It followed from that explanation that terrestrial predators (including parasites) as a group are limited in abundance by the supply of herbivores, because the latter constitute most of their food (some, of

course, prey on each other). The full set of HSS hypotheses was as follows (Hairston et al. 1960). After the first one, the others were stated to apply to terrestrial situations, and we specifically excluded successional stages, as well as aquatic systems.

1. Decomposers as a group must be food-limited. (p. 421)
2. Producers are neither herbivore-limited nor catastrophe-limited and must therefore be limited by their own exhaustion of a resource. (p. 422)
3. The usual condition is for populations of herbivores *not* to be limited by their food supply. (p. 422)
4. Although rigorous proof that herbivores are generally controlled by predation is lacking, supporting evidence is available. (p. 423)
5. The predators and parasites, in controlling the populations of herbivores, must thereby limit their own resources, and as a group they must be food-limited. (p. 423)
6. Therefore, interspecific competition for resources exists among producers, among carnivores, and among decomposers. (p. 423)

The paper has been controversial. It is opposed, for example, by advocates of density independence (Ehrlich & Birch 1967), advocates of the universality of competition in structuring communities (Gill 1981), advocates of the ecological importance of herbivore-resisting compounds (Murdoch 1966, Janzen 1969), and advocates of selection for ecosystemwide traits [Patten and Odum (1981) do not specifically cite our paper, but they argue strongly for group selection among ecosystems, and they object to hypotheses that require only Darwinian selection]. In addition to those objections, Murdoch (1966) claimed that the hypotheses were untestable and therefore unscientific. In our response (Slobodkin, Smith, & Hairston 1967), we listed means whereby each hypothesis could be falsified.

Despite the widespread opposition, the interrelated set of hypotheses appeals to enough different ecologists for discussion to continue; see, for example, Fretwell (1977, 1987), Strong (1982), Strong, Lawton, and Southwood (1984a), Menge et al. (1986a), and especially Oksanen (1988). Both Schoener (1983) and Connell (1983) included analyses of the frequency with which individual experiments did or did not confirm the existence of competition among herbivores. Such a "majority vote" was specifically excluded in our clarification (Slobodkin et al. 1967): "We were not making statements about most herbivores, or most carnivores, but about these trophic levels as wholes. Our statements, then, apply to the quantitatively dominant species but not necessarily to the numerical majority of species in any ecosystem"

(p. 109). The analyses of the literature could not not have falsified the hypothesis, but could have provided incidental support, as was the case with Schoener's analysis. That was the commonest misinterpretation, a second one being a failure to note our exclusion of all "herbivores" except those that consume the foliage or sap. Specifically, the hypothesis does not apply to seed eaters, nectar feeders, or pollen feeders.

The HSS hypotheses have recently aroused some interest; their eventual influence cannot be predicted. Perhaps, when the workings of different kinds of ecosystems are considered separately, terrestrial ecology will be shown to have been influenced.

The broken stick

The final hypothesis to be considered is an example of one that has shown extraordinary viability in the face of negative evidence regarding its validity. The background of the "broken stick" model is in the large amount of data on the relative abundances of species and in attempts to interpret the observation that when comprehensive collections of any taxonomic group of organisms are made in a specified area, a common pattern of relative abundances is revealed: A few species are quite abundant, and many species are rare, and that pattern is found in nearly all groups for which adequate data exist (Fig. 1.1). The first attempts to interpret this relationship involved trying to fit the relative abundance distributions to mathematically defined statistical formulae. The aim was to find a common formula, so that the diversities of different communities could be compared (Fisher, Corbet & Williams 1943; Preston 1948). It was hoped that an "index of diversity" would allow investigators to assess the relationship between the number of species and the number of individuals for different situations, specifying the relative "richness" of the fauna. The match of Fisher's logarithmic series to the data is shown in Figure 1.1; Preston's lognormal match of the same data is shown in Figure 1.2.

In 1957, MacArthur took a different and intuitively more satisfying approach to the regularity that the data showed. His idea was to consider how the community might be organized in several different ways, construct a mathematical model for each of the imaginary organizations, and compare them to real data to permit a choice among the models. Imagine that the total number of individuals has some correspondence to the total available niche space. The species can be

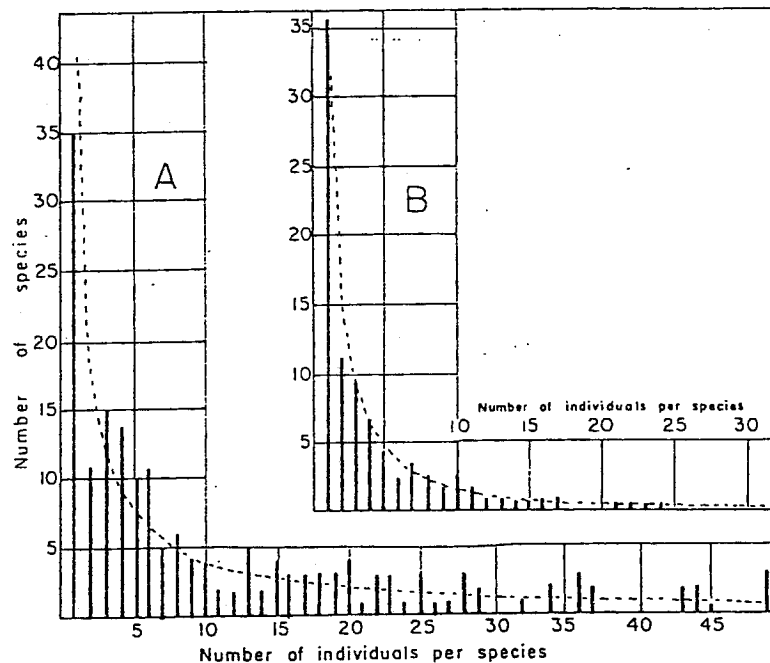


Figure 1.1. Species abundance distributions of Lepidoptera caught in a light trap at Rothamsted, England. The graphs show the number of species represented by different numbers of individuals. The broken lines were calculated from Fisher's logarithmic series. (From Fisher, Corbet, & Williams 1943.)

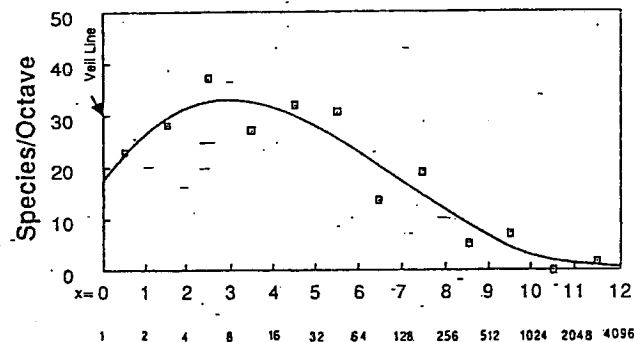


Figure 1.2. The same data as in Figure 1.1, with the abscissa on a scale of logarithms to the base 2, the successive intervals of which are called octaves. The line was calculated from Preston's lognormal distribution. (From Preston 1948.)

thought of as dividing up the niche space in several different ways. First, suppose that they divide the niche space in a nonoverlapping way and that the relative sizes of the different niches are determined by so many factors that the result is a random array of abundances. Such a nonoverlapping arrangement would be arrived at through competitive interactions among the species, each acquiring its part of the niche space by being superior there. MacArthur's image was of a line on which points are thrown at random until the number of segments between points is equal to the number of species involved. A second model, in which the abundances were allowed to overlap by throwing pairs of points at the line, was claimed to represent some common overlapping use of niche space by the different species, and a third model assumed independently acquired abundances. MacArthur (1957, 1960) found some data on birds that fit his first model and suggested that a commonly observed form of the relative abundance data, in which common species are too common and there are too many rare species, was due to the pooling of data from heterogeneous sources. That was shown not to be the case (Hairston 1959, 1964, 1969), as pooling data brings about a closer fit to the first model. The second model was incorrect mathematically (Pielou & Arnason 1966; Vandermeer & MacArthur 1966) and the third was never mentioned after the original paper.

The first model continues to fascinate some ecologists, despite the demonstration by Cohen (1968) that the mathematical formula could be derived from premises other than the original competition-based premises of MacArthur, and despite MacArthur's own public abandonment of the whole approach (MacArthur 1966). Perhaps it is the universal regularity of the numerical relationships of the abundances of coexisting species that is so fascinating, but it is more likely that the original approach of MacArthur in starting from biological relationships, rather than the statistical distribution, that continues to hold the attention. Some ecologists continue to cling to the model, ignoring its obvious failures (De Vita 1979; Sugihara 1980; Bush & Holmes 1983; May 1986).

Synthesis

This introductory chapter is intended to place ecological experimentation into the broader context of the whole field, including such issues as why we need experiments, what sorts of questions can

be answered by good experiments, how the questions have been asked in the past, and how hypotheses have been formulated in proposing answers to some of the questions.

The remainder of this book consists of a discussion of scientific requirements in ecological research, some of the special problems that ecologists face in meeting the requirements, and critiques of a number of ecological experiments in each of the environments in which they have been conducted. This is not an attempt to discuss all ecological experiments, a task that would require an encyclopedia. The experiments selected for treatment were chosen for many reasons: because they can be used to illustrate difficulties in the implementation of experiments in the field, because they illustrate failures to follow acceptable design, because they illustrate outstanding scientific procedure, because they illustrate especially important ecological phenomena, or because they are necessary to an understanding of the habitat in which they were performed.

2

Minimal requirements of experimental design in ecology

Introduction

Ecological experiments come in many forms, and at all levels of sophistication. At the simplest, they involve an ecologist thinking that something interesting is happening, and deciding to change the system to find out more about possible causes. If the change is followed by a dramatic "result," the curious ecologist may decide that the case is proved, and write the experiment up for publication. The change in the system would be classified as an experiment, but there would be problems in accepting the conclusion, unless certain precautions had been observed. There is the question whether or not a rare or even unique event has been observed. Can the result be repeated? Very few ecological field experiments have been repeated, but the use of a satisfactory experimental design can remove at least some of the uncertainty. Our ecologist is faced with a choice, either to carry out a simple manipulation to satisfy a perhaps uncritical curiosity or to invest more work and conduct an experiment that will convince the array of scientific colleagues.

It might seem superfluous to describe the requirements of experimental design, but they have been violated regularly enough in published works that all should be warned of errors. It is possible to do anything badly, even something as desirable as an experimental approach to ecology. One of the features that distinguish science from other approaches to understanding the world is the set of rules by which scientists operate. As best they can, the rules prevent scientists from yielding to temptation, and it is generally accepted that, in principle, our colleagues will catch us if we do not follow the rules. The most formal and widely accepted of the rules are those contained in the requirements of experimental design. The more complex the experiment, the more complex the requirements, but the minimum requirements are a knowledge of the initial conditions, adequate controls, and replication. These are explained next, and the explanation is followed

by two examples of good experiments. Finally, an example of an experiment in which all of the rules were violated is presented.

Initial conditions

Experiments should be designed to test hypotheses, and a satisfactory hypothesis must, at the very least, contain one or more overt or implied predictions, the role of which has been mentioned in Chapter 1. Our curious ecologist would not perform the simplest manipulation if there were no expectation of some result, however nebulous, and that expectation constitutes an implied prediction.

A prediction requires a clear statement of the conditions that must exist for it to be confirmed. Therefore, to carry out a satisfactory experiment, one must have full knowledge of the conditions existing before the experiment is begun. Otherwise, there is a limited assurance that the hypothesis is being tested properly. This would not appear to be a difficult requirement to meet in ecological research, as it is the natural situation that leads to the hypothesis. The commonest failing in this regard is inadequate description, a problem that can be avoided by a sufficient period of taking baseline data – following populations in locations to be used for controls and experimentals before any manipulations are begun. As one example of the use to which such preliminary information can be put, I have selected the work of Hurlbert and Mulla (1981). They planned to test the effect of the mosquitofish, *Gambusia affinis*, on the plankton community in small artificial ponds. Even though the ponds had been specially constructed as replicates, the baseline data of Hurlbert and Mulla showed that the variability in population densities was great. They used that information to assign ponds to treatments so that each treatment received the full range of densities, thus avoiding chance assignment of high densities to one treatment and low densities to another. The choice between random assignment (the hypothetical ideal) and deliberate stratification is one that is not easily settled (Hurlbert 1984).

In the process of making observations on initial conditions, it is inevitable that variation will be encountered. This variation imposes a decision on what part of the environment will actually be used in the proposed experiment. There is a trade-off between including a wide range of conditions, so as to make any conclusions general, and using a restricted set of conditions, to avoid so much variability that analysis of the results yields no conclusions, or at best weak conclusions. The

principle to be followed is not to become so enmeshed in experimental design that common sense cannot dictate the choice. A knowledge of the natural history of the organisms involved is absolutely essential in making the choice.

The requirement of adequate knowledge of initial conditions has important implications for the validity of so-called natural experiments. Inasmuch as those "experiments" are recognized only when they are completed, or in progress at the earliest, it is impossible to be certain of the conditions that existed before such an experiment began. It then becomes necessary to make assumptions about those conditions, and any conclusions reached on the basis of "natural experiments" are thereby weakened to the point of being hypotheses, and they should be stated as such.

Controls

The necessity for controls is obvious, but there are examples in the ecological literature of experiments without them, and there are more examples of inadequate controls. In ecological experiments, there are so many factors that must be allowed to fluctuate naturally that it is essential to know what would have happened if the experimental manipulation had not been performed. A simple example will demonstrate the necessity. The example is of my own experiment testing for the existence and strength of interspecific competition between two species of terrestrial salamanders in the Great Smoky Mountains (Hairston 1980a). On the basis of the altitudinal distributions of *Plethodon jordani* and *P. glutinosus* it was originally postulated that they were in strong competition (Hairston 1951), and experiments were subsequently proposed to test the hypothesis. These experiments would involve the removal of each species (from separate plots), with the expectation that the other would respond favorably (Hairston 1973). The expectation was met, the population of *P. jordani* increasing on the experimental plots during the first three years after the start of the removal of *P. glutinosus* (Fig. 2.1). The population on the control plots, however, showed an almost identical increase during the same three years. The elementary message is that without the controls, a positive result on the density of *P. jordani* might mistakenly have been claimed. As an aside, the means constitute temporal pseudoreplication, as discussed later. That is, they imply more independent observations than were available. In the statistical analysis of this experi-

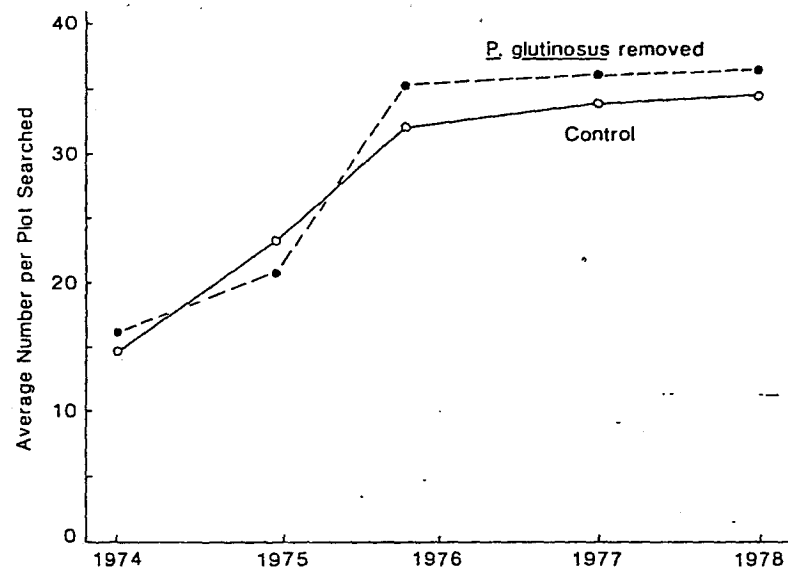


Figure 2.1. Annual average numbers of *Plethodon jordani* per plot search on plots from which *P. glutinosus* were removed (broken line) and on control plots (solid line). Data from the experiment in the Great Smoky Mountains.

ment, the problem was resolved by a procedure that avoids the issue (Hairston 1980a,b).

It is obvious that control areas should be as much like experimental areas as is possible. This usually can be established by preliminary study, such as repeated sampling of all areas to be included in the experiment. The analysis of these samples will also permit estimation of the number of replicates that may be needed, as discussed later. The inability to obtain this information in "natural experiments" further weakens conclusions about their results.

In many complex experiments it is necessary to add controls for specific effects beyond observing the natural situation. In an experiment related to the simple removals described earlier, it was necessary to transfer specimens of *P. jordani* from one mountain range to another in the southern Appalachians. The purpose was to test for the effect of the introduced form on the local population of *P. glutinosus*, when it was substituted for the local *P. jordani*. An additional control was required, in order to confirm that the introduced *P. jordani* could sur-

vive and reproduce in the new area, in the absence of competition from either native form. Thus, more plots were established, from which both local species of *Plethodon* were removed, while introducing *P. jordani* from the distant mountain range (Hairston 1980b, 1983). The introduced salamanders survived on the control plots, thus validating the results of the experiment. In experiments using enclosures or exclosures, more elaborate controls may be necessary, in order to check for the effect of the manipulation on the outcome of the experiments.

Replication

Variability is all-pervasive in nature, and it is manifestly impossible to find identical locations for the different treatments to be employed in a field experiment. Therefore, in order to avoid possible unrecognized systematic differences among locations, it is necessary to assign locations to treatments randomly; that is, each time an assignment is made to a treatment, all locations should have an equal chance of being assigned. This procedure has the additional merit of avoiding possible bias by the experimenter. The only problem can occur in small experiments with few replicates, in which case chance may segregate one treatment to neighboring locations. Deliberate interspersion will avoid this problem, but reduces the purity of any statistical inference on subsequent analysis of the results. This subject is treated in detail by Hurlbert (1984).

The natural variability must be taken into account in interpreting the results, and the only means of assessing variability is to have two or more representatives of each treatment, including the controls. Several early experiments on root competition in forests suffered from failure to provide replicates; see, for example, Tournay and Kienholz (1931). Despite the apparently dramatic results of cutting roots around one plot, a comparison of the replicates in one such experiment in a 31-year-old loblolly pine stand (Korstian & Coile 1938) demonstrates the necessity of having them. At the start of the experiment in 1932, the two experimental plots to be trenched to minimize competition from tree roots had 16 and 6 species, and 119 and 80 individual plants, respectively. The control plot for the first had 21 species and 65 individuals, and the control plot for the second had 5 species and 48 individuals. During the next four years, the flora of the trenched plots increased in species and numbers: from 16 to 24 species in the first

plot, and from 6 to 16 species in the second; from 119 to 741 individuals in the first plot, and from 80 to 1,213 individuals in the second. The two control plots decreased moderately in both number of species and number of individuals. The changes in species composition differed greatly on the experimental plots. On the first, the dominant species after four years, amounting to 26 percent of the total, was absent at the start; on the second, the dominant species at the start (90%) was also dominant at the end (82%). Thus, both the initial composition and the response to the same treatment varied greatly among replicates. It should be noted that the authors did provide sufficient details for this analysis to be made. The current dictates of space limitation frequently prevent the reader from obtaining such information.

If controls and experimentals were essentially alike at the start, a greater similarity among controls alone and among experimentals alone than exists between the groups at the end of an experiment is the first indication that the experimental manipulation has had an effect. The quantitative measure of those comparisons is the province of statistics. Although the subject is outside the scope of this book, it cannot be stressed too strongly that an understanding of the principles of statistics is absolutely essential before any kind of ecological experiment is undertaken. Throughout this book, statistical terms are used occasionally. A reader who is not familiar with statistics should nevertheless be able to understand the general meaning from the context. Knowledge of the exact technical meanings of the terms is not essential to understanding the point being made. Among the available array of texts, Sokal and Rohlf (1981) is authoritative, and Fisher (1960) and Siegel (1956) provide useful information in readily understandable form.

In planning an experiment, it is necessary to know how much replication to include. This depends on the amount of variability in the data; and correct assessment of the needed replication is a problem that can be solved only with the help of statistics. The estimate should be made during the period of establishing the initial conditions. It is quite possible that the amount of replication that is necessary is more than the time and other resources of the experimenter will allow. In that case, the planned experiment may have to be reduced in complexity, with reduced sophistication of the questions being asked. The trade-offs involved in such decisions are discussed in more detail in Chapter 3.

At this point, it is recognized that some large-scale ecological experiments are carried out under conditions in which replication cannot be achieved. A notable example is the experiment on whole lakes by Schindler (1974). His aerial photograph is thoroughly convincing that adding phosphate increases the phytoplankton biomass. Statistical "significance" was not determinable in that case, and was not attempted. There is nothing wrong with such experiments. What would be wrong would be a misleading attempt to use statistics to bolster the conclusion. The problem is discussed, with examples, by Stewart-Oaten, Murdoch, and Parker (1986).

Systematic variability: blocks

For some field experiments, it is known or strongly suspected that the locations of some parts of the experiment (plots) are different in an important way from others. Yet it may not be possible to locate sufficient replicates without encountering these differences. In such cases it is frequently possible to design the experiment so that all treatments, including controls, are represented in each kind of location. In my experiments on *Desmognathus* (Hairston 1986) it was necessary to place some plots along each of three streams. I worried that the streams might support different densities of the four species involved, and therefore I placed one plot of each of the three treatments on each stream, thereby establishing three blocks. The effect of any differences between streams could be removed from the analysis of results statistically, making for a more precise answer to the question of differences between treatments. Before starting the manipulations, there were five counts of the number of salamanders seen during night searches on each of the nine plots. For two of these sets of searches, the counts of *D. monticola* were as follows:

Search 2: The plots on stream A had 1, 4, and 0; those on stream B had 7, 5, and 7; those on stream C had 3, 2, and 1.

Search 4: The plots on stream A had 4, 3, and 2; those on stream B had 16, 4, and 5; those on stream C had 4, 6, and 8.

Data such as these permit an assessment of differences between blocks (streams, in this example), whether or not experimental manipulations

have taken place. The actual assessment is a problem in statistics, but for the present it is sufficient to state that one of the two sets of searches showed that there was less than one chance in twenty that the differences were due to variation between plots on individual streams. Such a probability level usually is accepted as "significant" statistically. It was the data from search 2 that showed this difference – not at all obvious from a simple inspection of the numbers. There were, however, no statistically significant differences among the plots that were destined to be parts of the different treatments – a result that undoubtedly came from placing one plot for each treatment on each stream. The statistical tests would have been impossible without more than one plot per stream. When the experiments were concluded, it was found that the significant difference between blocks occurred no more frequently over all searches than would be expected from chance. Thus, the original worry need not have affected the design of the experiment, but that conclusion could be made only after the fact, and the block design was appreciated by a critical reader.

The use of blocks partially achieves the interspersion of treatments that was mentioned earlier, and randomization within blocks further ensures this.

Systematic variability: stratification of data

Different well-defined parts of a population may differ much less within the parts than a random sample of the whole population would reveal. An obvious example is provided by many attributes of the sexes. If an ecological experiment is designed to test a hypothesis about growth, the sizes of individuals are important. Among many species the sexes differ in size, and the data must be kept separate, because the combined data would incorporate an unnecessary amount of variation, and because changes in sex ratio at different times during the experiment could easily produce misleading results.

Keeping the data separate is called stratification, and there are many ways in which different subgroups can be recognized and the data stratified accordingly. An example is treated in some detail later, with children of different ages having different levels of parasitization – a common observation for many parasites. Obviously, the different age classes must be examined separately for the data to be meaningful. Although the overall proportion positive is less easily interpreted than the stratified data, it is needed in some compilations; and for it to be

correct, the different age classes (strata) must be incorporated in proportion to their representation in the population.

Groups within a population may respond differently to an experimental manipulation. For example, in the experimental removal of *Plethodon glutinosus* described earlier, the mean densities of *P. jordani* on experimental and control plots did diverge in the last three years (Fig. 2.1), but the difference was not significant statistically because the differences between plots within treatments were great enough to obscure, in a statistical sense, the mean difference between treatments. However, the observed difference was entirely due to increases in the numbers of salamanders in the two youngest year classes, which created an interesting and statistically significant difference in the proportion of those young individuals. The obvious interpretation is that the removal of *P. glutinosus* permitted more reproductive output by the adult *P. jordani* than could be attained on the control plots. Stratifying the data by age classes revealed an important difference that was obscured in the gross population numbers. The implication of that result is that if the experiment had been continued for a sufficient number of years, the population means would also have been significantly different, as they were for the reciprocal experiment (see Chapter 4).

This raises the question of how long an experiment should be continued, or when one should be terminated. In the current example, a prior decision had been made that the experiment should last for the minimum duration of the life cycle of *P. glutinosus*, which has the longer generation of the two species involved. The duration of an experiment should be determined in advance, because of twin temptations: to stop when the results are pleasing, or to continue until they become so. Either procedure would violate proper scientific practice. The choice depends on the natural history of the organisms involved. Long-lived species should be given time to respond, as should those whose irregular pattern of reproductive success may dictate a long experiment so that bursts of successful reproduction will be included. For species in these two categories, brief experiments only rarely can give populational results; behavioral or physiological responses are the most one can expect. These were ignored by Bender, Case, and Gilpin (1984), who created an imaginary dichotomy between "pulse" and "press" experiments. In the real world, experiments can have any of an array of durations, with a variety of kinds of responses. Ecologists who are not thoroughly familiar with the organisms involved risk wasting a great deal of time.

Pitfalls in ecological experiments

Ecologists conducting experiments in the field, no matter how carefully the experiments have been planned, sometimes encounter unexpected difficulties either in the execution of the work or in the analysis of the results. These difficulties fall under several headings.

Pseudoreplication

Proper replication requires that replicates be independent of each other. That is, no measurement, count, or other observation on one replicate should have any influence on a similar observation on any other replicate. That is one of the most fundamental assumptions in the mathematics of statistical analysis, and yet there have been many studies in ecology in which that assumption has been violated (Hurlbert 1984). Hurlbert coined the term "pseudoreplication" to cover all of the various ways in which independence of replicates was not ensured in different studies.

One of the most obvious violations of the assumption of independence occurs when an experiment is continued over several sets of observations. Although the counts or measurements at any one time are truly independent of each other, the same counts or measurements on the same plots or other experimental units at successive times clearly are not independent, and must not be used as such in the statistical analysis of the experiment. The importance of the point lies in the fact that the statistical significance of an observed difference between the mean values of two treatments, say experimental and controls, depends in part on the number of replicates involved. Thus, inflating that number with nonindependent observations constitutes pseudoreplication.

One form of pseudoreplication occurs when plots or other replicates of a given treatment are placed together, and apart from the replicates of another treatment (or control). This may happen deliberately, as for the sake of convenience, or accidentally, as when random assignment of locations results in replicates not being interspersed with those of other treatments. In either case, the replicates are less than truly independent, and the statistical probability calculated from the data therefore is suspect. These points were discussed earlier, and by Hurlbert (1984).

The commonest form of pseudoreplication is the use of several samples within a treatment unit to calculate the statistical probability of the effect of the treatment. Thus, counts of the individuals in different samples from a tank, or from a plot, are not completely independent of each other, because the measures reflect both the treatment to which the unit has been subjected and the local characteristics of the plot. In principle, it is not possible to separate the two influences. The same problem applies to measurement of the number of individuals within a plot. The solution is not as simple as Hurlbert suggested, and to understand the controversy it is necessary to have enough statistical background to include a knowledge of correlation and analysis of variance (ANOVA).

Ecological experiments often include several, or even many, observations on each experimental unit to which a treatment has been applied. For example, an experiment on density dependence might have two cages containing ten individuals and two cages containing twenty individuals. At the end of the experiment, each survivor is weighed. How many observations of the effect of density on growth have been made? In a strict sense, the answer is one per survivor, perhaps as many as sixty, but if the individuals are competing they are influencing one another's growth, and these observations are not independent. Within a unit, individuals may have a negative covariance. In the extreme case of only two individuals in a unit, if one grows large, the other must be small, if they are limited by resources. On the other hand, suppose one of the replicates of the low-density treatment is in a favorable habitat and the other replicate is in a less favorable habitat. In this case, individuals in a unit are expected to have a positive covariance. Only in the case of zero covariance among individuals in a unit can all measurements be taken as independent estimates of the effect of treatment. When intraunit correlations are positive, the use of all individuals in tests of significance runs the risk of a type I error, rejecting the null hypothesis when it is in fact true. Standard statistical practice would be to use the mean response of individuals within an experimental unit. In this case, one would have only four observations, two on the effect of each treatment on growth. When intraunit correlations are negative (as expected when responses reflect competition), use of only means may be too conservative, and there is a risk of a type II error, that is, accepting the null hypothesis when it is in fact false. In this case, the use of individual observations is in fact justified.

The F ratio of the mean square between replicates divided by the mean square among individuals within units is a measure of the sign of the correlation. If $F > 1$, then one should use population means; if $F \leq 1$ then the use of individual observations is justified (H. M. Wilbur pers. commun.)

Unwanted effects of experimentation and observation

The requirements of experimental design or of the question being asked have, in some instances, caused experimenters to use techniques that have affected the outcomes of the experiments. The commonest of these are "cage effects." These were first observed when attempts were made to include or exclude predators or grazers from experiments in the rocky intertidal habitat. For example, Connell (1961a) suggested that cages inhibited the grazing potential of caged limpets, and Dayton (1971) reported that algae growing on cages prevented desiccation under them and improved the survival of predatory nemertean. An experiment on colonization by marine algae (Sousa 1979) yielded inconclusive results because cages designed to exclude herbivores became covered with algae and barnacles, preventing the establishment of algae within the cages. Shading and interference with the circulation of water or air have also been attributed to the presence of cages or even of roofless enclosures. Moreover, the use of fences to keep animals within experimental areas may well have the result of keeping their predators out or of selectively permitting flying predators access to the fenced area, while excluding earthbound predators. These effects are widely recognized at present as being undesirable properties of otherwise well-designed experiments.

It is commonly thought by ornithologists that mammal and snake predators follow the trails of investigators to nests that are visited regularly. The literature is less than unanimous. In two studies, one by Lenington (1979) on red-winged blackbirds (*Agelaius phoeniceus*) and one by Westmoreland and Best (1985) on mourning doves (*Zenaidura macroura*), evidence was presented that survival was reduced by extra predation on frequently visited nests. Willis (1973) recorded a slight increase in predation early in the nesting cycle of bicolor antbirds (*Gymnophthalmus bicolor*) as an apparent result of his visits, but Gottfried and Thompson (1978) could not detect significant extra predation as a result of their visits to Japanese quail eggs placed in nests in a planned experiment. Proper precautions appear to negate this effect of the

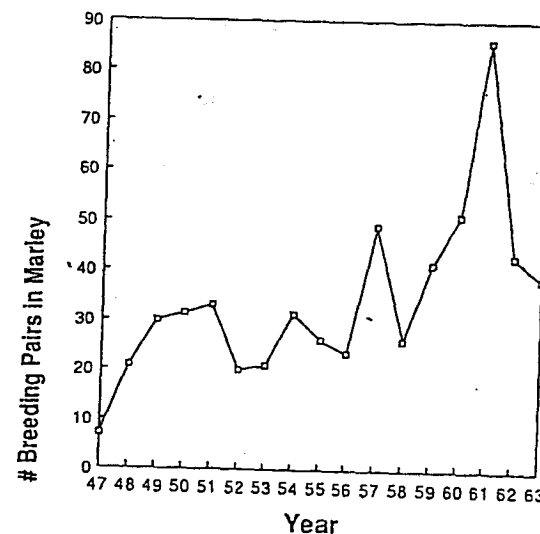


Figure 2.2. Fluctuations in the population of the great tit in Marley Wood, near Oxford, England. There was a surplus of nest boxes throughout the study. (From Perrins 1965.)

investigation itself (Nolan 1978). Another form of unintentional investigator effect on nesting success is documented in detail by Fetterolf (1983). He showed conclusively that routine study of the nests and chicks of ring-billed gulls (*Larus delawarensis*) caused significant decreases in number of eggs laid, hatching success, and fledging success. Those decreases were due to the effects of investigators' visits on the behavior of both adults and chicks.

Investigators working with hole-nesting birds sometimes place nest boxes in the habitat to facilitate the study. In the well-known long-term work on the great tit (*Parus major*) in Marley Wood, near Oxford, 200 nest boxes were put up in 1947 (Gibb 1950). The additional nest sites apparently shifted the limiting resource from nest holes to food and resulted in an increase in the population (see also Dhondt, Eyckerman, & Huble 1979). Although not mentioned by Perrins (1965) or Lack (1966), the change from a stable to an unstable limiting resource apparently resulted in an increase in the magnitude of fluctuations in the population (Fig. 2.2).

Trapping animals with bait can have several different unwanted effects on the outcome of the study. Some individuals become accus-

tomed to the food provided in the trap, and thus become "trap-happy" – preventing the trap from taking an unbiased sample of the population. Others have the opposite reaction and become "trap-shy." The effect of such a behavioral reaction to the technique is rarely quantifiable, but it must be substantial, and it could lead to an erroneous interpretation of the results. Moreover, the time spent in a trap can well lead to reduced predation because of being protected by the trap, or to increased exposure to the elements, if provision is not made for nest material or for protective cover.

These examples of known or potential effects of the investigation itself, beyond any manipulations that are carried out, are not meant to be exhaustive, but to show the interested reader the kinds of problems that can arise and to stimulate careful thought prior to the start of a study.

Changes in controls

Natural areas have a tendency to vary in different directions, and such a divergence can cause problems in a long-term experiment. In the experiment involving transplanting salamanders (*Plethodon jordani*) described earlier, I predicted that one of the transferred populations would have a negative effect on the newly exposed population of *P. glutinosus* (Hairston 1973). The experiment was followed for eight years, with two plots being established where many individuals were introduced, and two plots remaining as unmanipulated controls. Unfortunately, the mean counts of *P. glutinosus* on the controls declined during most of the study (Hairston 1983), an observation that caused difficulty in interpreting the steeper decline in the mean counts of *P. glutinosus* on the experimental plots. In this case, there is a statistical test for comparing the rates of decline, but one can imagine situations in which the problem would be insoluble. This example is unlike the situation, already described, in which numbers on both control and experimental plots increased. In that case, there was not a significant difference in the rates of increase.

As a different kind of example, consider the results of an experiment in the control of a disease: schistosomiasis. One of the problems in controlling many tropical diseases is cost. These tropical countries are poor and lack the resources to do some of the obvious things. In this case, it was thought that because acquiring schistosomiasis requires active behavior on the part of people (getting into water with infective

Table 2.1. *Prevalence of schistosomiasis in four age classes in two villages*

Age (yr)	Control village			Experimental village		
	No. examined	No. positive	Prevalence (%)	No. examined	No. positive	Prevalence (%)
<3	56	1	1.9	46	0	0
3-6	115	13	11.3	92	7	7.6
7-10	109	43	39.5	90	35	38.9
11-14	79	47	59.5	74	35	46.7

Note: See text for full explanation.

snails; infecting snails by promiscuous disposal of feces), an aggressive campaign of health education might reduce the incidence of the disease.

In order to find out if such a program would have an effect, an experiment was set up in two villages with approximately equal proportions of children infected: 77 of 303 children (25.4%) in the village that was to be the focus of an educational campaign, and 104 of 359 (29.0%) in the village destined to be a control. The small difference is not significant statistically. For the purpose of this discussion, the absence of replication is ignored. The test consisted of examinations of children under 15 years of age, that group being accessible, as well as being the part of the population most receptive to health education. After 2.5 years of effort, the children were reexamined, with the following results: In the experimental village, 43 of 333 (12.9%) were found to be infected; in the control village, 75 of 322 (23.3%). The probability that the difference was due to chance is less than 1 in 1,000, a satisfying result, especially because the control village had not changed by a significant amount.

This crude analysis ignores the importance of age structure among the children, as well as the fact that the prevalence of schistosomiasis (percentage positive) increases with age. Broken down by age classes, the initial conditions are as shown in Table 2.1. The data show the sensitivities of these age classes to any measure reducing the rate of transmission. Thus, in the experiment, most of the members of each of the three youngest age classes would have moved into the next older class, and if transmission had been reduced, they would have diluted the older group with negatives. The data also show the importance of

Table 2.2. Prevalence of schistosomiasis in each age class in the two villages after 2.5 years of health education in the experimental village

Age (yr)	Control village				Experimental village			
	No. examined	No. positive	Prevalence (%)	Change (%)	No. examined	No. positive	Prevalence (%)	Change (%)
<3	47	0	0	-1.9	62	0	0	0
3-6	84	5	5.9	-5.3	106	1	0.9	-6.7
7-10	112	29	25.9	-13.6	94	15	16.0	-22.9
11-14	79	41	51.9	-7.6	71	27	38.0	-7.9

age structure in determining the crude prevalence rates reported earlier. A change in the representations of the different age classes in the population examined would alter the crude prevalence rate. An increased representation of the older children would increase the calculated prevalence of schistosomiasis, and vice versa. That is what happened, as the breakdown of the final data shows (Table 2.2). There was little difference between the two villages in the decrease in percentage positive at any age. The difference in overall percentages infected came almost entirely from a change in the relative representations of the age classes. In the experimental village, the two younger age classes increased from 45.6 percent of the total to 50.4 percent; in the control village, the representation of the same two age classes decreased from 47.6 to 40.7 percent. The result was that in the experimental village the overall parasite rate was diluted with the normal number of negatives among the younger children, whereas in the control village the parasite rate was inflated by the larger number of positives among the older children.

The change in age structure among the children examined could have brought about a statistically significant difference without any change in parasite rate in any age class. That can be confirmed by calculating the expected number infected at each age at the end of the experiment, assuming no change during the experiment. For the experimental village, the expected number is 77.67, or 23.3 percent of the 333 examined; for the control village, it is 101.62, or 31.6 percent of the 322 examined. That difference would have been "significant" statistically. Any claim of success in reducing the transmission of schistosomiasis would therefore have been spurious.

Examples of good design

The work of D. C. Smith (1983) is the first example discussed here of a field experiment designed and conducted in an excellent manner. Most experiments in fresh water have involved the use of enclosed spaces, with the ever-present risk that the enclosures themselves have influenced the outcomes of the experiments. This study, in contrast, was performed in naturally isolated bodies of water. They were rock pools near the shore of Lake Superior on two small islands associated with Isle Royale. The species of interest was the chorus frog, *Pseudacris triseriata*, which breeds in these pools.

Two factors endangered the tadpoles from the physical standpoint: desiccation in the smaller pools, and being swept out by wave action during storms for those pools low on the shore. Smith observed that there was virtually no survival of the tadpoles in either event.

No such accidental loss occurred in the larger pools far away from the shore, where tadpoles were rare or absent, and Smith considered two possible sources of tadpole mortality: predation and intraspecific competition. He identified three species of potentially important predators: larvae of the salamander *Ambystoma laterale*, nymphs of the dragonfly *Anax junius* (he has informed me that it was misidentified and is actually *Aeshna juncea*), and the dytiscid beetle *Rhantus binotatus*. His study of the initial conditions showed that *Ambystoma* and *Anax* were confined to the largest pools high above the water's edge and that *Pseudacris* tadpoles were found in unexpectedly few of these pools, and in low densities where they occurred. No such negative relationship was observed between *Pseudacris* and *Rhantus*, and although the beetles readily ate all sizes of tadpoles in the laboratory, only the smallest ones were affected in the field. Replicated laboratory experiments, in which two tadpoles of equal size were confined with one predator, established that *Ambystoma* could not eat large tadpoles (Fig. 2.3). The timing of breeding meant that there were few salamander larvae large enough to consume tadpoles, less than 5 percent of which were of a vulnerable size. The dragonfly nymphs were able to eat a larger array of sizes of tadpoles (Fig. 2.3), none of which were too large for the large nymphs.

In the field, *Anax* appeared to be abundant enough to have a strong impact on the tadpole populations, and Smith conducted two reciprocal experiments to test that hypothesis. He selected two large pools high up on the shore that contained dragonfly nymphs but no tadpoles, and two large pools lower down that contained tadpoles but no drag-

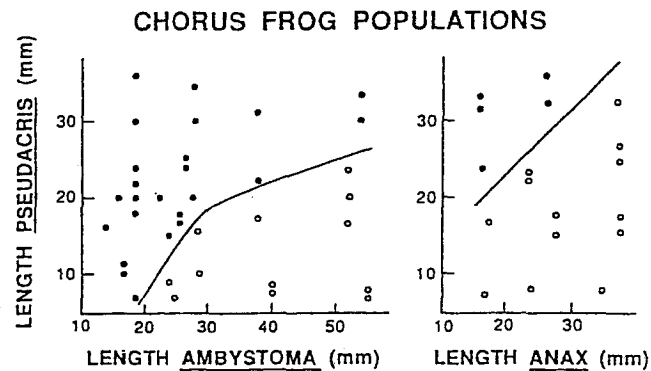


Figure 2.3. Relationships among total length of *P. triseriata* tadpoles, total length of predators, and vulnerability of tadpoles to predation in 200-ml laboratory containers. Tadpoles above the lines escape predation. Lines drawn by eye. Open circles indicate tadpoles eaten; solid circles indicate tadpoles surviving. (From D. C. Smith 1983.)

onflies. After measuring the volumes of all four pools, he removed the tadpoles from the lower pools and then reintroduced them at the same density for both pools. *Anax* nymphs were introduced into one of the two at a density comparable to the natural density. Nearly all *Anax* (at least 92.5%) were removed from one of the upper pools, and tadpoles were introduced into both at the same density as in the reciprocal experiment. The results were clear (Table 2.3). *Pseudacris* was eliminated by natural densities of *Anax*, but survived well where the predatory nymphs were absent or nearly so. The experiment revealed the reason for the virtual absence of tadpoles from seemingly acceptable pools that were large enough to nullify desiccation and far enough from shore not to be washed out by waves.

The next question concerned the factors controlling the populations of tadpoles in the remaining pools, which were transient enough to prevent successful colonization by the predators, but of sufficient duration to allow metamorphosis of the tadpoles to be completed. As part of determining the initial conditions, Smith sampled natural pools repeatedly. He found that growth was negatively correlated with density and that the length of the larval period was positively correlated with density. He carried out an experiment testing for the effects of density and food supply. Among the pools that qualified as acceptable, he established five blocks of four pools each. The pools within each

Table 2.3. Dragonfly predation experiments

Pool	Treatment	Pool volume (liters)	Natural numbers		Experimental numbers			
			<i>Anax</i>	<i>Pseudacris</i>	<i>Anax</i>	<i>Pseudacris</i>		
			19 June	18 June	19 June	25 June	2 July	10 July
<i>Experiment A: dragonfly addition</i>								
P1	No <i>Anax</i>	746	0	229	0	297	203	166
P2	<i>Anax</i>	330	0	102	33	132	17	0
			2 July	2 July	3 July	8 July	14 July	24 July
<i>Experiment B: dragonfly removal</i>								
P3	No <i>Anax</i>	283	92	0	7	112	58	57
P4	<i>Anax</i>	583	50	0	95	234	0	0

Note: Experiment A tests for the effect of *A. junius* on numbers of *P. triseriata* in pools initially occupied by *P. triseriata*. Experiment B tests for the effect of *A. junius* on *P. triseriata* in pools initially with *A. junius*. *A. junius* eliminated *P. triseriata* in both experiments. From Smith (1983).

Table 2.4. Data table for food and density experiment

Treatment		Survivorship (%)	Larval period (d)	Growth coefficient		Body length (mm)
Density	Food			First order	Second order	
L	H	90.0	46	0.321	-0.017	12.74
L	L	42.9	63	0.119	-0.001	11.22
H	H	71.2	52	0.239	-0.026	11.52
H	L	24.6	70	0.101	0.176	10.74

Note: H indicates high; L indicates low. Body lengths are of stage 42 tadpoles, defined here to be at metamorphosis.
From Smith (1983).

block were carefully matched according to size, exposure, relative depth, height on shore, and location. Each block contained four treatments of the two factors: high and low levels of both density and food. Assignment of treatments within each block was random. The densities were set within the natural range, with the high densities being 2-6/l and the low densities 0.3-0.8/l. Within each block, the ratio of high to low density was the same for both food levels, which consisted of the regular addition of a small amount of rabbit chow for the high food level and the naturally existing supply for the low food level.

The effects of the treatments were assessed as survivorship, duration of larval period, body length at metamorphosis, and growth (from a second-order polynomial growth curve, the first-order coefficient being determined by the rate of growth, and the second-order coefficient being the departure from constant, arithmetic growth). Density had a statistically significant effect on all of the variables; food affected survivorship, the first-order growth coefficient, and body length at metamorphosis. The actual mean values under the different treatments are in Table 2.4. There was no significant effect of an interaction between density and food. Larval period and growth rate were affected by block differences, a result showing the wisdom of using a block design, as the statistical manipulation removed the variance due to block effects before the remaining treatments were analyzed.

In this study, the author met the requirements of experimental design in providing initial conditions, adequate controls, and sufficient replication. The use of a block design spread the treatments over the

available array of conditions in an admirable way. He avoided the pitfalls described earlier, in not using pens, and in not using multiple measurements for each pool. It is noteworthy that the predation results were not analyzed statistically. It is not necessary to prove the obvious.

The second example of good experimentation was carried out in the laboratory (F. E. Smith 1963b). It is chosen as much for Smith's avoidance of circularity as for the experimentation per se. Smith pointed out that the principal difficulty in reconciling experimental data with the logistic model of population growth involved the universal occurrence of time lags in real population growth. Although they occur at all times, the lags are most obvious in the phenomenon of "overshooting" the carrying capacity of the experimental system. The reproductive process becomes irreversible, and having been initiated at a time when the population was well below the carrying capacity, the newly born young represent an excess. Their consumption of the limiting resource causes the carrying capacity itself to be lowered temporarily, and an oscillation is set up. The result is strong deviations from the logistic, especially in populations with an age structure.

He noted that many attributes of population growth had been studied by putting microorganisms into chemostats - continuous culture systems in which growth rates are determined by the rate of flow of nutrient resource into the system, compensated volumetrically by an equal outflow. Such a system is time-free. He devised an analogous system for animals with an age structure. His purpose was to test the hypothesis that without time lags the population could be described by the logistic equation. Smith's method was to grow *Daphnia* in such a way that continuous population growth was achieved at different rates of supply of algal suspension. The use of parthenogenetic animals allowed him to avoid dealing mathematically with males. He started with populations in 100 ml or less of culture fluid, which was changed daily and increased by a constant percentage. At the end of a week, each population was sorted to size, and each size class was reduced in numbers so as to return the system to its original volume and dilution. He kept this up until the size structure was "reasonably constant" - that is, until the stable age distribution could be assumed. Specimens not used to restart the cultures were preserved for future use.

For each population, he then had the imposed population growth rate per individual, $dN/N dt$, and the density of animals per 100 ml. These data could be used to test his original hypothesis. Because, from the logistic model, $dN/dt = rN(K - N)/K$, we know that as N , the

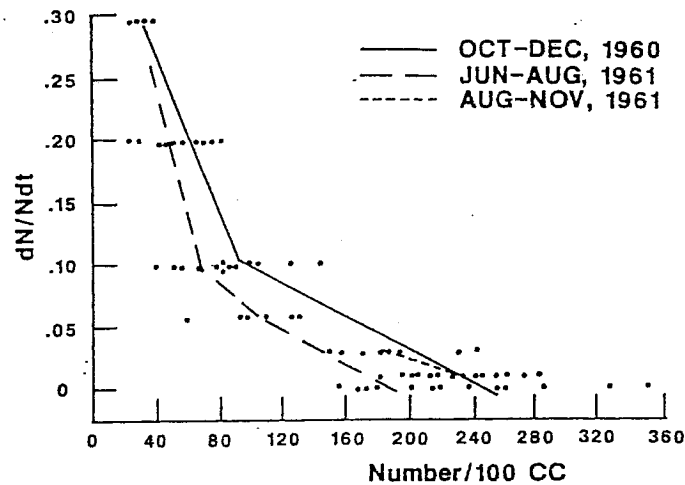


Figure 2.4. Observed densities of *Daphnia* at various specific rates of growth, using number of individuals as the measure of density. Data are combined from three sets of experiments, as shown. (From F. E. Smith 1963.)

number of individuals, increases, $dN/N dt$ decreases because $(K - N)/K$ becomes smaller and reaches zero when $N = K$, the carrying capacity of the system. The population growth rate per individual approaches r , the intrinsic rate of population growth, as N becomes very small. The relationship is linear. Therefore, if Smith's *Daphnia* populations had been following the logistic equation, his data points should have scattered around the straight line connecting r and K on a plot of $dN/N dt$ against N . They plainly did not do so (Fig. 2.4). Converting the numbers to dry weight (obtained from the animals preserved from the experiments) did not improve the fit, thus eliminating the possibility that the size structure gave inappropriate data.

He pointed out that in food-limited systems, the rate of supply of food and the rate of its utilization are items of more importance than are the carrying capacity and the unused part of the carrying capacity $(K - N)/K$. This being the case, the rate of utilization of food per unit biomass is not equivalent to the biomass as a proportion of the carrying capacity, because animals in a growing population use up more food per unit biomass than those at the carrying capacity, which would have just enough food to maintain themselves and replace deaths with

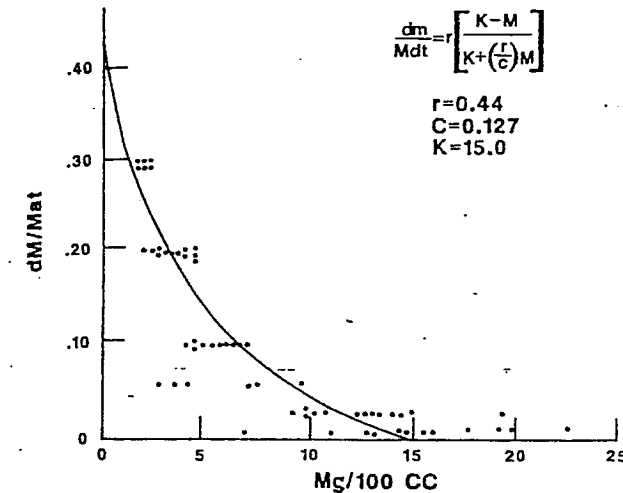


Figure 2.5. Dry-weight densities of *Daphnia* at various growth rates, compared with the proposed model. The data and curve were independently determined. (From F. E. Smith 1963.)

births. The necessary conversion of biomass into rate of food consumption requires the following modification of the logistic equation:

$$dM/dt = rM(K - M)/[K + (r/c)M] \quad (2.1)$$

where M is biomass; K is M at saturation, and c is rate of replacement of mass in the population at saturation. Smith gives the full derivation of all terms, for which his paper should be consulted.

Equation (2.1) has three constants: r , c , and K . In order to test the validity of his equation, Smith took the excellent step of measuring them independent of his expanding populations and thus was able to use the populations for the test. To obtain an estimate of r , forty-eight individuals were raised separately from birth in an excess of food, giving forty-eight independent rates of increase, the mean of which was 0.44/day, with a standard error of 0.005. A mean value of K was obtained from two populations maintained at zero growth rate for more than two months. The two values were 15.1 mg/100 ml and 14.6 mg/100 ml, with a mean of 14.85 mg. The constant c was obtained from death rates and the loss of dry weight by starved individuals. The estimate was 0.127/day. Using these independently obtained constants, the model was compared to the observations on the expanding cultures (Fig. 2.5). The fit was quite good, and most ecologists would

have been satisfied that the model had been validated, but in subsequent experiments using different algal densities, the model was found to be inadequate and Smith resorted to a different model (F. E. Smith, pers. commun.). This final twist to the account reveals the self-correcting nature of science in a way that dramatizes the experimental approach.

It is perhaps not fair to compare these experiments with field experiments, but it should be noted that the requirements of experimental design were met in all particulars, and the unusual rigor in the complete independence of the various measurements makes the work especially worthy of emulation.

An experiment that lacked important elements of design

The work of Morris and Grant (1972) is an example of work in which several of the requirements of experimental design were not met. I would have hesitated to cite it as an experiment were it not for Schoener's acceptance of the paper as an experiment in which interspecific competition was demonstrated (Schoener 1983). The purpose of the work was to look for competition between two species of microtine rodents, the meadow vole, *Microtus pennsylvanicus*, and the red-back vole, *Clethrionomys gapperi*, with *Clethrionomys* excluding *Microtus* from some habitats. Certain aspects of the ecological distribution of these voles suggested the initial hypothesis of exclusion of *Microtus* by *Clethrionomys*. *Microtus* is a grassland species, whereas *Clethrionomys* is typically a woodland species. On islands where no *Clethrionomys* are present, *Microtus* is less selective of its habitat than on the mainland, where *Clethrionomys* is present. The plan was to give *Microtus* the opportunity to occupy habitats from which *Clethrionomys* had been excluded experimentally.

The experiment was conducted as follows: A two-acre plot containing both grassland and woodland was surrounded by galvanized metal 3 ft (0.91 m) high, 1 ft being sunk in the ground. Within this plot, all small mammals were removed by trapping in May 1968. Those captured were *Microtus*, ground squirrels (*Spermophilus tridecemlineatus*), jumping mice (*Zapus hudsonius*), and shrews (*Sorex cinereus*) (Morris 1972). From the standpoint of establishing the initial conditions of the experiment, it thus consisted of unoccupied, enclosed habitat. As far as the experimental enclosure was concerned, the initial trapping data could have answered their question, because no *Cleth-*

riomys were present, and the distribution of *Microtus* in May must have shown where they were in the two habitats. The information is not given in this paper, nor in others on the same study (Grant 1972; Morris 1972). The absence of *Clethrionomys* from the enclosure raises the question of the suitability of the location for the experiment. The point may be relevant to the difficulty in establishing that species in a later phase of the study.

One of the flaws in the design of this experiment was the use of the same enclosure as the control for another experiment, testing the effect of the insecticide endrin on the population of another small rodent, *Peromyscus maniculatus*. Thus, when twelve *Microtus* were introduced after the general removals, twenty-two *Peromyscus* were also introduced. This was described by Morris and Grant (1972, p. 277): "Initially a few *Peromyscus maniculatus* were also introduced to the enclosure, in connection with another experiment (Morris 1972), but since most escaped immediately after introduction their generally brief presence is not considered a complicating factor." The paper referred to, however, shows that ten more were introduced (along with ten more *Microtus*) within the next month and that at least some of them remained throughout the first season of work, including some recruitment. *Peromyscus* varies geographically in its habitat, and the geographic race used was not specified. The possibility exists that even though less abundant than *Microtus*, it may have influenced the distribution of the latter. The number known to have been present at each trapping period has been added above the histogram (Fig. 2.6). *Microtus* increased in abundance to a total of sixty-three in the last trapping period. At no time were more captured in the woodland part of the enclosure than expected from the number of traps there, and in one period (of six), significantly fewer captures were made there than expected (Fig. 2.6).

No control enclosure was established, nor were there any replicates of the experimental enclosure. These were two serious omissions. We do not know what happened in the absence of the experimental manipulation, and we do not have any means of estimating the variation that may have occurred between enclosures treated alike.

At the beginning of the second season (June 1969), both *Clethrionomys* and *Microtus* were introduced into their appropriate habitats in the enclosure at densities typical of unrestrained populations. The actual numbers are not given, but the *Microtus* data for the four trapping periods are shown in Figure 2.7, and presumably *Clethrionomys*.

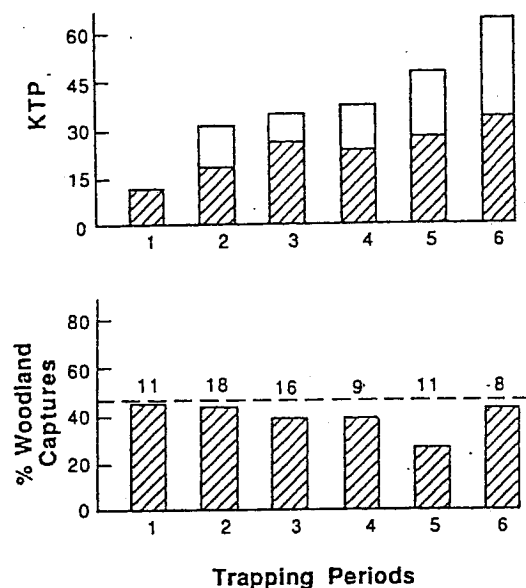


Figure 2.6. Known trappable population (KTP) of *Microtus* and percentage captures in the woodland in the first experiment. Unshaded portions of histograms indicate the number of animals trapped for the first time in each trapping period. The broken line indicates the percentage of traps in the woodland. Figures above the lower histogram show the number of *Peromyscus* known to be present. (From Morris & Grant 1972 and Morris 1972.)

had increased to thirty-seven in late July between the second and third trapping periods, as that number was reintroduced in August between the third and fourth periods. The reintroduction was followed by a "catastrophic decline," only seven being caught in the subsequent trapping period. The authors attribute the failure of *Microtus* to move out of the woodland (Fig. 2.7) to this reduction in numbers of *Clethrionomys*. The inconclusive nature of the results of that year's experiment was recognized, and the work was continued for a third season.

In 1970, there were twenty-two *Microtus* left from the experiment of 1969. These increased in abundance in the absence of *Clethrionomys* through August (Fig. 2.8). At that time, twelve *Clethrionomys* were introduced, and introductions were continued in an attempt to keep the number at that level. In 1968 and 1969, the experiments were ended in early to mid September, but in 1970, the observations were

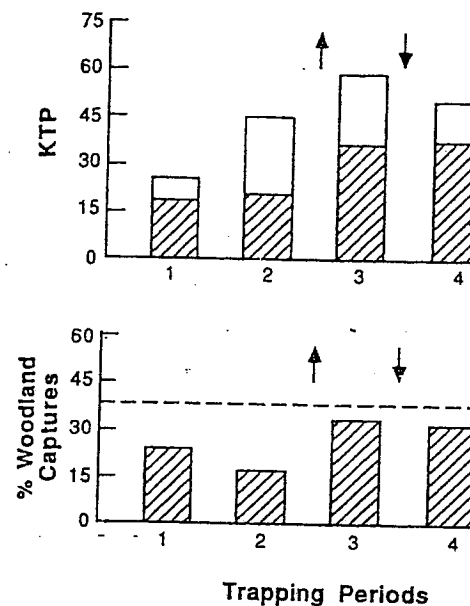


Figure 2.7. Known trappable population (KTP) of *Microtus* and percentage captures in the woodland in the second experiment. Upward arrows indicate the time of complete removal of *Clethrionomys*, and downward arrows indicate the time of reintroduction of *Clethrionomys*. Other information as in Figure 2.6. (From Morris & Grant 1972.)

continued until 3 December. As that was the only year with autumn data, comparisons are impossible.

The authors conclude that the results of the three experiments support the hypothesis of exclusion of *Microtus* by *Clethrionomys*. The data can equally well be interpreted as the result of intraspecific crowding by *Microtus*, if indeed any conclusion can legitimately be made.

To recapitulate, the flaws in this study were as follows:

1. The initial conditions were not given, though known from the initial trapping.
2. The presence of *Peromyscus* throughout the first season was not fully described.
3. There was no control. There were no *Clethrionomys* in the enclosure to begin with, yet their presence was in some sense considered to be a "control."

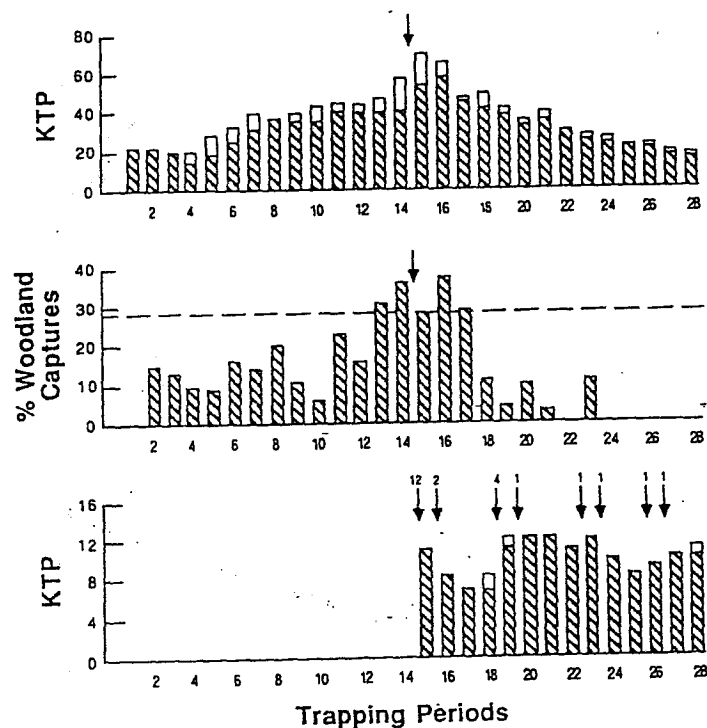


Figure 2.8. Known trappable populations (KTP) of *Microtus* (top histogram) and *Clethrionomys* (bottom histogram) during the third experiment. The middle histogram shows the percentage of captures of *Microtus* in the woodland. Arrows show times of introduction of *Clethrionomys*, and for the bottom figure, the number introduced. Other information as in Figure 2.6. (From Morris & Grant 1972.)

4. There was no replication.
5. There was an obvious alternative explanation in the high numbers of *Microtus* that were present when the confirmatory results were obtained in each year.

Synthesis

The requirements of experimental design do not need to be modified for ecological experiments, but certain of them do need to be emphasized. The requirement of clearly stating the initial conditions must be met, or there will be no possibility of using the conclusions

from the experiment in making true predictions, because a prediction can be expected to be true only if the initial conditions are known.

"No control, no conclusion" is axiomatic in all experimentation, but it may not be easy to have and maintain satisfactory controls. This is especially true in field situations, where natural variation over time may cause the controls to become different from other treatments spontaneously. Complex experiments, of course, may require ingenious controls to take care of the various contingencies.

The great headache in ecological experimentation is natural variability of conditions, and the principal defense is sufficient replication. There has been a tendency to begin an experiment before enough time has been spent in measuring the variability and thus determining the necessary amount of replication. Some of the problems can be avoided by the use of a block design, or by stratifying the data so that significantly different parts of a population are analyzed independently.

Despite care in planning, there are still unexpected pitfalls in ecological experiments. The first of these concerns the problem of pseudoreplication. Baldly stated, this is the inflation of degrees of freedom beyond what is warranted by the design of the experiment. It happens when counts or other measures are made repeatedly in time, thus invalidating the assumption of independence. It also occurs where treatment and place are confounded, as when several samples are taken within treatment plots, and the variation among those samples is used in the statistical analysis. According to the strictest application of rules, we would be required to discard valuable information.

A second pitfall concerns the unwanted side effects of the experimentation itself. This may take the form of "cage effects." The dynamics of the population being followed may be altered through the experimenter's presence influencing the behavior of individuals of either the subject species or its predators or competitors.

Changes in controls during the course of an experiment may make them no longer valid. Natural areas have a tendency to vary in different directions, and such an event in a long-term experiment can cause problems. A second kind of problem in controls can be caused by inattention to such matters as changes in representation in stratified data, which can make important differences in the interpretation of results.

This chapter has presented descriptions of two studies that illustrate proper attention to experimental and other scientific requirements, as well as a description of one study in which so many rules were broken that the conclusions cannot be accepted.

Trade-offs in ecological experimentation

Introduction

Ecological experimentation has some characteristics in common with mathematical modeling in ecology, which has been stated by Levins (1968) to face the impossibility of simultaneously maximizing precision, realism, and generality [see a discussion of this problem by Hunt & Doyle (1984)]. Ecologists conducting experiments face analogous choices, in that providing confidence in a relevant ecological process may preclude its general application, or the requirements of a sophisticated experimental design may severely decrease realism, or the use of an elaborate design in the field may put the necessary amount of replication beyond the resources of the investigator. In each case, the ecologist faces a difficult choice. All of the choices should be made deliberately, because to let them be made by default can lead to a misinterpretation or to such an unfortunate effect as replication insufficient to yield a convincing (that is, statistically significant) result.

Generality versus confidence

Conclusions reached from ecological studies increase in value as they can be extended to more and more situations. Therefore, any factors reducing this generality should also incorporate sufficient benefits to offset the loss. Experience over a wide range of habitats might be offset by a thorough knowledge of the ecological processes taking place in one or a few. Anyone undertaking ecological experiments is faced with this dilemma. The experiments are needed to test hypotheses about the processes, but they necessarily limit the results of the tests to one or a few localities. One of the first field experiments, that of Connell (1961a,b), provides an example. The vertical zonation of marine organisms on rocky coasts is a common observation, and Connell's experiments on an island in the estuary of the River Clyde in Scotland showed that the lower limit of distribution of the barnacle

Chthamalus stellatus was determined by competition with another barnacle, *Balanus balanoides*. The upper limit of *B. balanoides* was set by physical factors, especially desiccation, to which *C. stellatus* was less susceptible. Predation by the snail *Thais lapillus* was not important at levels where ecologically significant competition between the barnacle species took place. Lower down on the shore, it consumed an appreciable proportion of the barnacles. The clear and convincing picture of the causes of zonation that came from Connell's work did not apply to other areas, even to such a place as Puget Sound in the state of Washington, where the same genera of barnacles and predatory snails were present. In that locality, *Balanus glandula* occupied the uppermost zone, whereas *B. cariosus* was found lower down (Connell 1970, 1971). *Chthamalus dalli* was found scattered throughout the intertidal zone. *Balanus glandula* was excluded from all but the uppermost zone by predation from *Thais emarginata*, which ranged higher on the shore than did *T. lapillus* in Scotland. *Balanus glandula* settled regularly over a wide range of the intertidal zone in large numbers each year and was shown to be competitively superior to *C. dalli* in caging experiments. The young *B. glandula* that had settled below the highest level were as regularly consumed. Presented with a dependable food supply, *T. emarginata* had become adapted to feeding higher in the intertidal than other species in the genus. Connell's explanation of the difference in ecological interactions was that in Scotland, where settling was less regular, such an adaptation would have led to starvation in some years. Apparently, in Puget Sound, *Chthamalus dalli* escapes predation because *B. glandula* is the favored prey, and *B. cariosus* becomes too large for snail predation after two years of growth. That did not happen often enough for interspecific competition among the barnacles to be a factor in determining their distribution.

The experiments in these two areas have revealed in each area the ecological interactions determining zonation. The interactions were so different that generalizing from either would be impossible; and generalizing from observation of the zonation alone would almost certainly be erroneous. Slobodkin (1986) has implied that we should not attempt to find solutions to large questions in ecology because of the complexity of the organisms and the world in which they live. Although the solutions of two decades ago were obviously too simplistic, I am not ready to abandon the hope of syntheses, at least for some environments. The field might yet produce another Darwin.

A second example will serve to illustrate the trade-off between gen-

eralization and local experimentation in a different way. The organisms are two species of terrestrial salamanders, *Plethodon jordani* and *P. glutinosus*, and an explanation of their altitudinal distributions in the southern Appalachians has been the object of research over a period of sixty years. *Plethodon glutinosus* is widely distributed over the eastern United States, and consists of a number of geographic forms that are distinguishable by color pattern, or enzymatically, or both (Highton 1983). In the southern Appalachians, it occupies lower elevations than does the endemic *P. jordani*, which is partially or completely isolated at higher altitudes on a number of local mountain ranges. Early reports (Pope 1928; Bailey 1937) concentrated on geographic data, but Pope (pers. commun.) believed that altitudinal distributions would not be easy to interpret because the recorded locations of both species included a wide range of altitudes.

My study in the Black Mountains of North Carolina showed *P. glutinosus* to be present at altitudes from 915 to 1,433 m, while *P. jordani* was found from 915 to 2,013 m (Hairston 1949). The overlap was thus 518 m, entirely consistent with literature records. However, my data had been taken in the form of detailed transects on different slopes of the mountains and those records showed that the altitudinal overlap was no more than 70 m on any single transect. The total reflected the variety of directions in which the slopes faced, with overlap at lower elevations on north-facing slopes and at higher elevations on south-facing ones. My interpretation was that the two species were kept apart by competition. Similar transects on other mountains in the southern Appalachians subsequently revealed that the altitudinal distributions of the two species differed among mountains. The situation in the Great Smoky Mountains of North Carolina and Tennessee was like that in the Black Mountains; in the Balsam Mountains of North Carolina, which are between the Blacks and the Smokies, the species overlapped by at least 1,200 m (Hairston 1951). That observation required a different interpretation for the salamanders in the Balsams, and I concluded that competition was much less intense there than in the Blacks or Smokies. Extensive field work had led to two hypotheses about the interactions between the two species.

Experimental tests were proposed (Hairston 1973), and the experiments confirmed the hypotheses in that competition was shown to be more intense in the Great Smoky Mountains than in the Balsam Mountains (Hairston 1980a). The confirmation was satisfying, but still it is not known with certainty if the conclusions derived from the

experiments can be extended to the remaining mountain ranges jointly occupied by the two species. The altitudinal distribution in the Black Mountains leads to the expectation that there is intense competition between *Plethodon jordani* and *P. glutinosus*, but the situation is complicated by the presence of a third species of large plethodon, *P. vonahlossee*, which presumably has requirements similar to those of the other two. It is not found in either the Balsam Mountains or the Great Smoky Mountains. Similarly, *P. jordani* and *P. glutinosus* coexist throughout the South Mountains, a small range isolated to the east of the rest of the southern Appalachians. The broad overlap in the South Mountains suggests that competition is less intense there than in the Smokies, as it is in the Balsams. It should be noted that *P. jordani* attains a considerably larger size there than it does elsewhere (R. Highton pers. commun.), and that could alter its relationship with *P. glutinosus*, which is larger than *P. jordani* in all other localities. These differences among localities, which are typical of many such situations, should cause us to be cautious about generalizing from even well-studied situations. We do not know how many detailed experiments must be carried out before the generality can be accepted. Perhaps some clear-sighted ecologist of the future will be able to use the accumulation of experimental evidence in at least one kind of habitat and produce the hoped-for synthesis.

Realism versus sophistication of experimental design

Ecological relationships are so complicated that in attempting to deal with them, we must, like Hercules fighting Hydra, put some of the heads under rocks while we contest with the others.

Dealing with more than one or two ecological questions at once requires a complex experimental design. The demands of statistical analysis are such that in complex experiments, variability within treatments should be kept nearly constant from treatment to treatment. Such an experiment permits a number of questions to be answered at once and also allows for the assessment of interactions between factors. The price to be paid is having a large number of experimental units in which the experimenter has a measure of control over the sources of environmental variability. The requirements are most frequently met in the laboratory, but there have been some examples of elaborate experiments that were carried out under seminatural conditions in the field.

Laboratory experiments

Laboratory experiments involve at least two conditions that raise questions about their ecological meaning. The first is the widely recognized simplified environments that must be used. This is an advantage, in that the conditions permit the experimenter to confirm or eliminate the importance of some factors that have been advocated as important from the theoretical standpoint. A good example is the large set of experiments on competition among an array of species of *Drosophila* conducted by Gilpin, Carpenter, and Pomerantz (1986). They concluded that competition was most likely exploitative, with some evidence for generalized interference, and none for specific interference. As might be expected from those results, nearly all combinations revealed transitivity of competitive success. That is, if species *A* wins over species *B*, and *B* wins over *C*, then *A* practically always wins over *C*. For *C* to win over *A* would require some form of highly specific competitive ability against *A*, an ability not involved in exploitative ability or generalized interference, such as secretion of a metabolite generally toxic to other species of *Drosophila*.

It is also noteworthy that Gilpin and associates found almost no evidence of niche partitioning in their simple laboratory systems. Coexistence among ten species competing at once never involved more than three species at the arbitrary end of an experiment; in the whole set of experiments, only five of the species were ever present at the end, and the authors argued that two of those would have gone extinct had the experiments continued longer. It is true, of course, that not many niche axes were available for the flies to partition, which is a partial explanation of the finding of the importance of exploitative competition. That was the nearly inevitable consequence of the simple laboratory environment.

The second condition that raises questions regarding the ecological meaning of laboratory experiments is the restricted array of species that can be used, especially the restricted array of *kinds* of species available for such work. Experiments on these species were of interest in the early days of ecological experiments, such as those by Gause (1934) or Crombie (1945, 1946) when the applicability of simple mathematical models was being tested. Practical considerations require that the species used in laboratory populations be both small and easily cultured. Thus, most ecological experiments in the laboratory have used such organisms as phage, bacteria, protozoans, algae, rotifers, cla-

docerans, and a restricted array of terrestrial arthropods: grain pests, fruit flies, a few mosquitos, and a few mites. These species, which seem to be preadapted to laboratory conditions, tend to be the adaptable "weeds." They lack the highly specific requirements that imply coevolution in organized communities. Few of them have been studied in the field to an extent that would allow us to suspect the factors determining their distribution and abundance. That is especially true of interspecific interactions, either competition or predation, that may have an influence on the sizes of their populations. The example of competition among the twenty-eight species of *Drosophila* from around the world demonstrates only that many of them compete in laboratory cream bottles, where they have little choice. The experiments tested no hypotheses derived from the ecology of the species, and no such hypothesis could have been tested, as that ecology is virtually unknown. It was not to be expected that they would reveal much of interest that could be applied to field systems where the species have evolved in the presence of competitors or predators.

In fact, it is difficult to find any population experiments conducted in the laboratory that were designed to test hypotheses generated from the ecology of the organisms as understood from field studies. The explanation advanced here is that field studies suggesting the importance of competition or predation have involved species that do not lend themselves to laboratory experimentation. Laboratory experiments testing field-generated hypotheses have involved not populations but individuals whose physiology or behavior was being investigated. Examples of that kind of work are the studies by Muller, Muller, and Haines (1964) on the effect of allelochemicals and Landa and Rabinowitz (1983) on differential susceptibility of plant species to herbivore feeding. The species involved could be kept alive during the experiments, but no attempt was made to culture them, and population phenomena were not involved. There is no intent to denigrate such experiments. Many of them are revealing, but they are not relevant to the present subject.

Field experiments

Work conducted under field conditions almost automatically involves much greater realism than that conducted in the laboratory. The experiments are carried out where the natural situation contains the basis for proposing the hypothesis to be tested. Unless the postu-

lated ecological phenomena are irregular in occurrence (Wiens 1977; Dunham 1980), any failure to demonstrate them means rejection of the hypothesis of their existence, or else means that the experiment was not adequate in design or execution. The requirements for any experiment depend on the question or set of questions to be answered. They are described in Chapter 2. For present purposes, it is sufficient to note that the number of experimental units depends both on the complexity of the design and on the variability of the data that are expected to be collected. Field experiments will automatically involve greater variability than laboratory experiments, and it can be assumed that the experimenter is limited in time and resources to the extent that the need for enough experimental units (plots, pools, or cages) to overcome the variability problem will limit the complexity of design and thus limit the sophistication of the kinds of questions that can be attacked simultaneously. Suppose, for example, that D. C. Smith (1983) whose experiment on tadpoles is described in Chapter 2. had needed to explore the effects of three levels of density or three levels of food. Neither would have been thought unusual as being of general interest. To have retained the same design would have required manipulating and monitoring half again as many pools, meaning perhaps more effort than was possible. To have explored three levels of both factors would have required 2.25 times as much effort. Keeping the effort approximately constant would have meant reducing the number of blocks, and hence the number of replicates, from five to either three or two. Anyone who has engaged in field experimentation will appreciate the danger of loss of statistical significance from the loss of replicates per treatment.

A second class of problems encountered in field experimentation can compound those already mentioned. These problems concern the necessary size of the experimental units. Except for trees, sessile or sedentary organisms pose few problems, the size of experimental units being determined by density and the number of individuals needed for adequate populations. It is the motile animals that cause difficulties. The necessity of knowing in advance the extent of movements is well illustrated by the experience of Schroder and Rosenzweig (1975). They investigated possible competition between the heteromyid rodents *Dipodomys merriami* and *D. ordii*. These granivorous species were similar in size and in terms of seeds eaten in central New Mexico, and although they showed differences in preferred vegetation, they were caught frequently at the same trap locations on grids where both kinds of vegetation were present.

The plan was to remove each species to two different levels on matched grids, retaining two grids as controls. A response of either species to the lowered population density of the other was to be interpreted as evidence of interspecific competition. Each grid covered 16.2 ha and consisted of 331 live traps set 25 m apart in a concentric hexagonal array. The diameter of the hexagon was 500 m, and none of the grids was fenced. The ten grids were placed at wide intervals over a total area of 3,108 ha. Captured rodents were marked with numbered ear tags, and those not to be removed experimentally were released at the trap sites. The most remarkable finding was the rapidity with which each species returned to its initial density on grids from which conspecific individuals had been removed. The opposite species responded to the removals to a negligible extent.

Intuitively, the sizes of the grids would appear to have been more than adequate for the experiment, as most small mammals are quite sedentary. That was not the case with these two species of *Dipodomys*, even though they reside in burrows. The new animals were adults that were distinguishable from the remaining residents only in the absence of ear tags, and hence the immigration could not be attributed to juvenile wandering in search of unoccupied habitat. Although recaptures between the widely spaced grids were rare, two instances show the remarkable vagility of *D. ordii*, at least. Between February and May, one individual had moved a distance of 9.5 km, and another had moved 5 km in one week. There was much more movement than had been anticipated, and despite the large size of the grids, they were inadequate to keep the populations reduced long enough to show an interspecific effect. Presumably, it would not have been feasible to enlarge the grids sufficiently without sacrificing important elements of experimental design, even had Schroder and Rosenzweig been aware of the extent of movements of individual animals.

Managed environments in the field

There are, of course, intermediate situations between the largely controlled conditions in laboratory experiments and the completely open conditions under which many field experiments are conducted. These may involve enclosures to keep out predators or competitors, with accompanying confinement of the experimental species and the general effect of such barriers on the physical environment, or they may involve outdoor containers in which there is control over all species introduced during the experiment, with accompanying loss of

realism. The outstanding example of the use of outdoor containers to achieve definitive results is the work of H. M. Wilbur and his students (Morin 1983a; Morin et al. 1983; Wilbur et al. 1983). The use of numerous cattle tanks as surrogate pools permitted the establishment of carefully replicated but complex communities of amphibians and the accompanying array of plankton and macrophytes. Block designs controlled for potential effects of location within the hexagonal array of tanks. The elegance of experimental design was followed by sophisticated statistical analyses of the results, and the interaction of competition and predation in structuring the communities of tadpoles and plankters was revealed in impressive fashion. Pure competition among the tadpoles resulted in the group of metamorphosed frogs being dominated by *Scaphiopus holbrooki* and *Bufo terrestris*, with *Hyla crucifer* and *H. gratiosa* being nearly eliminated. With increasing numbers of the predator *Notophthalmus*, first *Bufo* and then *Scaphiopus* declined in representation, and *Hyla crucifer* assumed increasing prominence. Of equal interest was the relationship between competition and predation in their effects on the biomass of froglets produced. The presence of intermediate numbers of *Notophthalmus* reduced competition, and the tanks produced significantly greater masses of metamorphs than did those without predators or those with the maximum number.

There was some obvious loss of realism, in that natural pools do not remain the same size and depth. Wilbur (1987) added differential gradual draining as a manipulation in otherwise similar experiments, but the vertical sides of the cattle tanks remained an unrealistic element. Another unrealistic aspect of those experiments was the prevention of immigration and emigration, properties not shared with natural pools. Perhaps more important was the necessary limitation of the array of amphibian species used in the experiments. The most abundant four to six species of anuran tadpoles were chosen from a potentially interacting group of more than twenty species, and one or two species of predatory urodeles were chosen from an array of five or more. Thus, whereas we know a great deal about the ecological interactions in the tanks, the degree of correspondence between those events and competition and predation in the natural ponds is only partly known from an experiment undertaken by Morin (1983b) in the pond that inspired the work. Working with the two species of predators used in the tank experiments, he found that *Notophthalmus viridescens* preyed extensively on the eggs of *Ambystoma tigrinum*, reducing survival to hatching from 98.3 percent among eggs in the pond that were protected from

Notophthalmus to 3.5 percent among unprotected eggs. In the experimental tanks, *Ambystoma* larvae preyed on *Notophthalmus* larvae, and reduced their production to about 27 percent of the control numbers. Without the work in the natural pond, the mutual predation would have been unknown.

The natural predation by *Ambystoma* on anuran tadpoles is unknown, but it seems unlikely that it is as great as was observed in the experiments. A density of four *Ambystoma* larvae per tank (stated to be within the natural range of densities) eliminated all 1,200 tadpoles from each of three tanks and permitted only a total of 13 to reach metamorphosis from the remaining three tanks. Data were not provided on field survival rates, but 0.0018 seems intuitively to be low.

These experiments have resulted in a wealth of information about the effects of intraspecific and interspecific competition on growth, survival and duration of the larval period, as well as the potential of predation in ameliorating those effects. The difficulty is in translating what can be the interactions in nature into what they actually are. Seminatural experiments such as these provide one possible transition between variable and otherwise difficult natural situations and the unrealistically simple environments of the laboratory.

Sophistication of experimental design versus adequate replication in the field

A recurring theme in previous sections of this book has been that natural situations are variable and that the variability imposes a necessity to have sufficient replication to distinguish statistically between average differences among units that have been treated differently. In combination with the requirement that the units be large enough to accommodate the natural movements of a sufficient number of the animals being investigated, the need for multiple units imposes a strain on the resources of any investigator. The experiment of Menge et al. (1986a,b) involved monitoring thirty-two plots, each 0.25 m² in area, for three years. Their conservative estimate was that about 11,410 person-hours had been devoted to establishing and maintaining the study, and they acknowledged that they were unable to avoid pseudoreplication because of constraints on time and budget. For most of the experimental conditions, they maintained only two replicate plots. Fortunately, the effects of the manipulations were large, and two replicates were sufficient to show a number of interesting ecological inter-

actions. With a greater, and not unusual, amount of variation between replicates, many of these effects could easily have been undetected, as in the equally sophisticated experiments of Hall et al. (1970) described in Chapter 7.

It is thus essential to establish the basic variability during the planning of an experiment, so that the choice of design can be made between an elaborate experiment that will test for several effects and their interactions, and a simpler experiment that will give fewer answers but will have sufficient replication to ensure statistically satisfactory results. As an example of the difficulty in making the decisions, I have chosen the well-designed experiments of Stiven and Kuenzler (1979). They addressed the question of what affected the populations of two species of mollusk in salt marshes in the vicinity of Morehead City, North Carolina. The mollusks were the snail *Littorina irrorata*, which feeds by scraping surfaces, and the mussel *Geukensia demissa*, which is a filter feeder. Their hypothesis was that the animals were competing intraspecifically and that the most important resources were the deposits of detritus from the only important plant on the marsh, *Spartina alterniflora*, the salt marsh grass, and the periphyton on stems and the marsh surface. In the years 1974 through 1976 they carried out an intensive sampling program in three salt marshes, one of which (Calico Marsh) was heavily influenced by sewage effluent. In all three marshes, both species were strongly clumped in distribution, with mean densities being small fractions of the variances (Table 3.1). The distributions precluded the use of normal statistics in comparing the marshes, but the differences were striking, especially for *Littorina*. That observation forced a difficult choice in adhering to their experimental design.

The plan was to use nine treatments, consisting of three densities of the mollusks (normal, half normal, and twice normal) and three amounts of *Spartina* litter for each density. For use as a block, nine enclosures 10 m by 0.75 m were established in each marsh, there being no replication of any treatment within blocks. All mollusks of both species were removed for measurement and reintroduction. Unfortunately, the mean numbers removed in the different marshes differed by a factor of seven for *Littorina* and more than six for *Geukensia*, making the establishment of a single normal control density for all three blocks (marshes) impossible. Use of the observed mean density within each block would mean giving up the replication between blocks, and to use the same density for all would mean that the base-

Table 3.1. Sampling data on two salt marsh mollusks over a period of two years

Marsh	Month	<i>Littorina</i>			<i>Geukensia</i>		
		N	Mean	Variance	N	Mean	Variance
Tar Landing	10/74	25	18.6	506.25	25	3.6	20.25
	6/75	30	14.5	346.8	30	2.8	36.3
	10/75	30	11.3	202.8	30	5.7	218.7
	6/76	30	13.2	388.87	30	0.1	0.3
	10/76	30	11.4	346.7	30	0.4	2.7
Causeway	11/74	20	0.8	7.2	20	4.3	80.0
	6/75	30	2.3	24.3	30	2.0	58.8
	10/75	30	0.5	2.7	30	7.5	346.8
	6/76	30	0.7	14.7	30	2.3	19.2
	10/76	30	0.4	1.2	30	1.6	19.2
Calico	10/74	25	42.2	2304	25	2.4	20.25
	6/75	30	30.9	1598.7	30	2.4	58.8
	10/75	30	28.4	1555.2	30	1.6	14.7
	6/76	24	19.8	952.56	24	4.4	15.36
	10/76	30	30.7	4465.2	30	0.5	4.8

Adapted from Stiven and Kuenzler (1979).

line density, whatever was chosen, would be unnatural for two of the blocks. They chose the second alternative. For one block, the high density returned to the enclosures was below the mean number removed, and for another the low density introduced was nearly double the mean number removed. These statements apply to *Littorina*, but matters were little better for *Geukensia* (Table 3.2).

The results of these experiments are difficult to interpret. As far as survival was concerned, both species survived significantly better in Calico Marsh than in the other two (Fig. 3.1). This result was independent of the experimental manipulations, because Calico Marsh was intermediate between the others in premanipulation density. The sewage effluent may have had an effect on survival, or the greater distance from open bay water may have reduced the number of predators. Survival was related to the experimental manipulations in a few cases. *Littorina* survived better at lower densities in Tar Landing Marsh — the area where all experimental densities were below the natural density, a result that appears to be fortuitous. Similarly, survival of *Geukensia* improved in the expected directions in Causeway Marsh for

Table 3.2. Existing densities per enclosure, and experimental manipulations of two salt marsh mollusks

	<i>Littorina</i>			<i>Geukensia</i>		
	Tar Landing	Causeway	Calico	Tar Landing	Causeway	Calico
Mean no. removed per enclosure	342.2	46.9	150.2	21.8	8.2	53.2
No. returned at:						
Half density	75	75	75	14	14	14
Full density	150	150	150	28	28	28
Double density	300	300	300	56	56	56

Note: Each enclosure was 10 m by 0.75 m.
Adapted from Stiven and Kuenzler (1979).

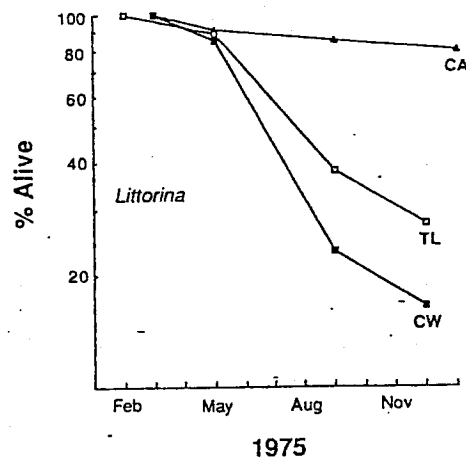


Figure 3.1. Survivorship trends for experimental populations of *Littorina* in Calico (CA), Tar Landing (TL), and Causeway (CW) marshes during the first experimental year, all treatments combined. (From Stiven & Kuenzler 1979.)

both density and added detritus, but note that all experimental densities there were three times that planned, relative to the local density.

It is easy to criticize this study with the help of hindsight, but the dilemma posed by the great variability in local densities of the mol-

lusks would have required abandoning what was an excellent experimental design. In hindsight, it can be suggested that confining the experiment to one marsh would have answered the questions posed, but only for that marsh, and generality would have been lost. The actual densities were so different at the locations of the blocks that the use of more than one marsh would not have been feasible, even if the experiment had been redesigned so that fewer densities of mollusks or litter would be used, in the interest of increasing the number of replicates. The lesson to be learned is that it may be difficult or even impossible to devise experiments to answer every question we need answered.

Synthesis

Ecological experiments are nearly always conducted under conditions that force us to choose among alternative strategies for solving our problems. Some of the trade-offs come from the act of experimenting in the first place. Others come from limitations of time and other resources of the experimenter; still others come from the choice between levels of management of the environment in which the experiment is to be carried out.

Similar observations in different locations tempt us to provide a common explanation, but the hypothesis thus generated usually can be tested experimentally in only one place. A successful outcome gives confidence in the hypothesis, but the confidence may not be extendable to other areas. Two examples have been given in which it was known from separate experiments that generalizing from either would be incorrect or in which there was reason to question that further experiments in specific different locations would give the same results, even though the basic observations were similar to those in the locations of the experiments.

Experiments carried out in the field provide the assurance that results are applicable in the real world from which was derived the hypotheses that were tested. The price to be paid for the realism is that normally rather simple experimental designs must be followed, because the variability of most factors requires a number of replicates, and for motile animals the size of experimental units must be large enough to ensure a relatively small proportion of emigration and immigration. Experiments that are designed to detect several ecological phenomena and their interactions require low variances among

replicates and control by the experimenter over the sources of such variance. That nearly always means that the experiments must be carried out in the laboratory. The disadvantages of laboratory experiments involve the simple environments that must be used and the limited kinds of species that can be carried through a sufficient number of generations for completion of the experiment. Both of these features greatly reduce the realism of the study. Seminatural experiments may offer a reasonable compromise between field experiments and those conducted in the laboratory. They allow both a larger degree of experimenter control than is true in the field and a greater degree of complexity and hence realism within the experimental units than can be achieved in the laboratory.

A third trade-off faced by ecologists is the choice between adequate replication to assure statistical significance of results and a sophisticated design that theoretically will answer a complex set of questions. Assuming that resources and effort are available to the usual extent, the variation ordinarily encountered in the field may require more replication than can be sustained in a complex experiment. Nature has no stake in our understanding its interactions, and it may not be possible to carry out all desirable experiments.

4

Experiments in forests

Introduction

Humans are terrestrial. Our experience is almost exclusively with that environment, and we should therefore have an intuitive understanding of terrestrial situations that is superior to our grasp of aquatic situations, whether freshwater or marine. For that reason, I have begun this description and criticism of ecological experiments with those performed on land. One of the themes of this book is that many generalizations in ecology are unlikely to hold across environments that differ in major ways. A consideration of the results of observations and experiments suggests that on land, interactions among the elements of the biota will be distinctive for at least the following categories: forest, prairie, desert, and tundra. Further subdivisions may be necessary.

A conventional and logical arrangement of the experiments to be described is to consider them in the context of the trophic level on which they were conducted, starting with the decomposers. It is recognized that it is not always possible to make the assignment cleanly. For example, in taking prey, most predators do not discriminate between herbivores and smaller predators, but for a considerable majority of the experiments, the assignments are reasonable. No attempt is made to describe all of the experiments on any trophic level. Those experiments selected either illustrate the interactions of greatest interest, or are examples of problems that face experimental ecologists, or represent good or bad experimental design.

Forests present practical difficulties for a number of kinds of experiments, and fewer have been conducted in forests than in more accessible habitats. In considering the results of experiments in old fields and other successional areas once forested, it is necessary to keep in mind the successional nature of the communities, because some of the processes hasten the end of the stage, and others tend to stabilize the situation. Descriptions of experiments in successional communities and in arid environments are deferred to later chapters.