6
Ecological Experimentation: Strengths and Conceptual Problems

DAVID TILMAN

Introduction

The ecological world is a multi-causal system in which patterns result from the direct effects of physical factors, from intraspecific and interspecific biotic interactions, and from indirect and feedback effects of one species or system element on another. As such, there is no single, simple approach that can ever unambiguously demonstrate how or why a particular process, physical factor, or species has an effect on another element of the ecosystem. Although this chapter will critically evaluate the potential contributions of experimentation to ecological understanding, I must stress at the outset that ecological research requires a synthetic approach in which observational, experimental, and theoretical approaches are pursued in a simultaneous, coordinated, interactive manner.

It is important to define the central goal of ecological research before discussing how such work should be done. I consider ecology to be the study of the causes of patterns in nature. As such, it is critically important that natural patterns be adequately described. Such observational studies can be used not only to establish the relations that exist among species and between species and various elements of the physical and chemical habitat, but to generate hypotheses that may explain these patterns. The formulation of such hypotheses can often suggest further data that could be collected to help distinguish between the predictions of competing hypotheses. This approach, the comparative approach, has been a powerful tool in many branches of ecology, especially in evolutionary ecology. However, a criticism that has been repeatedly leveled against the comparative approach is that there are an infinite number of alternative hypotheses that could explain any given set of observations (Jeffreys, 1931, as quoted in Hutchinson, 1978, p.3). The collection of additional observational data can never lead to the rejection of all but one of these. Put differently, the comparative approach can only provide weak inference. Hairston (1980, p. 824) alluded to this problem when, in discussing some of the problems with field experiments, he stated “None of these difficulties, however, are as great as are the problems of weak inference, in which the maximum hope is for a correct prediction, which is frequently a prediction of information already in hand, and may in fact be correct
for the wrong reason. We should not forget that Ptolemy's calculations correctly predicted the positions of the planets. Ptolemy, you may recall, assumed that the earth was at the center of the solar system.

An observational or comparative approach is only powerful in those rare cases (which may be nonexistent in ecology) in which there can be strong inference, i.e., for which there are a few mutually exclusive alternative hypotheses. In such a case, observations that led to the rejection of all but one of the alternatives would prove that the one not rejected was correct. However, because of the feedback effects of one species or system element on another, and because of the number of interacting and interdependent elements, there are few, if any, cases in which strong inference can be used in ecology. Detailed observations are clearly an integral part of ecological research, because they are necessary to define the patterns which exist in nature. Long-term observations, as detailed in the chapters by Franklin, Davis, and Taylor, are much more useful than short-term observations. However, observations, by themselves, are incapable of determining the causes of the patterns we see in nature.

Another approach to ecological research is the development of theory. Theory, whether it be the loose verbal sort of "theorizing" common to the field or rigorous mathematics, is simply the process of determining the logical implications of a series of simplifying assumptions. Theory is a critical part of ecological research because it allows us to formulate testable predictions and thus to distinguish between various alternative sets of simplifying assumptions. By its nature, it encourages us to seek simplicity, which has been a hallmark of the scientific method since William of Ockham coined his famous razor in the 1300s. However, theoretical ecology cannot stand by itself. No matter how pleasing the assumptions of a theory and no matter how elegant its predictions, a theory must be tested using real data collected in nature.

The other approach to ecological research is manipulative experimentation (sensu Hurlbert, 1984). An experiment is a deliberate manipulation of one or more variables done in such a way that other variables are held constant or rendered unimportant through suitable experimental design. In the confines of the laboratory, other variables are often held constant through careful control and replicates are used to determine experimental errors. In the field, it is often impossible and may not be desirable to hold all nonexperimental variables constant. Instead, randomization, replication, and interdispersion of treatments provides a control for nonexperimental variables. Hurlbert (1984) has criticized numerous ecological "experiments" for their lack of replicates and controls, and for the resulting misapplication of statistical tests during data analysis. I will not reiterate his criticisms, but do wish to stress that improperly designed or incorrectly analyzed experiments can lead to incorrect conclusions.

Manipulative experiments differ markedly from observational studies because experiments, if properly designed, can determine if there were a causal link between a manipulated variable and some measured response. Observational studies cannot. No matter how good the correlation might be between two variables, only appropriately randomized, replicated, interdispersed, and controlled
experiments can distinguish between correlation and causation. As such, experimentation is essential for conceptual advances in ecology.

The potential power of experimentation has led it to be an increasingly popular approach among ecologists. Its popularity is evident in the research grants that are being funded and the papers that are being published in response to peer review. Schoener (1983) looked at a narrow set of ecological field experiments, those designed to determine if interspecific competition were occurring in nature. In doing his review, Schoener noted that most such studies were new: "Indeed, when plotted, growth of all [experimental competition] studies combined appears greater than exponential" (Schoener, 1983, p. 241). "Rare until recently, field-experimental studies of interspecific competition now number well over 150" (Schoener, 1983, p. 276). Connell (1983a), who independently reviewed a more limited subset of this literature, found 72 different papers reporting the results of experimental studies of interspecific competition in the field. Sih et al. (1985) surveyed 20 years (1965–1984) of seven major ecological journals to find field manipulation experiments on predation and found 139 such papers. The number of such studies seems to have grown in a logistic manner from 1965 to 1984, and may have reached a plateau of about 25 per year for these seven journals combined (Sih et al., 1985). As a further measure of the popularity of experimentation in ecology, but mainly to address some questions that are raised later in this chapter, I surveyed a subset of Ecology issues (2 randomly chosen issues per year for 1977–1984; all issues for 1985-Feb. 1987). Each paper was classified as to the duration of the study, and as to whether it was based on or included a strong component of direct observation, an inferred chronosequence, paleoecological data, manipulative experiments, or theory. Of the total of 749 papers in the sample, 387 (52%) were either mainly experimental or had an experimental component. Of these, 180 papers had a field manipulation experiment as their central purpose. There were 99 papers (13%) that were mainly conceptual or theoretical papers, most of which also included an observational or experimental component. The remaining papers were mainly observational. Thus, as shown by the specialized reviews of Schoener (1983), Connell (1983a), and Sih et al. (1983), and by my more general review, field experimentation is now a major component of ecological research.

I have argued, above, that ecology requires a mixture of observational, experimental, and theoretical approaches. Although ecology was overly biased toward observational research 20 years ago, there are now major experimental and theoretical components. This balance is a measure of the maturation of ecology as a science.

All, though, is not well in the ecological enterprise. The advent of the experimental approach has brought with it a new series of methodological and conceptual problems that are hindering major advances in ecological understanding. The methodological problems, which have already been addressed by Hurlbert (1984), concern the statistics of experimental design and analysis. In this chapter, I will address the conceptual problems. These conceptual issues must be resolved before ecological field experiments can provide the inferential strength that they seem to promise. Despite the vast number of field experiments that have been
performed to date, few of them have dealt properly with the conceptual issues I raise below, and thus almost all of them are open to interpretations that differ dramatically from those reached by their authors. Because few of the authors have acknowledged these flaws in their experiments, an inordinate weight has been put on the value of these experimental results. This acceptance has led many ecologists to reject, as spurious correlations, contrary patterns suggested by observational studies, even though these contrary patterns may be as indicative of causation as the improperly conceptualized experiments. Such conceptual problems are a major factor hampering the ability of the experimental approach to contribute maximally to ecology. As is detailed in the following section, many of these conceptual problems have occurred because ecologists have used short-term experiments to address long-term questions.

Conceptual Problems for Field Experiments

There are four main conceptual problems that many field experimentalists have not considered properly when designing experiments and interpreting experimental results. These are (1) transient dynamics, (2) indirect or feedback effects, (3) environmental variability, and (4) multiple stable equilibria and site history. All of these problems occur because of the disparity between the extreme complexity of nature and the simplicity of the conceptual approaches upon which many experiments are based. Until now, most experimentalists have been satisfied with finding a statistically significant result from their experiment. However, a well-replicated, well-controlled, suitably randomized, and interdispersed field experiment that yielded highly statistically significant results could be easily misinterpreted if its conceptual basis were faulty. I will discuss the four major conceptual problems and provide examples of each, in the order given above.

Transient Dynamics

All experiments are manipulations, i.e., perturbations to an existing system. The system may or may not have been at equilibrium at the time of the experiment. Assuming that the perturbation is not a one-time pulse, but a continuous manipulation, the perturbation could cause the system to assume eventually a new equilibrium or to oscillate with a different mean amplitude or frequency. However, one of the most general features of a perturbed system is that the system can undergo complex dynamics during the transition from its original state to the new experimentally imposed state. These dynamics are its transient dynamics. The more complex a system is, the more likely it is to have complex transient dynamics during the period between the perturbation and the attainment of a new state.

Imagine a simple physical system—a spring fastened to a ceiling. By attaching objects of known mass, it would be possible to calibrate this spring to establish a relationship between its length and the mass of the object attached to it. This
is the principle behind a spring scale. Clearly, such a calibration would be valid only once the spring-object system had stopped bouncing, i.e., had reached an equilibrium. No one would consider trying to calibrate it as it bounced around immediately after a new mass was added. The damped oscillations it would show after a new object was added are a simple example of transient dynamics, i.e., the dynamics of a system as it goes from one state to a new experimentally imposed state.

Because of the age and size dependence of birth and death in most species, the number of interacting species, the temporal and spatial variability of the physical habitat, time lags, feedback effects, and a variety of other processes, ecological systems are much more complex than a spring hanging from a ceiling. They are likely to have complex dynamic responses to almost any experimental manipulation. The duration of such transient responses depends on numerous factors, including the reproductive rates and life spans of the organisms in relation to the duration of the time lags, the intensity of their interactions, and the loss rates the different species experience. The appropriate amount of time for an experiment to proceed depends on the hypothesis being tested. Hypotheses can be formulated to deal directly with the immediate transient dynamic responses to an experimental manipulation as well as with the new state that may be attained after the period of transient dynamics has passed. Unfortunately, few ecological experiments have addressed the question of the appropriate length of time for an experiment to proceed in order for it to test the hypothesis of interest. Most field experiments have been based on loose, verbal hypotheses, and most of them have been interested implicitly in the new state a system might attain after a manipulation, not in its transient dynamics. As such, they have been attempts to determine if a treatment would move a variable of interest from one state (its pretreatment average, usually estimated by the controls) to a new state (its post-treatment average).

Have the ecological field experiments that have been performed to date been of sufficient duration to test the hypothesis upon which they have been based? There is no clear answer to this question, but two lines of evidence suggest that they may have been much too short. The first line of evidence comes from my sampling of papers published in *Ecology* during the past decade. Of the 180 papers reporting manipulative experiments performed in the field, 40% lasted less than one year (generally a single field season) and 86% lasted three or fewer field seasons. Only 7% of these studies lasted 5 or more field seasons (Fig. 6.1). This pattern means that, of the total of 749 papers included in my sample of the past 10 years of *Ecology*, only 1.7% of the studies were field experiments that lasted at least 5 field seasons. (Interestingly, the authors of these long-term experimental studies are well-known ecologists, including Jim Brown, Nelson Hairston, Sr., Gene Likens, Robert Paine, and John Weins. However, such correspondence between ecological fame and performing long-term experiments, because it is based on observational data, does not prove a cause and effect relationship between long-term experimentation and renown.)
What are the dynamics of responses to various experimental treatments? How might the conclusions reached in a given study depend on the length of time that a study was performed? We will not know the answers to these questions until many more long-term experiments have been performed, but the few that have been done suggest that even a 5-year study may be too short.

For instance, the classic trenching experiments that Toumey and Kienholz (1931) performed showed that trenching had a dramatic effect on the understory vegetation. This effect led them to conclude that the overstory trees were inhibiting the growth of understory plants through competition for a soil resource. What they observed, though, were species responses that changed from year to year (Toumey and Kienholz, 1931, pp. 29-30):

A distinct succession occurred on the trenched plot from 1922 to 1930. The chief points in this succession are

a. An increased luxuriance—abundance, density, and number of species—during that period.

b. The gradual change from a complete dominance of herbaceous vegetation to an increasing dominance of woody vegetation.

c. The gradual coming in of mesic species such as violet, balsam fir, willow, etc.
Some species, important at first (grass, *Lobelia*), had almost completely disappeared by 1930. Some species, absent at first, came in later, but have now almost disappeared (sorrel). Some species, absent at first, came in later in very large numbers and have remained (white violet, aster, goldenrod).

Lutz (1945) resampled these plots in 1943, 21 years after they had been trenched. He found that hemlock had displaced most other species, including white pine, from the understory. Species richness had steadily increased during the Tourney and Kienholz observations, but fell dramatically during the next 13 years (Fig. 6.2). Lutz (1945) concluded that these changes contradicted many of the generalizations that Tourney and Kienholz had made after 8 years of observation. Lutz (1931, p. 202) stated:

Tourney and Kienholz have already utilized the data from trenched plots in the Yale Forest to demonstrate the significance of root competition as an ecological force. Little need be added to their 1931 report on this point. However, the examination in 1943 has made it clear that soil moisture and root competition, important as they are, do not explain why hemlock has persisted and grown vigorously while the white pine has been dying (and is now dead) ... The conclusion seems evident that radiation intensity was too low for the relatively shade-intolerant white pine but was favorable for the shade-tolerant hemlock. Thus, while plot 5 illustrates strikingly the importance of root competition it illustrates with equal force the importance of solar radiation.

This classic study shows that a simple manipulation, trenching, led to a whole sequence of changes in the composition of that community. If the work had been done for one year, it could have been concluded that trenching, which reduced the severity of root competition, favored grasses. At 4 or 6 years, the main effect would have been to favor sorrel, at 8 years, blackberry, and at 21 years, hemlock. As Tourney and Kienholz noted, their manipulation led to a sequence of events much like a succession. These successional dynamics are the transient dynamics caused by the manipulation.

Jim Brown and his co-workers have long-term observations on the effects of experimental removals of various small mammal species from plots in the Sonoran and Chihuahuan Deserts of Arizona (Brown et al., 1986). One of the more striking features of the experiments are the transient dynamics that they have observed following manipulation (Brown et al., 1986):

A simple perturbation, such as the removal of rodents, sets in motion a complex series of changes that ripple through the community, affecting an increasing number of species. [We] are still observing pronounced changes in plants and other organisms at least seven years after exclusion of rodents began. (p. 58)

We were surprised to observe the long time lags, highly asymmetrical relationships, slight compensation for absent species, and substantial competition between distantly related taxa that our manipulations have clearly demonstrated. In retrospect, it is easy to come up with realistic hypotheses to explain these results, but these just emphasize how naive and unrealistic our initial ideas were. (p. 57)

Their work, again, begs the question of what the "result" of an ecological experiment is. The result is time-dependent. It is time-dependent because their manipu-
lation, the removal of rodents, led to a period of readjustments in the densities of other small mammal species, of several ant species, and of numerous plant species.

I have performed several large, well-replicated field manipulation experiments that are now in their sixth year at Cedar Creek Natural History Area (Tilman, 1987a, 1988). One of the most striking features of these experiments is the dynamic nature of the response. A treatment, such as annual nitrogen addition, led to highly statistically significant responses of species in every year of the experiment. However, the qualitative pattern of the species' response changed from year to year (Tilman, 1987a). For instance, vegetation in a prairie opening of a stand of undisturbed, native oak savanna showed quite different responses from the first through the fifth years of experimental nitrogen addition (Fig. 6.3). *Poa pratensis* increased significantly in abundance following N addition compared to controls in the first year of the experiment, increased even more the second year, and then declined. By the fifth year of the experiment, it was significantly less abundant in the N addition replicates than in the controls. *Panicum perlongum* increased four fold in its biomass the first year following N addition, and then decreased to a level that was more than ten fold lower than the controls. *Ambrosia coronopifolia* increased seven fold following N addition, and then declined to its former abundance. *Lathyrus venosus* was unaffected by the treatment during the first two years, increased in abundance the next two, and then
Native Oak Savanna

- **Ambrosia coronopilolia**

- **Panicum perlongum**
  - High Nitrogen Treatment
  - Unmanipulated Control

- **Sorghastrum nutans**

- **Poa pratensis**

- **Panicum oligosanthes**

- **Lathyrus venosus**

- **Carex sp.**

- **Rubus sp.**

**Absolute Abundance (g/m²)**

**Time (year)**
declined. *Rubus* sp., which became the dominant of the high nitrogen plots by 1986, did not respond at all to N addition the first year. *Panicum oligosanthes* responded the first three years by increasing in abundance, and responded the next two by decreasing in abundance. *Sorghastrum nutans* initially increased in abundance following N addition, and then fell to a much lower level than the controls. Every year of the Cedar Creek experiments, there have been species that responded significantly to the treatments. However, the pattern of their response has changed from year to year. Quite different conclusions would have been reached if the experiment had lasted one year, or three years, or five years. Indeed, these and other similar experiments performed at Cedar Creek have shown a consistent result (Tilman, 1987a): the species that were favored immediately by the treatments were invariably displaced later on. Those that increased most dramatically in response to N addition the first year were the species that were rapidly displaced by the second, third, or fourth years. All of these responses were statistically significant. Thus, the "outcome" of these experiments is time dependent. Even after five years, they have not yet reached a new "equilibrium" state imposed by the treatments. Rather, they are still in a state of flux, i.e., in a period of transient dynamics.

The Cedar Creek experiments have shown that many variables other than the biomass of individual species have transient dynamics. For instance, species richness also showed a transient response to nitrogen addition in the Cedar Creek experiments. In both the 25-year old field and in the native savanna, species richness increased significantly with the rate of nitrogen addition in the first year of the experiment, but decreased significantly with it in later years (Fig. 6.4).

The Park Grass Experiments of Rothamsted, England, are the most long-term of all ecological experiments (Lawes and Gilbert, 1880; Lawes et al., 1882; Brenchley and Warington, 1958, Tilman, 1982). These plots, which have been subjected to the same pattern of nutrient addition for the past 130 years, have shown that the dynamics of competitive displacement in such pastures are slow in comparison to most ecological studies of competition. They also show that nutrient addition can lead to a long period of transient dominance by species that are later displaced. For instance, a plot that received complete mineral fertilizer, with nitrogen applied as NH₄, was dominated by *Dactylis* for the first 15 years, with *Agrostis* briefly being the dominant before *Halecrus lanatus* displaced almost all other species (Fig. 6.5). These dynamics may be just the transient dynamics

---

**Figure 6.3.** Each part of this figure shows the mean (as a dot or square) and the standard error of the mean (as vertical bars) for the absolute abundance of a species in control plots (open squares, broken lines) and plots receiving the highest rate of nitrogen addition (solid dots, solid lines) in a prairie opening in native oak savanna at Cedar Creek Natural History Area, Minnesota. See Tilman (1987a, 1988) for a further description of the fields and experiments.
Figure 6.4. Response of species richness to different rates of nitrogen addition. See Tilman (1987a) for a description of the fields and experiments. (A)-(C). Each dot shows the species richness in an individual plot in a 25 yr old field in a particular year. Each solid line is a linear regression of species richness on the annual rate of nitrogen addition, for a given year. Note that species richness increased significantly with the rate of nitrogen addition in 1982, and decreased with the rate of nitrogen addition in subsequent years. (D)-(F). A similar pattern occurred in the nitrogen addition gradient plots in the prairie opening in native oak savanna.
of competitive displacement, but may also be caused by slow changes in soil pH in response to ammonium addition, with lower pH favoring _Holcus lanatus_ (Tilman, 1982).

These studies have shown that the "outcome" of a field manipulation experiment is time dependent. It is time dependent because all experimental manipulations are perturbations, and perturbations can lead to a period of transient dynamics. There is the distinct chance that any field manipulation experiment will lead to a period of transient dynamics. Such transient dynamics would be caused by differences in maximal growth rates, colonization rates, initial conditions, age or size structures of the populations, and a variety of other factors. Transient dynamics represent experimentally imposed "successions" that, unfortunately, have all too often been interpreted as "climaxes," and thus led to inappropriate conclusions. The ecological literature is heavily biased toward 1.2, and 3 year field experiments (Fig. 6.1), perhaps because this is the duration of most graduate student research or most research grants. The apparent power of the experimental approach has led many ecologists to put too great an emphasis on short-term experiments. Ecologists would draw very different conclusions if given the data collected after one year of experimentation than if given that collected after 3 or 5 or 20 years. Despite this fact, few experimentally-based papers have considered the potential time-dependence of the results reported.

If my results for old field herbs at Cedar Creek, Jim Brown's results for desert granivores, and other long-term studies are indicative of the prevalence of tran-
sient dynamics, it is likely that many of the conclusions that have been based upon short-term experiments may be misleading. There is a good chance that ecological experiments—in all areas of ecology—have been misinterpreted because it was assumed that a statistically significant response to a manipulation was indicative of the eventual effect of the manipulation. Such misinterpretations may have biased and confused the science of ecology. A well-replicated field experiment can give statistically significant short-term results that are the exact opposite of the eventually obtained, and equally significant, long-term results. It is critically important that ecologists recognize that short-term experiments mainly give information on transient dynamics, and that transient dynamics can be the opposite of the long-term effects of an experimental manipulation.

The appropriate length of time for an experiment depends on the life spans of the species of interest and the hypothesis being tested. Transient dynamics, if their transient nature is recognized and if they are properly interpreted, can give many insights into the functioning of a system. Unfortunately, we cannot know if experimental results are transients until an experiment has run for a sufficiently long time for the experiment to demonstrate that the results are not transients. Moreover, at our present state of understanding nature, many of the hypotheses of greatest interest to us are hypotheses concerning the average, long-term state of a system. Many processes that can lead to transient dynamics are unimportant in determining the eventual, long-term, average state of a system. The existence of transient dynamics argues strongly for the need for more long-term experimental research and for explicit theories that distinguish between transient and long-term phenomena.

**INDIRECT EFFECTS**

Brown et al. (1986) quote a passage from Darwin (1859) that showed that Darwin appreciated the complex linkages that exist between all elements of a food web. Based on his observations, Darwin hypothesized that domestic cats could be having a major, though indirect, effect on the abundance of red clover and heart-ease. This effect, he suggested, would come because cats prey on mice, mice prey on humble-bees, and humble-bees are the obligate pollinator of these flowering plants. Darwin noted that the bees were more numerous near villages, presumably because of high cat densities. The growth of interest in pairwise interspecific interactions, and ecologists' fascination with simple theories such as the Lotka-Volterra models of competition or predator-prey interactions, led many ecologists to ignore such complex links of food webs and try to interpret patterns in nature as if they were caused by interactions between a single pair of species.

However, as indicated by recent theory (e.g., Levine, 1976; Holt, 1977; Lawlor, 1979; Vandermeer, 1980; Patten, 1982; Bender et al., 1984), there are many ways that one species may influence another in a multispecies community. One species may affect another through some direct pairwise interaction, such as predation or mutualism, as well as through indirect effects that are mediated through intermediate species. Lawlor (1979), for instance, showed that, in
theory, four warbler species that only interacted directly with each other as competitors could function, in a four-species community, as if some pairs were mutualists and other pairs did not interact. A density manipulation field experiment designed to determine how one warbler species affected the other would detect the total effect of one species on another. This total effect would have two components—the direct (competitive) effect of one species on another and various indirect effects of the first species on the second that were caused by effects propagated through other species in the food web. As such, the simple type of experiment that would be directly suggested by the classical Lotka-Volterra view of competition is not an appropriate test of whether or not two species are competing in nature. As soon as there are more than two species interacting in a community—i.e., in all natural communities—there is the chance of indirect effects. Unless these are recognized and experiments are designed to allow for them, replicated, randomized, interdispersed, and controlled experiments can give statistically significant results that do not test the hypothesis posed.

There have been few, if any, field experiments that have been suitably designed and replicated to overcome the potential effects of indirect effects. Bender et al. (1984) demonstrated that it would be necessary to manipulate the densities of every species in separate, replicated experiments just to demonstrate that a given pair of species were competing, if the usual density manipulation approach were used. As an alternative, Bender et al. (1984) suggested that ecologists use short-term “pulse” experiments that look at the immediate transient dynamic response to a manipulation. They showed that two such experiments could, in theory, allow detection of a simple direct interaction between two species. However, the validity of their conclusion depends, as they pointed out, on the assumption that all interactions are caused by a direct density effect with no time lags. Competitive interactions are almost never caused by a direct density effect, but are mediated through some resource or allelopathic compound. The age and size structures of most species mean that time lags are quite likely. As such, “pulse” experiments are of limited usefulness. However, the philosophical basis of their work is critically important, because it is an attempt to design experiments that can be properly interpreted given the complexity of natural food webs and the time scales upon which organisms respond to manipulations.

Many of the field manipulation experiments that have been performed have been “designed” to determine the existence or importance of one of the classical pairwise interspecific interactions, such as competition or predation or mutualism. Schoener (1983) and Connell (1983a) reviewed more than 150 different papers that reported the results of field manipulation experiments that were “designed” to determine if a given pair of species were competing with each other in the field. In reviewing these papers, they concluded that interspecific competition was a powerful force structuring natural communities. However, these experiments were based on a hypothesis that was logically flawed, as Connell (1983b) recognized. They were based on logic derived from the Lotka-Volterra model of *pairwise* competition. They assumed that if two species were competing in nature, an increase in the density of one would lead to a decrease in the density
or growth rate of the other, and vice versa. This assumption, though, ignores indirect effects and assumes that these two species exist in an ecological vacuum, free of other species that could indirectly influence the total effect of one species on another. These 150 experiments were a noble but misguided attempt to determine the “importance” of competition in nature. In commenting on their work, I said (Tilman, 1987b, p. 771):

The 150+ density manipulation experiments that have been performed have determined the total effect of one species on another, not the mechanism of interaction. Perhaps the cases that Schoener and Connell listed as indicative of competition actually were indicative of it. Perhaps, as the work of Lawlor suggests, a much greater proportion of the species were directly competing, but the effect of their direct competition was masked by indirect effects. Perhaps many interactions were just “apparent competition,” caused by shared predators (Holt 1977). Unfortunately, even after more than 150 field manipulations of population densities, we are still uncertain.

The same criticism that I offered against those experiments applies to any field manipulation experiment that is used to infer simple cause-and-effect relations in a complex system with feedback loops. Field experiments, whatever their duration and whatever phenomena are being studied, need to be interpreted with caution. The result directly observed in a field manipulation experiment is the total effect of the manipulation. The total effect may be determined mainly by a simple cause-and-effect relation between two variables, or it may be determined by a string or web of cause-and-effect relations, many of which do not involve the two variables.

Jim Brown set out to study competition among desert rodents believing in MacArthur’s niche overlap models and expecting to see the simple, direct effects of seed competition among rodents. What his experiments have actually shown, though, is “a degree of complexity that neither the simple pairwise models nor the geographical comparisons had prepared us to expect” (Brown et al., 1986, p. 57). Much of this complexity comes from the numerous indirect effects that their experiments have detected. They have found, for instance, that the removal of particular species of seed-eating rodents causes changes in the species composition of the desert annuals, and that these changes in species composition in turn affect other species. In one case, they found that the immediate effect of a density manipulation experiment suggested that rodents were competing with ants for seed. However, as the experiments proceeded for more years, the two species were found to function as mutualists. They interpreted this mutualism as an indirect effect (Fig. 6.6). The mutualism occurred because rodents feed preferentially on large seeds and ants on small seeds. Rodents would thus reduce the density of large-seeded plants, freeing up small seeded plants from competition, allowing them to produce more seeds, and thus favoring ants. Ants have a similar effect, with increased ant density favoring large seeded plants, and thus rodents. Brown et al. (1986, p. 42) found many other indirect effects. They concluded that “Perhaps most importantly, our experiments demonstrate several kinds of strong indirect interactions, in which species influence each other through intermediary species.”
In total, the presence of indirect effects means that experiments should be designed and interpreted differently than has usually been the case. As Brown et al.’s work has shown, the actual long-term functional relationship that exists between various species in a natural community depends both on the direct interactions among various pairs of species and on indirect effects that are propagated through other species. To understand nature will require that we incorporate indirect effects into our experimental plans, and that we allow our experiments to proceed for a sufficiently long period of time we can observe the long-term total effect of one species on another.

**ENVIRONMENTAL VARIABILITY**

Weather is notoriously variable and unpredictable. An unusually dry year can be followed by an unusually wet year, etc. Rainfall, average temperature, length of the growing season, timing of vernal and autumnal frosts, and a variety of other climatic variables have all been shown to influence plant productivity, the local abundances of plant species, insect densities, and a variety of ecosystem variables and processes. Despite this dependence on climate, few, if any, field experiments have been replicated in time. Thus, we do not know if the conclusions we draw from an experiment started in one year can apply to the same experiment started in a different year. Hurlbert (1984) stressed that experimental replicates must be properly interdispersed. Ecologists do not interdisperse their experiments through time. This limits the strength of inferences that can be drawn from the experiments. For a wide variety of experiments, this difficulty could be alleviated or at least greatly reduced if the experiments were allowed to proceed for a long period of time. The longer that an experiment proceeds, the more its results represent a response to the average climatic conditions that exist in a geographic region. If long-term experiments are monitored on an annual basis, it would be possible to use a regression approach to determine the extent to which year-to-year fluctuations in experimental results could be explained by climatic variation at the same time that the experiments, as a whole, determined the long-term
average response to the treatments. Thus a long-term experiment performed in a changing environment should provide more insights than a purely observational study performed in that habitat.

Although I have never published in an agronomic journal. I have been told by several agriculturalists that results of field trials are almost never considered acceptable for publication until they have been replicated in both time and space—repeated for at least two or three years in at least two or three different fields. Dr. Orvin Burnside, Head of Agronomy at the University of Minnesota, told me that a paper reporting the results of a less well replicated experiment would be rejected by any major agronomic journal. Many agronomic studies, he stressed, last 5 to 10 years. Even longer term experiments, he asserted, are necessary now that reduced tillage is allowing perennial weeds to become permanently established in fields.

The agronomic policy of only considering a field study publishable if it has been replicated in both space and time is based on the frequently observed year-to-year and field-to-field variation in crop yields that depends on climate and substrate. I find it amazing that agronomists, who study annual plants grown in monoculture, require that at least two full life cycles of their plants be studied before results are considered publishable, whereas ecologists are willing to publish a 3-month field study of organisms that may live 100 or more years. There will always be cases in which a short-term experiment provides results of sufficiently novel interest to warrant publication. However, is there any intellectually justifiable grounds for 40% of the manipulative field experiments published by ecologists to last less than a year? A cynic might say that agricultural journals require 2 or more years of data because field trials actually matter. They are used to determine which varieties and which agricultural methods are best suited for a given region. Does the publication of such a large proportion of short-term experiments mean that ecological research does not “matter”? I believe that we owe our discipline the same respect and care that is reflected in the agronomist’s stringent requirements for field trials. Our discipline would advance more rapidly if we had fewer, but more carefully planned, more thoroughly replicated, longer term experiments than if we continue on our present course. This advance would require that each of us, as a scientist, focus on a few issues that we would spend our careers pursuing using suitably long-term observational and experimental approaches.

I asked Dr. Burnside how it was that agronomic journals could be so demanding. He stressed that faculty in agriculture have long-term research support. They are relatively free from the immediate pressures of having to find a “hot” idea in order to sell their next research project. He expressed doubt that much long-term ecological research could be accomplished until ecologists have long-term, assured funding.

Many excellent field studies, which could have contributed greatly to ecology if they had become long-term studies, have been lost because they failed to obtain a renewal of their funding. Given the low proportion of submitted grants that are currently being funded in ecology, and the unavoidably stochastic nature of the

peer review process, it is not surprising that there have been so few long-term ecological studies. The major conceptual advances that are needed if we are to deal with both fundamental and applied issues in ecology demand that we have new sources of funds available to support sustained ecological research. It is important to note that many worthwhile projects could be sustained, albeit at a minimal level, for much less than it costs to initiate them. The present method of funding ecological research has too great a stochastic element. This stochasticity assures that the average life of a project will be short, for few projects can survive a single field season without at least some minimal support for maintaining experimental treatments and collecting critical data.

**Multiple Stable Equilibria and Site History**

I want to briefly mention one other potential problem that could cause field experiments to be improperly interpreted, and that is the occurrence of multiple stable equilibria. As pointed out by Lewontin (1969), Strobeck (1973), and May (1977, 1979), simple models of multispecies communities can have multiple stable equilibria. The presence of multiple stable equilibria means that the current state of an ecosystem, even at equilibrium, depends both on the biotic interactions among the species and the ecosystem's history. It might be possible, for instance, for a grass, if initially abundant, to prevent the establishment of trees in a field, or for trees, if initially abundant, to form a closed canopy that would displace the grass. If these two different states were persistent in the absence of major disturbances, they would represent alternative stable states. The one that was present in a given locality would be determined by the area's history. A small perturbation would not change it, but a large perturbation could cause the system to flip to a different equilibrium. Clearly, the potential for ecosystems to have multiple stable equilibria is of great importance, for they imply that the long-term state of a system could be dramatically changed by a single environmental perturbation. Connell and Sousa (1983) found little evidence supporting the existence of multiple stable equilibria, but few appropriate data have yet been collected. The abilities of different tree species to invade communities dominated by various combination of other species, such as occurred during the post-glacial expansion of the last 14,000 years in North America (e.g., Davis, 1981), also suggests that initial establishment may be an unimportant determinant of the eventual composition of a community. Nonetheless, the possibility that there may be multiple stable equilibria means that it is important that experiments be well replicated and that initial conditions be determined in all plots. It is also necessary that experiments be allowed to proceed for a sufficiently long period so that their equilibrial or quasi-equilibrial outcome be observed. Different initial conditions can cause transient dynamics that could be easily misinterpreted as multiple stable equilibria.

Even if ecosystems are not found to have multiple stable equilibria, in the strict mathematical sense, the history of a research site is likely to be an important determinant of its response to natural and experimental manipulations. The
transient dynamics exhibited by an ecosystem are highly dependent on starting conditions and site history. A portion of an old field that was logged and then grazed may respond quite differently to a manipulation than another portion that was plowed after logging. Because many experiments can only proceed for a short time, and thus report transient dynamics, it is important that site history and pretreatment conditions be fully documented.

Summary

I have argued that an experimental approach is essential if we are to ever understand cause-and-effect relations in ecology. Only experimentation can separate spurious correlation from causation. However, because of the complexity of natural communities, it is easy to misinterpret experimental results. The main problem that has occurred, I believe, comes from the short duration of most field experiments. Although experimentation has become increasingly popular in ecology, most experiments have been of short duration, with almost 70% of field experiments lasting two or fewer field seasons (Fig 6.1). The few long-term experiments that have been performed suggest that the initial results of an experimental manipulation are often transient dynamics that may be unrelated to, and can be the exact opposite of, the long-term results of the experiments. Available studies of old fields and pastures, of desert small mammal communities, and of forested habitats suggest that experiments need to proceed for 10 to 50 or more years before they can show the actual long-term effects of the treatments. The shortness of existing ecological field experiments may have biased our view of the workings of nature so as to give too great an emphasis on life history traits and interactions that are important determinants of transient dynamics but less important in the long-term. Ecology needs a greater emphasis on long-term experiments if we are to understand the broad forces that structure nature.

Long-term experiments have an additional benefit: the longer an experiment, the greater are the serendipitous insights gained from it. For example, the Park Grass Experiments of Rothamsted, England, which have been in operation since 1856, were designed simply to determine the effects of different nutrients on hay production, but these experiments have also revealed that: (1) plant species are specialized on different ratios of limiting resources (Tilman, 1982); (2) plant species diversity decreases with productivity (Silvertown, 1980); (3) plants can evolve morphologies adapted to local conditions within a period as short as 100 years (Snaydon and Davies, 1972; Snaydon, 1976); (4) plant and microbial nutrient uptake can have a major effect on soil pH (Johnston et al., 1986); and (5) pH has a major impact on litter decomposition and thus soil development (Johnston, 1986). Brown et al. (1986) repeatedly emphasized that they were continually surprised by their results and that the insights they have gained from their experiments depended on them being long term.

Long-term experiments are expensive, both in their cost to operate and in the large portion of an individual's career that must be devoted to them. Long-term
experiments could be complemented by short-term experiments explicitly designed to study the transient dynamics of a system. However, transient dynamics must be interpreted properly. This will require that we develop explicit dynamic theories that can predict transient dynamics and thus allow us to properly design and interpret short-term experiments.

The lack of long-term experimentation and improper interpretation of short-term experiments may be the most important factors limiting the ability of ecologists to understand the causes of patterns in the natural world and to predict the responses of natural ecosystems to various disturbances. If ecology is to make significant advances, we must find ways to encourage long-term research, including long-term experimentation. Moreover, we must develop an approach that involves the simultaneous use of experimental, observational, and theoretical approaches. An important step would be for individuals reviewing articles for journals to determine if a study was sufficiently long-term for it to provide a reasonable test of the hypotheses posed. Because there are few meaningful ecological questions than can be answered in 1, 2, or 3 years, a higher proportion of research grants throughout all of the ecological sciences should be awarded for periods of 5 or more years. Moreover, long-term projects, when being considered for renewal, though they must be subject to the full rigors of peer review, should be protected from the fleeting whims of the bandwagons that seem to attract a share of resources that is disproportionate to their intellectual merit. By the time that an idea is perceived to be "hot" it has probably already passed the peak in what it can contribute to our discipline. It is an unavoidable fact of nature that many of the most fundamental ecological questions will require one, two, or more decades of concerted research to be resolved. Short cuts and quick fixes, though always appealing, are the Sirens’ song that have slowed significant advances in ecology.

Acknowledgments: I thank Gene Likens for his comments of this paper and Andrea Larsen for performing the survey of Ecology papers, drafting figures, and assisting with other aspects of the production of this manuscript. This research was supported by a National Science Foundation grant (BSR-8612104) for Long-Term Ecological Research at Cedar Creek Natural History Area, Minnesota.

REFERENCES

Lawes, J., J. Gilbert, and M. Masters (1882). Agricultural, botanical and chemical results of experiments on the mixed herbage of permanent meadow, conducted for more than twenty years on the same land. II. The botanical results. *Philosophical Transactions of the Royal Society* 173:1181–1413.


