chapter 1

Overview: Laboratory Experiments, Field Experiments, and Natural Experiments

Jared Diamond

INTRODUCTION

Community ecologists seek to understand species abundances and distributions. These dependent variables are ultimately controlled by two sets of independent variables: physical environmental factors and other species.

The effects of both sets of factors on individuals of a given species can either be witnessed directly (e.g., as physiological stress, acts of predation, and acts of interference competition) or inferred from other observations (e.g., inferences of exploitation competition from observations of resource utilization). While such observations of individual organisms reveal mechanisms by which the independent variables act on a species, they cannot reveal whether a species' abundance and distribution are thereby affected. To answer this population-level question, one must measure the change in abundance and distribution associated with changes in the independent variables. Experiments of various sorts are a major tool for this purpose.

Community ecologists, like astronomers and geologists, must rely on more types of experiments than do chemists or molecular biologists. Three types are traditionally distinguished: laboratory experiments (LEs), or perturbations produced by the experimenter in the laboratory; field experiments (FEs), or perturbations produced by the experimenter in the field; and natural experiments (NEs), or natural perturbations occurring in the field. In practice, LEs, FEs, and NEs form a continuum.

These three types of experiments differ greatly in ways that set what species can be used and what problems can be addressed. Some species and problems can be conveniently studied by one type of experiment but cannot be studied at all by another type. For instance, in my own research on birds I use laboratory experiments to study food processing by avian intestine, field experiments to study bower decoration by bowerbirds, and natural experiments to study bird distributions in New Guinea national parks. While in these cases the reasons dictating the type of experiment are obvious, the trade-offs are not always as clear and are often overlooked, misunderstood, or controversial. Understanding these trade-offs is essential to appropriate experimental design in community ecology and hence forms a logical starting point for this book.

This chapter begins by considering eight axes along which the types of experiments differ in their relative merits. I then discuss each type of experiment, illustrating its merits and limitations and range of utility in more detail and mentioning possible methodological improvements. The final section then synthesizes this discussion to reach a simple answer to the question, Which is the best type of experiment in community ecology? and to examine why acceptance of this obvious answer has been delayed. The relative lengths of the sections on each type should not be construed as proportional to its virtues. In particular, the strengths and weaknesses of LEs, and the strengths of FEs and NEs, are straightforward and easily described. Much more space has to be devoted to the complicated problem of limitations on FEs and NEs, and what to do about them.

EIGHT AXES FOR TRADE-OFFS

Table 1.1 summarizes, for community ecology, how the three types of experiments differ in their relative merits along each of the eight axes. (Especially laboratory experiments have different merits in other areas of ecology.) For this purpose I distinguish two subtypes of natural experiments. (a) Natural trajectory experiments (NTEs) are comparisons of the same community at various times before, during, and after a witnessed perturbation by nature or by humans other than ecologists. Examples include a storm, volcanic eruption, or the introduction or local extermination of a species. This subtype of natural experiment is discussed at length in Chapter 4. (b) Natural snapshot experiments (NSEs) are comparisons of communities assumed to have reached a quasi-steady state with respect to the perturbing variable (e.g., islands with and without a certain predator, each island having been in that state from the time of earliest observations).

In Table 1.1 I am concerned with intrinsic practical limitations of each type of experiment, rather than with common but curable deficiencies in experimental design. The assignments of Table 1.1 are sketched briefly in this section and in more detail in succeeding sections. The eight axes are as follows.

1. ***Regulation of independent variables*** Ecologists regulate actively (often, hold constant) many or all significant independent variables in LEs, one or a few variables in FEs, none in NEs.

2. ***Matching of sites*** To compensate for any lack of regulation of independent variables, ecologists attempt to minimize intersite differences in unregulated variables by three methods: replication of sites, selection of sites so as initially to have the same values of unregulated variables (except for the naturally perturbed variable in an NE), and randomization and interspersion of control and experimental sites in FEs and LEs. This matching is likely to be least successful in NSEs, where sites may be distant from each other, and where the status of sites as "experimental" or "control" (e.g., as having or lacking a certain predator) is given by nature rather than assigned by the experimenter. Matching is likely to be more successful in NTEs, where the same site

Table 1.1 COMPARISON OF THE STRENGTHS AND WEAKNESSES OF DIFFERENT TYPES OF EXPERIMENTS IN ECOLOGY

Axis	Type of experiment			
	LE	FE	NTE	NSE
1. Regulation of independent variables	Highest	Medium/low	None	None
2. Site matching	Highest	Medium	Medium/low	Lowest
3. Ability to follow trajectory	Yes	Yes	Yes	No
4. Maximum temporal scale	Lowest	Lowest	Highest	Highest
5. Maximum spatial scale	Lowest	Low	Highest	Highest
6. Scope (range of manipulations)	Lowest	Medium/low	Medium/high	Highest
7. Realism	None/low	High	Highest	Highest
8. Generality	None	Low	High	High

LE = laboratory experiment; FE = field experiment; NTE = natural trajectory experiment; NSE = natural snapshot experiment.

may be studied before and during a perturbation; still more successful in FEs, where sites may be randomly assigned and interspersed, as well as close or abutting; and most successful in LEs, where the "sites" may be uniform bottles studied at the same lab bench on the same day.

*3. **Ability to follow response trajectories resulting from perturbations*** LEs, FEs, and NTEs routinely follow response trajectories after experimental perturbations. NSEs by definition deal with snapshots or quasi-steady states and do not follow trajectories.

*4. **Maximum temporal scale*** The experimenter must exert on-going effort to keep LEs and FEs running correctly. Hence LEs and FEs are usually restricted to durations of a few years. NTEs documented by historical or fossil evidence afford the sole opportunity to follow longer response trajectories, including ones lasting millions of years (Chapter 7). NSEs permit one to examine a snapshot of a trajectory after a long time, though not the trajectory itself.

*5. **Maximum spatial scale*** Practical problems limit the spatial scale of LEs to around 0.01 hectare (the size of a constant-temperature laboratory room), and FEs usually to less than 1 ha (usually much less, occasionally more). NEs are unlimited as to spatial scale and often use large islands or continents.

*6. **Range of species and manipulations that can be studied*** As discussed in Chapter 8, one's choice of temporal and spatial scales determines what one can study. The just-mentioned extreme limitations of temporal and spatial scales, and numerous other practical problems of maintaining plant and animal populations in the laboratory, restrict LE studies of community ecology to only a tiny fraction of the species and processes existing in nature. Similar but not quite so severe restrictions due to scale, plus other practical problems and severe legal and moral constraints, knock FEs out of consideration for a large range of species manipulations, processes, and sites. NEs escape scale problems, escape legal and moral restrictions by studying an already existing situation, and are limited in practice only by the experimenter's ability to find suitable natural situations for comparison.

*7. **Realism*** By realism is meant, Are there any natural community and any natural perturbation, even a single one, to which the results of the experiment apply or can be readily extrapolated? NEs are completely realistic: no extrapolation is needed, because they already study natural communities and usually study natural perturbations. The realism of FEs is high but not complete: they too study a natural community, but the experimental perturbation may or may not mimic a natural one. LEs are usually unrealistic by intent: they mostly utilize a highly simplified and regulated community unlike any natural one. There is thus usually an acute question whether the result of an LE applies to any real community at all. (I stress again that this statement applies to community ecology; LEs are often much more realistic in physiological or behavioral ecology.)

*8. **Generality*** A result possesses realism if it is immediately known to apply to at least one community; generality, if it is immediately known to apply to many communities. One's ultimate goal is a conclusion with at least some generality, rather than one that applies to just one site in one year. LEs, because they lack realism, also lack generality. NEs and FEs both aim to compare multiple real experimental and control communities. However, NEs obtain their replicates by explicitly sampling natural variation among real communities, while FEs explicitly seek to minimize this variation and to sample adjacent, preferably identical, replicates at a single site. There is thus an acute question concerning the extent to which the result of an FE possesses generality and applies beyond that single site in that year. In effect, FEs compared to NEs gain confidence in the conclusions at one site (axes 1 and 2 above), at the cost of generality.

LABORATORY EXPERIMENTS (LEs)

Examples of LEs

LEs are the experimental approach familiar to chemists, who are often initially puzzled as to why ecologists ever resort to any other approach.

In the laboratory the experimenter can regulate the whole abiotic environment, especially the ecologically critical variables of light, temperature, water, substrate, and nutrients that vary outdoors somewhat unpredictably in space and time. The experimenter can also regulate the biological environment, and in practice most LEs are done on synthetic communities composed of a few (usually two) chosen species. The LE technique is often employed on microorganisms or small arthropods kept in bottles, hence ecologists' somewhat derisive term "bottle experiment." However, similar experiments are also often conducted on plants in greenhouses. As will be discussed in the next section, the dichotomy between LEs using indoor synthetic communities and FEs using outdoor natural communities is artificial; LEs and FEs form a continuum.

An early bottle experiment (*sensu stricto*) was by Pearl and Parker (1922), who used *Drosophila* grown in bottles to demonstrate the decline in *per capita* reproduction with increasing population density. Perhaps the most frequently cited bottle experiments are Gause's (1934) demonstrations of exploitation competition, interference competition, and predator-prey dynamics, using cultures of yeast and protozoans. Other well-known bottle experiments have studied competing beetle species (Park 1962), predator and prey mites (Huffaker 1958), and competing predatory wasps and prey beetles (Utida 1957). LEs on a larger scale are the numerous studies of plants in greenhouses (de Wit 1960, Harper 1977), such as Donald's (1958) growing of two plant species in pots with aerial or soil partitions in order to separate the contributions of root and shoot interactions to competitive depression of growth.

LEs on caged vertebrates have mostly studied behavioral and physiological responses of individual animals, not the dynamics of reproducing populations. These responses have community ecological significance, but the studies fall within the fields of behavioral or physiological ecology rather than community ecology. Examples include Sheppard's (1971) and Heller's (1971) studies of aggressive interactions between two chipmunk species that field observations had suggested to be competitors and Klopfer's (1973) use of a gymnasium filled with oak and pine boughs to examine habitat selection in birds.

The present volume includes LEs on *Drosophila*, fish, and marine invertebrates (Chapters 2, 21, and 31).

Advantages of LEs

LEs are obviously by far the best type of ecological experiment in regard to regulation of independent variables (axis 1 of Table 1.1) and site matching (axis 2, creation of uniform sites adjacent to each other on the same lab bench). Like mathematical theory, LEs take specified simple starting conditions and reveal a range of possible outcomes, which field biologists can then evaluate for the LEs' relevance to actual communities. Famous examples include the central role of bottle experiments by Pearl, Nicholson, and others in our understanding of population growth and regulation; the role of Gause's LEs in stimulating Lack and subsequent field biologists to evaluate competitive exclusion; the influence of Gause's LEs also on the theories and field studies of predator-prey dynamics; and the role of Huffaker's LEs in stimulating field biologists to evaluate the importance of environmental patchiness for predator-prey dynamics (Chapter 11). These instances suffice to make clear the enormous influence that LEs have had for coalescing the fundamental principles of ecology.

Limitations of LEs

Obviously, in most other respects LEs are equally the procedure with the most serious drawbacks, which may be summarized as extreme unrealism and extremely restricted scope. In regard to realism, real communities depend on the direct and indirect effects of a large set of species whose boundaries are hard to predict (Chapter 3). Real communities also depend on physical environmental parameters whose spatial and temporal variances are not mere noise but critical factors in community organization (Chapters 13 and 14). By sweeping away these complexities in order to study a particular process, LEs renounce the goal of illuminating particular actual communities. That is, they are unrealistic and devoid of generality (axes 7 and 8 of Table 1.1). We often have no idea whether the results of an LE apply to any natural community. For

)n *Droso-*
'hapters 2,

)f ecologi-
f indepen-
ı and site
sites adja-
ıch). Like
ed simple
)f possible
nen evalu-
ımunities.
ɔle of bot-
ınd others
rowth and
timulating
o evaluate
)f Gause's
es of pred-
Huffaker's
ıa the
s foı pred-
These in-
ıous influ-
scing the

LEs are
st serious
ed as ex-
ted scope.
depend on
ırge set of
to predict
depend on
ose spatial.
noise but
ion (Chap-
hese com-
r process,
particular
unrealistic
8 of Table
the results
unity. For

instance, the bottle experiments of Gause and his successors told nothing as to whether or not competition is important in natural populations of the yeasts and protozoans studied. One does not even know what variables are worth studying in LEs until their importance in nature has been studied by other means.

In regard to scope, the size of laboratories (axis 5), the modest number of years one can afford to run one LE (axis 4), and the difficulty or impossibility of keeping populations of most species alive and reproducing for many generations in the laboratory all act to restrict LEs to studying community ecology on only a tiny fraction of the world's species (axis 6). (It is difficult to imagine an LE-based paper entitled "Competition between populations of vultures, hyenas, and lions: a 10-generation cage experiment.") Thus, LEs must treat small-sized, short-lived organisms as models of larger ones, despite the systematic ecological differences existing between small and large organisms (Chapter 28).

For these reasons, currently LEs are perhaps the least used of the experimental traditions in community ecology. but most relevant.
for conceptual generation i.e.:

Suggested Improvements in LEs

There are at least four ways in which LEs' overwhelming advantage in regulation of independent variables and in site uniformity could be more widely utilized.

1. LEs could be immediately applied to more species. Ecologists have not bothered to do extended laboratory population studies of most suitable species.
2. LEs could be run on a larger spatial scale than is traditional. Greenhouses exist for studying plants; they could be used to maintain populations of small animals not currently studied. Construction of large indoor facilities is expensive, but could in some cases be at least as illuminating as expensive outdoor projects.
3. LEs have almost exclusively studied synthetic communities. There are important natural communities that would lend themselves to the LE approach, such as communities of microorganisms in soil brought into the laboratory.
4. The synthetic communities studied by most LEs have consisted of two species. Perhaps the most important expansion of LEs will consist of creating synthetic communities far more complex and hence far more realistic than those constructed to date. Steps in this direction include Neill's (1975) studies of four crustacean species in all possible combinations; Vandermeer's (1969) experiments on four protozoan species alone and in combinations; Stiven's (1971) studies of 2-, 3-, and 5-species communities of hydras and an amoeba pathogen; and Huffaker's (1958) analysis of predator-prey coexistence in physically complex containers. Gilpin and his colleagues (Chapter 2) take this approach (which they term community reconstitution) a major step further by measuring growth rates of 28 *Drosophila* species singly, in all pairwise combinations, and in a 10-species set and by varying temperature, food quantity, and substrate texture. Such LEs could play a leading role in refining the questions asked by field ecologists, sensitizing ecologists to the kinds of complex outcomes worth looking for in the field, and guiding the design of better FEs and NEs.

FIELD EXPERIMENTS (FEs)

Examples of FEs

FEs differ from LEs in that they are conducted outdoors, and typically also in that they operate on natural rather than synthetic communities. The experimenter usually manipulates only one or a few independent variables. The most common manipulations consist of locally eliminating a species, locally introducing a species (e.g., to a small island), or erecting a fence or cage. The experimenter effectively selects initial values of other independent variables through site selection but does not hold them constant or regulate their trajectories thereafter, in contrast to LEs. There is, of course, a risk that any difference observed in the dependent variable's trajectory between the manipulated and unmanipulated plots is not due to the manipulation but to the natural differences expected between any two field plots. To reduce this risk, the experimenter often arranges replicate manipulated and unmanipulated plots in a randomized, interspersed spatial array.

FEs and LEs intergrade along a continuum, depending on the degrees to which the physical environment is regulated and to which species community composition is synthetic or natural (cf. Schoener 1983b, p. 242). Hybrid studies in which an experimenter creates identical adjacent synthetic communities outdoors instead of seeking similar adjacent natural communities include agricultural plots, numerous other outdoor studies of plant species grown in particular combinations (Chapter 12), and studies of artificial ponds stocked with particular fish species (Chapter 21). Wilbur and Travis (1984) created artificial ponds and allowed them to be colonized naturally by species. It becomes arbitrary where to draw the line among such experiments between LEs and FEs. FEs and NTEs also intergrade: the same manipulation may be launched by an experimental ecologist, by someone else, or by accident (p. 13).

One of the first FEs was performed by Darwin (1859), who demonstrated that mowing or the introduction of grazing animals increases plant species diversity on a lawn (by preventing some species from outcompeting others). Well-known modern FEs have removed limpets to demonstrate their grazing impact on algae (Jones 1948), removed or fenced-out starfish or predatory snails to test their impact on barnacle or mussel prey (Connell 1961a, Paine 1966), removed territory-holding birds to reveal the existence of nonbreeding surplus birds (Hensley and Cope 1951), fumigated mangrove trees to monitor recolonization by arthropods (Wilson and Simberloff 1969), and fenced field mice and noted effects of dispersal on population density (Krebs et al. 1969). The most extensive introduction experiment is by Schoener and Schoener (1983c; see also Chapter 33), who placed lizards on 30 Bahamian islets without natural lizard populations and thereby demonstrated that lizards could reproduce successfully on islets much smaller than those supporting lizards naturally. The present volume includes FEs on terrestrial arthropods (Chapters 3, 11, 24, 25, 26, and 33), rodents (Chapter 3), lizards (Chapters 30 and 33), and plants (Chapters 12, 22, and 25); fresh-water and marine fish (Chapter 21); and marine plants (Chapter 32) and invertebrates (Chapters 30 and 31).

Advantages of FEs

In their merits and drawbacks FEs are intermediate between LEs and NEs along almost all of the eight axes (Table 1.1). There is no single axis along which FEs are the most advantageous type, and also no axis along which they are the least advantageous type. Broadly speaking, FEs purchase gains in realism, spatial scale, and scope over LEs at the cost of losses in regulatory control and site matching; they purchase gains in regulatory control and site matching over NEs at the cost of losses in temporal and spatial scale, scope, and generality. The concluding section of Chapter 3 assesses the advantages and limitations of FEs in the light of a long-term project on desert granivores.

In more detail, compared to LEs, FEs acquire realism by working with actual communities. FEs bring a modest gain in maximum spatial scale, from about 0.01 to 1 ha. (However, there are some larger FEs; perhaps the largest is the experimental isolation of a 100-ha Amazonian forest patch by Lovejoy et al. (1984).) Because of the gains in use of actual communities and larger spatial scale, FEs can study far more species than can LEs. For example, FEs regularly use not just herbs but also trees, and not just yeasts but also rodents.

Vis-à-vis NEs that compare sites naturally having or lacking a particular species, FEs that intentionally introduce or eliminate the species on randomly chosen plots have a major advantage: the species' presence or absence could not be due to some preexisting difference between the sites. Other small-scale manipulations of independent variables can be clearly achieved by FEs when and where one wishes them, including manipulations for which there is no close NE equivalent. For instance, I cannot think of an NE that would have revealed the role of dispersal in microtine population cycles as did the fences of

), rodents
1 33), and
-water and
ine plants
ers 30 and

intermedi-
all of the
ingle axis
eous type,
the least
FEs pur-
nd scope
tory con-
ns in reg-
NEs at the
ial scale,
section of
mitations
roj on

s acquire
munities.
n spatial
er, there
st is the
nazonian
Because
ities and
nore spe-
egularly
not just

naturally
FEs that
species
r advan-
ould not
between
ns of in-
eved by
ncluding
ose NE
f an NE
ersal in
nces of

Krebs et al. (1969), or an NE that would have removed territory-holding birds as cleanly as did the shotguns of Hensley and Cope (1951). NEs and FEs both choose replicate sites, but FEs have the advantage that the sites are likely to be much closer together as well as randomized and interspersed, hence much better matched for independent variables other than the one being intentionally varied.

Limitations of FEs

Limitations Compared to LEs A major sacrifice that FEs make for these gains vis-à-vis LEs is in the matching of sites (axis 2 of Table 1.1). Two types of problems are involved. First, a curable problem: ecologists have started to make widespread use of FEs only recently, with the 10 result that FEs are often still plagued by pseudoreplication of sites (Hurlbert 1984). Second, even in the best designed FEs, there are practical limits on site replication and selection. Sites in spatially heterogenous environments are likely to be poorly matched, and replication is often impossible due to limited resources or because large sites are being studied. For example, in Tinkle's (1982) pioneering study of competition among Arizonan lizards in riparian woodland, he and his field assistants were able to establish one control site of 0.67 ha and one adjacent experimental site of 0.96 ha, to perform repeated removals of two lizard species on the latter site, and to mark 1,244 individual lizards and make many ecological and morphological measurements on them. The control site had more trees, fewer large trees, more rocks, greater structural diversity, and initially denser lizard populations than the experimental site. One of the largest and most important FEs is the Minimum Critical Size project of the World Wildlife Fund and the Brazilian government; currently there is one experimental plot of 100 ha, three of 10 ha, and three of 1 ha in the Amazonian rainforest and there will be others, including one of 10,000 ha (Lovejoy et al. 1984, Lewin 1984). The siting of such large plots in a landscape being cleared for cattle ranching involves many practical considerations other than ones of site matching. It should be emphasized that the FEs of Tinkle (1982) and Lovejoy et al. (1984) are exceptionally well executed and in no way exemplify curable problems of poor design. Rather, they illustrate inevitable practical constraints on large, expensive, labor-intensive field experiments. It is obvious that LEs escape these difficulties of site nonuniformity, and also that FEs are not always able to reap the benefits of their potential advantage in site matching over NEs. In particular, some NTEs can provide better matched experimental and control sites than can large FEs (as will be discussed in the section on p. 13).

The other inherent limitation of FEs compared to LEs is that the experimenter cannot regulate most independent variables, including temperature, rainfall, light, wind, and abundances of many other species. Hence the outcome of an FE may vary with year, season, or geographical location, because the outcome depends on unregulated variables affecting both experimental and control sites. For instance, the removal of desert lizards (Dunham 1980, Smith 1981) and rodents (Morris and Grant 1972) either succeeded or failed completely to affect abundances of competing species, apparently depending on rainfall and hence food availability in the particular year of the experiment. Schoener's (1983b) review of FEs found 11 other studies showing year-to-year variation in the effect of competition. The outcome of removal experiments on fresh-water crustacea varied among seasons of the same year for unknown reasons (Lynch 1978). The frequency of such variable outcomes in FEs is unknown, because FEs are rarely run for enough generations of the species studied even to test for the possibility of such variation (Schoener 1983b). Outcomes may also vary geographically: reciprocal effects of removing ants and rodents on each others' abundance were clear at one Sonoran Desert site and one Chihuahuan Desert site, but unclear or undetectable at another Chihuahuan Desert site (Chapter 3).

In short, because most independent variables cannot be regulated in FEs, it is harder for FEs than for LEs to obtain a reproducible result or to identify the explanation for a varying result. Note that this fact does not mean that FEs are categorically inferior to LEs, but simply that these two methods are useful for different purposes. Variable outcomes due to variation in nature are a *realistic* finding of FEs; the finding does apply to that site at that time.

Limitations Compared to NEs Vis-à-vis NEs, FEs retain four disadvantages of LEs: the same practical limitations of temporal scale, only a modest expansion of spatial scale, restrictions (though less severe ones) on the range of manipulations that can be studied, and problems of generality that will be discussed later in the section on NEs (p. 14).

Consider first the problems of scale. Schoener's (1983b, Fig. 1) depiction of the duration of 164 FEs showed only four studies that lasted over 5 years. Median durations were around 2, 8, 9, and 16 months for FEs on freshwater species, marine species, terrestrial animals, and terrestrial plants, respectively. FEs on a spatial scale exceeding 1 ha are rare. By being condemned to operate on these tiny scales, FEs are blind to whole classes of phenomena detectable by NTEs or NSEs: for instance, genetic changes (evolutionary responses); phenomena that emerge on large spatial scales and that arise from such factors as population patchiness, immigration-extinction dynamics, properties of closed as opposed to open systems, and biogeographic tests of evolutionary potential; population responses in species with long generation times; and chains of indirect effects extending beyond a decade. These and other consequences of temporal and spatial scale are discussed at length in Chapter 8.

A further consequence of the limited temporal and spatial scale of FEs is in rendering them less able than NEs, for two reasons, to follow a perturbation to its equilibrial or steady-state result. First, most FEs examine a small site that is simply fenced off or otherwise distinguished from the community in which it is imbedded. Few FEs study an arena sufficiently large or isolated to make the question of equilibrium even meaningful. Instead, the best that one can hope to attain is a quasi-steady state between internal processes and boundary processes. Unusual in this respect are a few FEs that introduced lizards, spiders, or mammals to islands sufficiently isolated and large to support closed, self-sustaining populations (Chapter 33; Crowell 1983, Schoener and Schoener 1983c). Second, major sustained perturbations to a community are likely to produce a chain of indirect effects lasting too long for an FE to follow to completion. For instance, enclosures from which Brown and colleagues fenced out desert rodents had not reached a steady state after 7 years (Chapter 3). As another example, if interspecific competition is compatible with coexistence in most years but eliminates one species in the rare year of low resource levels, the result in a closed system after a long time would be competitive exclusion. This result would be detectable in an NSE but only rarely in a closed FE, and in the usual open FE immigration might prevent the result from occurring at all.

Practical restrictions on the scope of FEs are of at least eight sorts.

(1) One cannot introduce mobile species that would simply leave, such as birds, other volant species, and small-bodied species impossible to fence or cage. (2) One cannot introduce species requiring a large territory, such as top carnivores, and obtain a self-sustaining population or even a realistically foraging individual in a small enclosure. (3) Species removals in the field have yet to be reported and may be virtually impossible for many types of species (e.g., internal parasites, abundant small-bodied species, subterranean species, hard-to-catch volant species) and for many habitats (e.g., forest canopy). Removal experiments on terrestrial animals are virtually confined to low habitats and strata, such as forest floor and understory, desert, and grassland. FEs in the world's most species-rich communities, the tropical rainforests, are few. (4) Many species would be illegal or immoral to introduce (many predators, pathogens, parasites) or to remove (many birds and mammals). (5) The techniques available for fencing species in or out select entire groups of species on the basis of characteristics such as size (cf. the rodent fences used in Chapter 3) or inability to dig, crawl, or jump past fences. Fence techniques for selecting any arbitrary species or combinations of species are unavailable. (6) FEs cannot realistically simulate important natural disturbances, such as

g popula-
oener and
ined per-
produce a
for an FE
nclosures
nced out
state after
, if inter-
coexist-
pecies in
esult in a
compet-
ctable in
nd in the
nt the re-

FEs are

cies that
r volant
ssible to
species
nivores,
r even a
ll
have yet
possible
al para-
ubterra-
ies) and
Removal
virtually
as forest
nd. FEs
unities,
ny spe-
ntroduce
r to re-
(5) The
n or out
basis of
t fences
awl, or
electing
species
stically
such as

hurricanes, droughts, and frosts. (7) Few long-term FEs requiring on-going attention have been executed at sites far from biologists' homes. In particular, nearly all FEs are done in the temperate zones and on continents, and very few in the tropics or on islands. (8) FEs are usually forbidden in nature reserves and national parks, such as the sites for the studies of New Guinea and Galápagos birds described elsewhere in this book (Chapters 6 and 10). This limitation on FEs will grow in importance as more and more biological communities become restricted to reserves.

Thus, for many (probably most) individual species and most manipulations in many places, FEs are scarcely possible for practical, legal, or moral reasons.

Suggested Improvements in FEs

This section on FEs concludes with 10 suggestions for expanding their scope and power.

1. There is still much scope for ingenuity in broadening the range of species and problems studied by FEs. Just one example: most FEs have introduced or removed species singly, while study of the interesting questions posed by varying combinations of species has remained the exclusive domain of LEs and NEs (e.g., Chapter 2; Schoener 1975, Diamond 1975). Perhaps some of the practical restrictions just outlined in the preceding section could be overcome through ingenuity.

2. There is especially scope for designing LE-FE hybrids that employ prepared outdoor sites, or synthetic outdoor communities, or partial regulation of environmental variables such as water and light. For some purposes such hybrids will be more illuminating than pure LEs or pure FEs.

3. FEs often still employ poor experimental designs. For instance, an important paper by Hurlbert (1984) discusses at length the curable problems of site replication. Another example is the underutilization of partial factorial designs. As illustrated by Table 3.1 in Chapter 3, such designs permit one to increase the effective number of replicates of experimental plots and to reduce the required number of control plots.

4. Ecologists are still confused about the distinction between "pulse experiments," in which one briefly applies a perturbation and then watches the system relax, and the much commoner "press experiments," in which one applies a sustained perturbation. These two types of FEs have very different interpretations (Bender et al. 1984). For example, many of the species interactions revealed by press experiments are not direct effects, as the experimenter often believes, but chains of indirect effects (Chapters 3 and 20; Bender et al. 1984); additional experiments of the pulse type may have to be run to distinguish these types of effects.

5. Experimental interventions risk introducing artifacts. That is, an experimenter making an intervention presumes to know what salient feature of the intervention causes the observed end result. However, the salient feature may instead be some unforeseen or undetected feature or proximal consequence of the intervention. This risk is modest when a particular species is manipulated by hand, but is large when a fence or cage is erected to keep a particular species in or out, because fences can produce many other effects and can influence many other species (p. 15). In FE-based papers that employ fences, the evidence provided to discriminate among possible causes of an observed fence effect is often disappointingly incomplete. Some of the "epidemiological" criteria for clarifying a problem of causation in NSEs (p. 16) may also be applicable to this problem of causation in FEs.

6. Many FEs could be profitably run for longer than is currently usual, both to reveal the chains of indirect effects, to test for year-to-year variation in the outcome, and to search for outcomes (such as competitive exclusion) likely to be consummated only in an exceptional year.

7. More FEs could be run on closed natural communities, such as small islands, to study equilibria little perturbed by boundary processes.

artifacts?

8. One approach to the problem of how to generalize from an FE at a single site was employed by Lubchenco (Chapter 32). She identified different types of sites in an area, used FEs to assess species interactions at each site type, and then weighted the results by the overall frequency of each type.

9. Another approach to the problem of generality would be to integrate FEs much more closely with NEs. For example, an NE might indicate that the habitat range of species A is restricted by a competing species B present at site 1 but not at site 2; further evidence could be obtained from an FE (a local removal of species A) at site 1. As noted by Connell (1980), integration of FEs with NEs may provide an excellent means to study postulated evolutionary changes in strengths of competition (see Chapter 6 for discussion); the FE-NSE by Abramsky and Sellah (1982) represents a promising start.

10. Many ecologists feel defensive about whether ecology is a rigorous science and are determined to avoid anything smacking of nonrigor. In particular, it is often assumed that rigorous science consists exclusively of experiments carried out to test à priori hypotheses, while experiments carried out simply to see how the system responds are intellectually inferior, and nonexperimental observations unspeakably worse. This attitude grew out of a healthy reaction to ecology's earlier paucity of hypothesis testing and domination by descriptive studies. However, the most profitable approach in science depends on one's state of knowledge. With a little known system one has to begin by describing it, examining its natural variation, and poking it to see how it behaves before one can hope to frame more detailed hypotheses intelligently. Ecological communities are complicated, and our understanding of their connectivity is as rudimentary as is our experience of FEs. Thus, FEs of the "poke-it" type on even the best studied communities may still yield major surprises. (An example discussed for desert granivore communities in Chapter 3 is the change in fungal infections of plants resulting from manipulations of rodent densities.) I do not wish to be misconstrued as implying that hypothesis testing has no place in FEs at present. However, obsession with hypothesis testing should not blind us to the possible value as well of some manipulations unconstrained by prior hypotheses.

NATURAL EXPERIMENTS (NEs)

Examples of NEs

Natural experiments differ from FEs and LEs in that the experimenter does not establish the perturbation but instead selects sites where the perturbation is already running or has run. The perturbation may have been initiated naturally or by humans other than an experimental ecologist. Along with the experimental sites, the investigator selects control sites so that the two types of sites differ in the presence or absence of the perturbation but are as similar as possible in other respects.

NEs fall into two categories: natural snapshot experiments (NSEs), in which one observes only a final steady state or other snapshot of an old system, but not the trajectory that led to it, and natural trajectory experiments (NTEs), in which one observes the trajectory or can reconstruct it from historical or fossil records. Typical NTEs examine the trajectory following a natural disturbance (volcanic explosion, freeze, drought), a human-made disturbance (fire, eutrophication), or an invasion or extinction of a species. Chapter 4 will discuss and illustrate NTEs in detail. In this volume they are utilized in studies of phytophagous mites, trees, birds, and fish (Chapters 5, 10, 16, 21, and 24).

Typical NSEs are studies comparing the abundance, morphology, and habitat range of species A on multiple islands or at multiple sites, some of which have and others of which lack a competing or predator species B. For instance, Schoener and Toft (1983a; see also Chapter 33) found that spiders averaged about eleven times more abundant on 48 Bahamian islands without lizards than on 26 islands with lizards, because lizards prey on and also compete with spiders. Where char and trout occur sympatrically in a lake, each species has a narrower diet and occupies a narrower range of habitats (char in deeper water than trout) than in allopatry. These niche shifts, which are maximal at late-summer times of low resource levels, have been observed in dozens of Scandinavian and North American lakes with two different trout species and two char species (Chapter 21). In this volume NSEs are utilized in studies of desert and Mediterranean scrub plants, spiders, fish, lizards, and birds (Chapters 6, 10, 21, 23, 30, and 33).

NSEs and NTEs intergrade, depending on whether the initial state before the perturbation was observed or could be reconstructed, and on how well the trajectory from initial to present state is documented. NTEs and FEs also intergrade: a species may be introduced or locally removed experimentally by an ecologist, accidentally or intentionally by another human, by some indirect effect of humans (e.g., an introduced disease), or by a natural cause. At first one might expect a sharp distinction between FEs and NTEs, in that FEs have clean layout of experimental and control plots (randomized interspersed design), while NTEs do not. In practice, for large-scale FEs it may be impossible to have good matching of experimental and control plots, or to replicate the experimental plots, or to have a control plot at all. By contrast, site matching in some NTEs may be much cleaner. For example, when the peacock bass, an exotic piscivorous fish, reached Panama's Gatun Lake after introduction into a tributary river by a local businessman, it spread throughout the lake during the course of 5 years as a slow wave with a leading edge of subadults. Zaret and Paine (1973) were able to map the bass's advance from year to year and to deduce its effects from two types of comparisons: approximately bimonthly samples of community composition at several stations for 4 years before and 2 years after arrival of the bass front; and simultaneous comparison of sites already reached or not yet reached by bass, and matched in habitat. Both comparisons revealed similar dramatic effects of bass on prey fish, on bird and fish predators of the prey fish, and on insect and zooplankton and fish prey of the prey fish. It is doubtful that an FE (experimentally motivated introduction) in a lake as large as Lake Gatun would have achieved a cleaner or different design.

Advantages of NEs

In their merits and drawbacks NEs are at the opposite pole from LEs. NEs are the most advantageous experimental procedure in all five respects in which LEs are the most seriously flawed: temporal scale, spatial scale, scope, realism, and generality. Conversely, NEs are the worst procedure in the two respects in which LEs are the most advantageous: regulation of independent variables and site matching. NEs show most of these same merits and drawbacks in comparison with FEs, but to a less marked degree.

The major advantage of NEs is that they provide the sole means to study many perturbations and questions. This arises from NEs' combined advantages of temporal scale, spatial scale, and scope. NEs not only use tiny islands similar in size to the typical FE, but also routinely operate on much larger spatial scales than are possible for FEs or LEs: large islands and even continents. NTEs are the sole technique for following the trajectory of a perturbation beyond a few decades; even durations of one decade are costly and hence very rare for FEs. For example, historical or fossil evidence lets one reconstruct successional or evolutionary trajectories of about 50, 70, 2000, 3 million, and 70 million years resulting from the following perturbations, respectively: elimination of American Chestnut by chestnut blight, insularization of Barro Colorado (Willis and Eisenmann 1979), the sudden crash of hemlock in the northeastern United States 4800 B.P. (Davis 1981a), the closing of the Panama Seaway that launched the Great American Interchange (Marshall et al. 1982), and the extinction of all large terrestrial vertebrates at the Cretaceous-Tertiary boundary. While NSEs do not follow trajectories at all, they do let one examine snapshots reached after long times. For instance, NSEs comparing existing forest fragments of various sizes enabled Willis (1980) and Whitcomb et al. (1981) to deduce how a century of forest fragmentation affected bird distributions, while an FE studying the same problems in Amazonia has only just begun (Lovejoy et al. 1984). The modern mammal fauna of Tasmania and neighboring islands is a snapshot produced by 10,000 years of differential extinction following the sundering of late-Pleistocene land bridges (Hope 1973, Diamond 1984a).

As discussed in Chapter 8, the NEs' expanded spatial and temporal scales open up for study a whole range of problems (including evolutionary ones) that are inaccessible to FEs and LEs. Because of these scale effects and the other practical constraints on FEs previously listed, NEs are the sole way to examine perturbations that cannot, may not, or should not be created deliberately. Chapter 4 describes the species introductions and extinctions available for study through NTEs.

NEs have the further virtue of generality, because they sample a wider range of natural variation among sites than do FEs (LEs do not sample natural variation at all). To set up and maintain an FE involves effort, as in building and maintaining enclosures or removing species. The experimenter deliberately concentrates that effort on immediately adjacent sites (so that control and experimental sites will be well matched), and deliberately samples only a tiny fraction of natural variation among sites. This strategy increases confidence that the conclusion obtained does apply to that study site, but reduces confidence that the conclusion applies to other sites, as the study site might be atypical. NEs instead put effort into studying geographically scattered sites and thereby sample a greater range of natural variation. Any difference that emerges between sites with and without some particular factor is therefore more likely to be valid for a considerable range of sites. For example, Toft and Schoener (1983) were able in 22 days to census five spider species on 116 islands with and without lizards, and to establish that spiders were on the average an order of magnitude more abundant on islands without than with lizards. The sampled islands covered a range of areas, distances, and vegetation heights and complexities, which Toft and Schoener measured in order to disentangle the effect of lizard presence or absence from concurrent effects of these other variables. In the same field time they would have been able to remove most (not all) individual lizards on only 2 islands supporting them, or to introduce a saturating number of lizards to only a few islands lacking them; they would still have had to wait up to several years for spider densities to reach new equilibrium values on the manipulated islands; and, without an NE, they might have been unlucky and picked exceptional islands where lizards did not happen to limit spiders.

Limitations of NEs

Set against these major advantages over FEs and LEs (plus the advantage of realism over LEs), NEs suffer to varying degrees from several major disadvantages.

First, NSEs by definition are NEs where one knows only a single snapshot of a trajectory, not the whole trajectory. As stressed in Chapter 3, obviously this constitutes an enormous loss of information compared to NTEs, FEs, and LEs. The lost information includes not only length of time between causes and effects, but also event sequences that are often decisive in tracing out cause-effect chains. As Brown and colleagues (Chapter 3) put it, "the temporal sequence of events reveals a great deal about the processes by which species affect each other and about the patterns of connectance that link the fates of species."

Second, like FEs, NEs try to minimize differences between experimental and control sites with respect to variables other than the one whose effect is to be examined, by comparing many sites of each type matched for those other variables. However, this matching is likely to be less close for NEs than for FEs, because sites are given by nature rather than selected, and geographically scattered rather than adjacent. This consideration is more of a problem for NSEs than for NTEs: with NTEs one can compare the same site before and after perturbation, as well as perturbed and unperturbed sites at the same time. Thus, the sequence with regard to closeness of site matching tends to be LE > FE > NTE > NSE.

The remaining drawback of NSEs is the one most troublesome to many ecologists: the perturbation is only inferred, not observed or experimentally created. In FEs the experimenter applies the perturbation and randomly assigns sites to be perturbed or not. In NTEs the perturbation is also generally some unequivocally identified event, such as an eruption or the arrival of a disease. In the case of NSEs, however, nature provided the two sets of sites (e.g., those with and without species A). If one observes that species B is more abundant in the absence than in the presence of species A, that could mean that species A itself somehow reduces species B's abundance (e.g., by competing with or preying on it). A necessary (but not sufficient) condition for the correctness of that interpretation is that A's presence or absence at each site must be due to factors of no direct relevance to B's abundance, such as local historical accidents of A's arrival and disappearance. However, it may also be that the two sets of sites differed in some other factor that favored A's presence *and* B's rareness; or (less plausibly in most actual cases) that it was B's abundance that caused A's absence. ''There is no certainty that the only difference between the two mountains is the absence of one species. A predator may have been absent, or an essential food organism, soil nutrient, etc., may have been present beyond the boundary on the second mountain'' (Connell 1975, p. 462).

In statistical terms, an NSE yields a correlation between two observed differences that distinguish two sets of sites. Other evidence must be examined to decide what is cause and what is effect. This particular burden of extra evidence does not arise with LEs, FEs, and NTEs, where the manipulation itself is known to be somehow the cause of the differences between the two types of sites. Hence there is no doubt that, all other things being equal, a controlled perturbation provides clearer evidence than a preexisting difference between sites.

In practice, I see this extra burden of NSEs as a matter of degree, rather than as the absolute distinction that the contrasting words ''cause'' and ''correlation'' suggest. Often, the manipulation employed in an FE does not mimic a natural perturbation but is instead a means used to produce a proximate effect mimicking a natural perturbation. However, the manipulation is actually likely to produce a cluster of proximate effects. While the unnatural manipulation itself surely somehow led to the observed ultimate effects, that fact is not biologically interesting. The interesting question is which proximate effect produced the ultimate effect. Answering this question involves a burden of extra evidence in some but not all FEs. For instance, one of the commonest manipulations in an FE is to fence or cage the experimental sites. Fences and cages produce at least nine types of effects, including preventing the monitored species inside (the primary dependent variable) from emigrating, preventing some other species inside from emigrating, preventing a predator outside from entering, preventing a competitor from entering, preventing a herbivore or a species in some other trophic relationship from entering, providing a perch for hawks or other animals, providing a surface for attachment, providing some protection against wave action in the intertidal, and providing shade. Among six FE-based studies employing fences or cages Chapters 3, 11, 31, Connell (1961a), Paine (1966), and Krebs et al. (1969) attributed the results mainly to the fourth, third, fifth, third, third, and first of these factors respectively, but the decision in each case poses a burden of evidence and may not be easy.

?

but

Suggested Improvements in NTEs

NTEs offer a large and underutilized body of clean-cut ecological data (Chapter 4). While field experimentalists are laboriously manipulating species abundances on tiny plots, analogous on-going manipulations on a gigantic scale are receiving little attention (e.g., expansions of parasitic cowbirds in North America and the West Indies, decimations of native fish by introduced piscivores, and on-going elimination of American elms by disease). In Chapter 3, Figs. 3.1 to 3.8 illustrate how FEs involving experimental reductions in populations of selected mammalian herbivores are labor-intensive but illuminating as regards resulting changes in ants, plants, other mammals, and even pathogens. Did any ecologist seize the opportunity that an NTE provided to measure such changes while myxomatosis was decimating Australia's rabbits?

Suggested Improvements in NSEs

Because NSEs provide the sole way to study many types of manipulations, it is not possible to renounce them because of the increased uncertainty and extra burden of evidence involved. How can this burden be met more convincingly than is now usually the case?

Useful guidance may be obtained from epidemiology, a science that for a long time has had formalized, successful procedures for dealing with an analogous problem. "Epidemiology is concerned with the patterns of disease occurrence in human populations and of the factors that influence these patterns. . . . The major epidemiologic problem in evaluating a statistical relationship is to determine whether or not the association is indirect or of etiological significance" (Lilienfeld and Lilienfeld 1980, pp. 3 and 296). These goals parallel the goals of the ecologist who employs NSEs to understand the factors underlying species abundances and distributions, and for whom the main practical problem is to distinguish indirect associations from causal (etiologically significant) associations. Epidemiology texts illustrate how the practicing scientist's use of the word "cause" must differ from that employed by the pure statistician or logician. In effect, one's decision between a causal association and an indirect association depends on details of the association and on auxiliary evidence. As summarized by Lilienfeld and Lilienfeld (1980) in their chapter entitled "The Derivation of Biological Inferences from Epidemiologic Studies," this evidence is of the following sorts:

> randomized experimental interventions on human populations, but these are usually not feasible or ethical (no. 1; numbers refer to the corresponding evidence that ecologists can employ with NSEs, as explained below);
>
> "a natural experiment [that] closely simulates the conditions of a randomized, controlled study and thus offers a unique opportunity to establish a causal inference" (Lilienfeld and Lilienfeld 1980, p. 315; no. 2 below);
>
> consistency and strength of the association over various populations (nos. 3 and 4);
>
> comparisons of different but related types of populations (no. 5);
>
> existence of a dose-response relationship between the putative causal factor and the effect (no. 6);
>
> specificity of the association (nos. 7 and 8);
>
> biological plausibility of the putative causal association (no. 9); and
>
> determination of the causal sequence connecting the putative ultimate cause and the effect (nos. 9 and 10).

Let us now examine the corresponding evidence that ecologists can employ with NSEs, as illustrated by some cases where such evidence actually has been employed. Several of these lines of reasoning may also be used in FEs involving an unnatural perturbation (e.g., fencing) for which it is unclear what proximate consequence is responsible for the ultimate outcome.

1. ***Carry out an FE*** Just as FEs can be generalized by NSEs, it may be feasible in some cases to strengthen NSEs by FEs. For example, an NSE showed that the vole *Microtus pennsylvanicus* has lower abundance, incidence, and range of occupied habitats on islands with than without a predator, the shrew *Blarina brevicauda* (Lomolino 1984). This suggests reduction of vole populations by shrew predation. Lomolino strengthened this interpretation by an FE: he introduced shrews to three islands initially harboring voles but no shrews, and found that voles declined in abundance or became extinct. Similarly, Hixon (1980) used FEs to test inferences from NSEs about marine fish communities, while Roughgarden (Chapter 30) and Schoener (Chapter 33) did so for *Anolis* communities. Chapter 3 summarizes the extensive FEs that Brown and colleagues have performed on desert rodent communities, stimulated by Brown's earlier NSEs with these communities.

2. ***Carry out an NTE*** The clearest examples come from fish communities of lakes (Chapter 21). Comparisons of lakes with and without fish, and of lakes containing a certain fish species with and without a competing or predator fish species, often reveal striking shifts in mean size of zooplankton and in the abundance, diet, body size,

ionship
ınd the

and 8);

causal

'e con-
ınd the

ding evi-
NSEs, as
evidence
of these
ı FEs in-
fencing)
te conse-
ıutcome.

ı be en-
in ne
example,
pennsyl-
ıce, and
vith than
evicauda
n of vole
.omolino
E: he in-
harbor-
at voles
t. Simi-
ferences
s, while
(Chap-
hapter 3
wn and
nt com-
r NSEs

camples
Chapter
ıut fish,
ies with
pecies,
of zoo-
ly

---

and habitat preference of fish. These same types of changes have been witnessed directly in numerous NTEs following the intentional or accidental stocking of lakes with fish.

3. Is the association consistent over many sites? NSEs that compare only a single experimental site and single control site are as dissatisfying as the correspondingly flawed FEs. If outcomes are confirmed for many experimental sites and many control sites, one can be more confident that the cause of the outcome is either the identified difference between the sites or else something closely correlated with it. For example, Toft et al. (1982) were able to examine duck distributions on 236 ponds within a small area, censused for four consecutive years, in their analysis of how five duck species affected each other's presence and abundance. Diamond (Chapter 3) examined how the New Guinea warblers *Sericornis virgatus* and *S. nouhuysi* affected each other's altitudinal range by comparing 9 mountains with both species, 14 with *S. virgatus* alone, and 23 with *S. nouhuysi* alone.

[illegible]

4. Is the association consistent over related populations that are expected to behave similarly with respect to the putative causal link? Chipmunks of the *Eutamias quadrivittatus* group and *E. dorsalis* segregate by altitude where they coexist, but each species expands to occupy the other's altitudinal range on at least 18 mountains supporting only one of the two species. Not only is this pattern consistent over many sites, but it also is exhibited by two different species of the *quadrivittatus* group, *E. quadrivittatus* itself and *E. umbrinus* (Brown and Gibson 1983). Similarly, pairs of New Guinea honey-eater species of the *Melidectes belfordi* group segregate by altitude where they co-occur, while each species expands its niche by several thousand feet to occupy the whole altitudinal transect where it occurs alone (Chapter 6). This pattern is exhibited by all three pairwise combinations of the three species (*belfordi/ochromelas*, *ochromelas/[rufocrissalis]*, *[rufocrissalis]/belfordi*), as well as by all three allospecies of the *[rufocrissalis]* superspecies (*rufocrissalis*, *leucostephes*, *foersteri*). Schoener's (1975) study of habitat shifts in the three most widespread West Indian anoles encompassed 20 sites offering nearly all existing species combinations involving these species.

5. Does the association vary predictably over related populations that are expected to behave differently with respect to the putative causal link? Schoener and Toft (1983a; Toft and Schoener 1983) found that abundances of several diurnal spider species were about 10 times lower on islands with than without diurnal lizards. The authors attributed this difference to lizard predation on and competition with spiders. In support of this interpretation, they found little or no effect of lizard presence or absence on nocturnal spiders that would escape predation and competition, nor on a spiny, brightly (perhaps aposematically) colored spider likely to escape predation, and only an inconsistent effect on a spider species whose adults are nocturnal and whose diurnal juveniles may be too small compared to lizards to suffer competition. In Schoener's (1975) study of site-to-site habitat shifts among lizard species identified as potential competitors, the shifts were greater for species pairs with the same than different climatic preferences, greater for similar- than dissimilar-sized species and size classes, and greater for the effect of large lizards on small ones than vice versa. These variations in outcome are consistent with an interpretation of competition but would otherwise be hard to explain. Among New Guinea montane bird species pairs, altitudinal shifts are often observed in the niche of one species between sites differing in the presence and absence of a related species. These shifts decrease in frequency with the size ratio, taxonomic distance, and differences in diet and foraging technique between the two species, suggesting that the shifts arise from competition (Chapter 6).

point ?

6. Does the effect size increase with the magnitude of the putative causal factor ("dose-response relationship")? Among 30 sites, Yeaton and Cody (1974) found song sparrow territory size to vary nearly 20-fold, from 0.06 to about 1.15 ha. One observation suggesting varying interspecific competition as the explanation

point ?

of this variation is that territory size increased regularly with variations in the number of competing species from 0 to 13 (Spearman Rank Correlation Coefficient, 0.93). Negative associations (suggestive of competition) between distributions of duck species increased with interspecific niche overlap along spatial and temporal axes (Toft et al. 1982).

7. ***Specificity of association: does the association between two variables persist over a wide range of variation in other variables?*** If so, it becomes unlikely that the other variables are causally involved in the association. In effect, one uses NSEs to explore the natural range of variation over which the association applies. For example, in the Bismarck Archipelago the whistler *Pachycephala melanura* disperses much more readily than the very similar *P. pectoralis* but is absent as a resident on every island occupied by *P. pectoralis*. On islands lacking *P. pectoralis, P. melanura* occupies all available light levels from zones of bright sunlight to nearly permanent cloud cover, all available moisture levels from zones with a 4-month annual drought to constant mist, all available elevations from sea level to the highest summit at 4,650 ft, all available vegetation heights from gardens and coral scrub to the canopy of tall forest, all available habitat densities from open savanna to dense subalpine shrubbery, and all available island sizes from 1 ha to 150 mi^2. This range of variation compatible with residence by *P. melanura* increases the likelihood that its absence from all *pectoralis*-occupied islands is due to the presence of *pectoralis* rather than to some habitat or climatic factor (Diamond 1975).

8. ***Specificity of association: effects of other variables*** Another approach to the question of specificity of association is to measure the effects of site differences other than the one initially suspected of providing the causal explanation. For example, if one observes niche shifts in species A associated with the presence or absence of species B, one can measure effects of habitat structure, predation, resource levels, and island area or isolation on species A's niche, in order to test whether these variables rather than B's presence are the ultimate cause of the niche shifts. In their study of island variation in song sparrow territory size, Yeaton and Cody (1974) measured insect abundance and amount of foliage at each site. The former factor proved to bear no relation to territory size, while the latter factor varied in a direction opposite to that required to explain variations in territory size. Lomolino (1984) found that neither island area nor habitat characteristics explained the observed negative relation between vole abundance and shrew presence. Toft and Schoener (1983) found that island spider numbers increased with island area and vegetation height and decreased with island distance, but that the large effect of lizard presence or absence persisted over the range of vegetation heights, over a range of distances up to at least 2.3 km, and over a range of area from 50 to at least 10,000 m^2. Schoener (1975) identified the habitat parameters of greatest predictive value for lizard distribution, measured them at each of 20 sites for each size class of each of 10 lizard species studied, and used the measurements in several different ways to calculate what the habitat niche of a given species would be if other sympatric species were without effect. The calculation confirmed the existence of many site differences in a given species' niche related to the presence or absence of other species and unrelated to habitat variables. Depending on the particular case, the effects of other species became variously more, less, or equally conspicuous after correction for habitat parameters. Critics of NSEs often implicitly assume that, if habitat and resource variables were measured, effects of species interactions inferred from NSEs could only turn out to be weaker than claimed or else nonexistent. Schoener's (1975) result, and that of Yeaton and Cody, illustrate the obvious point that the effects may also turn out to be stronger.

9. ***Is there a biologically plausible mechanism for the relative causal relationship?*** In support of his inference that presence of shrews caused reduced abundance of voles on islands, Lomolino (1984) obtained evidence making predation a plausible mechanism: most shrew scats contained vole fur, shrews preferentially eat juvenile voles, and juveniles practically disappeared from vole populations following introductions of shrews. In support of their inference that niche shifts in altitudinally segregating chipmunk species arise from interspecific competition, Brown (1971a), Heller (1971), and Sheppard (1971) quantified interspecific aggression at the transition altitude and in laboratory cages. In support of their inference that island variation in song sparrow territory size was caused by varying numbers of competing species, Yeaton and Cody (1974) measured niche overlaps between the competitors and song sparrows along three niche axes, and measured contraction of song sparrow vertical foraging range from an island with 3 competing species to an island with 11 competing species.

10. Can the observed effect be predicted quantitatively from the putative causal relationship and its observed mechanism? On the assumption that song sparrows progressively expand their niche with decreases in number of competing species, Yeaton and Cody (1974) were able to predict the varying values of island song sparrow territory size by considering the species present on each island and their niche overlap with song sparrow.

Conclusion Uniquely among the types of ecological experiments, the manipulation in an NSE is not directly observed but is instead inferred from site differences. Such an inference bears the obvious burden of demonstrating that the claimed difference rather than some other difference between the sites really caused the outcome. One can never be certain. However, the preceding section has outlined a 10-point strategy, paralleling that used by epidemiologists for testing the putative explanation as well as competing explanations. Most NSEs employ only a small fraction of this strategy, generally point 9, occasionally point 3 as well. I have frequently cited the NSEs by Yeaton and Cody (1974), Schoener (1974), Toft et al. (1982), Toft and Schoener (1983), and Lomolino (1984) because they went to exceptional effort to test alternative causal explanations. Schoener's (1975) test of habitat variables that might confound NSE-based conclusions about lizard competition was so detailed that it has never been attacked by skeptics of competition, but it has also never been imitated by believers in competition. Wider adoption of such methods could reduce the skepticism that ecologists accustomed to LEs and FEs feel toward NSEs.

WHICH IS THE BEST TYPE OF EXPERIMENT?

It will now be clear that this is a silly question, because the answer varies with the species, process, and site studied. NEs tend to have advantages over LEs for studying species interactions among whales; the advantage is reversed for yeasts. If one wants to understand the effects of introducing predator A on rodent B, FEs are often the method of choice to study the changes in abundance after 1 year; NEs are the sole method capable of studying the genetic changes after 10,000 years. FEs are certain to result in expulsion or incarceration for scientists working in national parks or in certain countries; NEs are not. Thus, a pluralistic approach is essential if one wishes to address the range of questions arising in community ecology.

It is therefore puzzling that the ecological literature contains few discussions of the relative merits of all three types of experiments and frequent claims of principled superiority for one type. The failure to utilize the full array of appropriate methods represents a serious impoverishment for ecology today, and its origin warrants careful examination.

Certainly, an ultimate explanation is that ecologists rarely receive equal training in the three approaches. In the past many gifted natural historians lacked experience in laboratory techniques, field experimental methods, or statistics. Today, in part as a backlash, good training in field experiments is infrequently accompanied by equal experience in natural history or the laboratory. My impression is that principled claims of superiority for one type of experiment are most likely to be made by ecologists familiar mainly or solely with that type. However, we still must consider the arguments used to buttress the claims. Most of them seem to me to fall into four categories.

1. Focusing on one consideration One of the commonest arguments is to cite correctly the inherent advantages of method A, the inherent disadvantages of method B, and the particular shortcomings of a certain study employing method B, while ignoring the inherent disadvantages of method A, the inherent advantages of B, and so on. For instance, field experimentalists wishing to make a general case for FEs over NEs often cite the real advantage of FEs in manipulative control and site matching, while ignoring the real advantages of NEs in maximum spatial and temporal scales and in the range of manipulations that can be studied. Similarly, critics of LEs and FEs often cite their unrealism and limited scale, respectively, while ignoring their compensating advantages.

2. Pattern, effect, process, and mechanism There is much confusion about these terms. It is sometimes claimed that FEs reveal only "effects," while LEs alone can reveal "mechanisms"; or that NEs or nonexperimental observations yield only "patterns," while FEs alone can identify underlying "mechanisms" or "processes" (e.g., Wilbur and Travis 1984, pp. 113–114).

Actually, what any community ecological experiment attempts to get at is more accurately described as a treatment effect or a cause-effect sequence: the change in X (usually a species abundance or distribution) caused by a change in the physical environment or in the abundance of some other species. In such a case the expression "underlying mechanism or process" would apply to the physiological response or species interactions underlying this experimentally demonstrated change in abundance and distribution. As noted by Schoener (1974a, p. 28; 1983b, p. 275) for FEs (his comment is equally valid for LEs and NEs), the experiment itself rarely investigates the mechanism of the change, although most experimenters nevertheless proceed to offer an opinion about mechanism. Supplementary data, especially direct physiological measurements or else observations of behavior, will always be required to identify the mechanism: for example, to show whether species B reduces species A's abundance by preying on, infecting, parasitizing, driving off, eating the food of, or eating the leaves of species A. Whether the population consequences of this observed behavior or physiological response are then best revealed by LEs, FEs, or NEs depends on the considerations of Table 1.1 applied to each particular case.

3. Popperphilia In the past decade, a position in the philosophy of science—the falsificationist criterion espoused by Popper—has been abused by some ecologists to support their position in various controversies, such as what type of experiment possesses inherent superiority. The first problem is that, as already noted, this question is not worth posing in such a general form. The second problem is that, even if one espouses a falsificationist criterion, that does not guide one to a preference among LEs, FEs, NTEs, or NSEs. Finally, while Popper's philosophy in general and the falsificationist criterion in particular were formerly considered to be among the significant views within the philosophy of science, most professional philosophers other than Popper's disciples abandoned these views by about 20 years ago. For example, in the comprehensive presentation of major modern philosophies of science edited by Suppe (1977), the falsificationist criterion is cited only briefly in a few places to explain why particular modern philosophers discarded it. Similarly, Popper's overall philosophy is not among those presented at length and still given serious consideration, but is instead summarized briefly as background for understanding its shortcomings and its influence on modern philosophers such as Feyerabend and Lakatos. The recent explosion of Popperphilia among ecologists exemplifies what the philosopher Suppe (1977, p. 19) had in mind when he wrote, "It seems to be characteristic, but unfortunate, of science to continue holding philosophical positions long after they are discredited."

4. Models from other sciences A recurrent point of view among ecologists is that ecology would progress faster by imitating the rigorous approaches of hard sciences like physics and molecular biology (cf. Strong, Simberloff, Abele, and Thistle 1984, p. viii). In fact, different sciences face different methodological problems, and the successes of physics and molecular biology were made possible by those fields having found appropriate solutions to their particular problems. Ecology must cope with its own distinctive problems. Insofar as ecologists can profitably be guided by other sciences, the best models would be sciences facing problems similar to those of ecology. Molecular biology and some areas of physics are methodologically at the opposite pole from ecology and offer poor guidance. Better models are astronomy, geology, medical science, neuroethology, or vulcanology, which share with ecology practical limitations on experimental intervention, concern with historical or evolutionary problems, and conditionality of results. In those fields the recognition that laboratory studies, field studies, and the comparative approach have a place is no longer controversial, as it still is in ecology. Probably the single most important methodological advance contributing to the recent successes of neuroethology has been the intimate feedback among these approaches: the same investigator *regularly* applies laboratory and field approaches to the same problem (cf. Roeder 1967). For instance, the contributions of Hopkins (1980; Hopkins and Bass 1981) to electric communication in fish have been based on comparing the ecology and electric discharges of 23 fish species (NE), carrying out field playback experiments with discharge patterns (FE), and doing electrophysiological studies on electric organs in the laboratory (LE).

FUTURE PROSPECTS

Prospects for Each Methodology

All three experimental methods could be applied in ecology with more power than they usually are at present. To reiterate briefly the points made in this chapter:

- Laboratory experiments could be applied to a much wider range of species, could be run on a larger spatial scale, could utilize certain natural communities brought indoors, could utilize more complex and realistic synthetic communities, and could be extended to more LE-FE hybrids studied outdoors.
- Field experiments could be applied to more species and problems, could more consistently employ good design, could be run for longer times, could use closed natural communities, and could be integrated with NEs to assess generality. The differences between pulse and press experiments, and between direct and indirect effects, need wider appreciation.
- Natural snapshot experiments (NSEs) could use a whole series of techniques, paralleling those in use in epidemiology, for increasing the strength of inferences about cause. Far more perturbations launched by nature or by humans (NTEs) merit analysis.

weak

Effect Size

Most FEs and NEs still ask only *whether* or not an effect is detectable: for example, whether the presence of species A alters species B's abundance and distribution at the $p < 0.05$ level. It is astonishing how few FEs and NEs go on to study effect size: that is, by *how much* various densities of species A alter B's abundance (Toft and Shea 1983). Such information is essential before ecologists can start to construct explanatory, semiquantitative theories of ecological communities.

Prospects for an Integrated Research Strategy

Ecologists, like scientists in many other fields, can profit by applying different methodologies to the same system. Examples to date include several studies of competitive exclusion in which the same investigators combined two or three approaches. For Caribbean lizards, Roughgarden and colleagues (Chapter 30) combined FEs and NSEs, as did Abramsky and Sellah (1982) for Mediterranean rodents, Grant (1969a) for north temperate rodents, Brown and colleagues (Chapter 3) for desert rodents, and Hixon (1980) for marine fish. LEs were combined with FEs in studies of marine invertebrates by Buss (Chapter 31), and with NSEs in studies of chipmunks by Sheppard (1971) and Heller (1971). Harger (1972) used LEs, FEs, and NSEs to analyze mussel distributions. Evidence from both NSEs and NTEs as well as LEs and FEs was used by Werner (Chapter 21) and Colwell (Chapter 24) to understand communities of fresh-water fish and hummingbird mites, respectively. Schoener and colleagues (Chapter 33; Schoener and Schoener 1984) combined NEs, FEs, and LEs to elucidate why Bahamian lizards require a minimum island area; they compared natural lizard communities on islands of different areas, introduced lizards to lizard-free islets, and studied lizards' overwater dispersal abilities in the laboratory.

These combined studies are exceptionally convincing, because conclusions tested by different methodologies become more robust. These studies also provide a much more complete understanding of their systems than a single-method study could, because each methodology yields some information inaccessible to the others. Finally, the combined studies accelerate progress, as questions raised by results from one methodology lead to further experiments by a different methodology. Thus, the greater use of multiple methodologies in an integrated research strategy may prove valuable in community ecology.

ACKNOWLEDGMENTS

This chapter profited greatly from the criticisms of James Brown, Stuart Hurlbert, Peter Kareiva, and Thomas Schoener.